

ASSESSING THE ENVIRONMENTAL IMPACTS OF SHARED AUTONOMOUS
ELECTRIC VEHICLE SYSTEMS WITH VARYING ADOPTION LEVELS
USING AGENT-BASED MODELS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Mustafa Lokhandwala

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2019

Purdue University

West Lafayette, Indiana

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL**

Dr. Hua Cai, Chair

School of Industrial Engineering, Purdue University

Environmental and Ecological Engineering, Purdue University

Dr. Seokcheon Lee

School of Industrial Engineering, Purdue University

Dr. Roshanak Nateghi

School of Industrial Engineering, Purdue University

Environmental and Ecological Engineering, Purdue University

Dr. Konstantina Gkritza

School of Civil Engineering, Purdue University

Department of Agricultural and Biological Engineering, Purdue University

Dr. Hong Wan

Department of Industrial and Systems Engineering, North Carolina State
University

Approved by:

Dr. Steven Landry

Head of Industrial Engineering, Purdue University

To my advisor Dr. Hua Cai for getting me interested in research without saying word, and for being an amazing mentor.

To my dad, mom and brother Khizer, Munira and Mohammed Lokhandwala for their continuous support through these four years.

To Umema Bohari, for keeping me sane through the toughest moments of this degree.

I owe each of you big time!

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	xi
ABBREVIATIONS	xx
ABSTRACT	xxii
1 Introduction	1
1.1 Ride Sharing	4
1.1.1 Ride Matching	6
1.1.2 Heterogenous Rider Preference in Ride Sharing	8
1.2 Ride Sharing using AVs and non-AVs	9
1.3 Electric Vehicles and Integration with SAVs	15
1.4 Research Gaps	17
1.5 Research Structure	18
2 Parameterized Preference-based Shared Autonomous Electric Vehicle (PP- SAEV) Model	21
2.1 System Level Parameters	22
2.2 Agent Behaviors	24
2.3 Model Outputs	28
2.4 Model Assumptions	30
2.5 Data and Exploratory Analysis	41
2.6 Modeling Taxi Shifts (for traditional taxi)	43
2.7 Summary	44
3 Ride sharing using shared autonomous vehicles and traditional taxi cabs	45
3.1 Introduction	45
3.2 Simulation Scenarios	46
3.3 Results	47
3.3.1 Fleet Reduction	47
3.3.2 Increased Resource Utilization	52

	Page
3.3.3 Environmental Benefits	53
3.3.4 Spatial coverage change	55
3.3.5 Changes in efficiency of matching	56
3.4 Conclusion	58
4 Ride Sharing Adoption in Mixed Autonomous Vehicles and Traditional Vehicle Fleets of Varying Capacities	60
4.1 Introduction	60
4.2 Literature Review	63
4.3 Method	66
4.4 Case Study Results	69
4.4.1 Fleet Sizing	75
4.4.2 Analysis of Key Output Parameters	77
4.4.3 Studying the efficiency limitations of ride sharing	84
4.5 Conclusion	95
5 Understanding The Impact of Heterogeneous Rider Preferences on a Shared Autonomous Vehicle System	97
5.1 Introduction	97
5.2 Parameter Variation using Mixture Experiments	99
5.3 Case Study Results	102
5.3.1 System level implications	102
5.3.2 Individual rider types	110
5.4 Conclusions and future scope	116
6 Siting Charging Stations for Incremental Electric Vehicle Adoption in Traditional Fleets and Future Shared Autonomous Vehicle Fleets	118
6.1 Introduction	118
6.2 Literature Review	120
6.3 Method	131
6.3.1 Key Assumptions	132
6.3.2 Demand generation using PP-SAEV	134
6.3.3 Optimization of charging station locations	135
6.3.4 Objective function calculation	141
6.4 Case Study	143
6.4.1 Siting charging stations for the case study	146
6.4.2 Case Study Analysis	156
6.4.3 Environmental Impacts	162

	Page
6.5 Conclusion and Future Scope	164
7 Conclusion	168
7.1 Key Results from Case Studies	170
7.2 Insights for SAEV System Development	173
7.3 Limitations and Future Research Directions	174
8 References	180
9 Curriculum Vita	200
9.1 Education	200
9.1.1 Academic Qualifications	200
9.1.2 Publications	200
9.1.3 Presentations	201
9.1.4 Teaching Assistant Experience	202
9.1.5 Mentorship	202
9.1.6 Projects	203
9.2 Work Experience	203
9.3 Technical skills	204
9.4 Achievements	204
9.4.1 Grants	204
9.4.2 Awards	205
9.5 Leadership	205
9.6 Affiliations	206
A Pseudo Codes for PP-SAEV Model	207
A.1 Main Program and Input Data	207
A.1.1 Global Parameters	208
A.1.2 Main Algorithm	210
A.2 Rider Algorithm and Data	211
A.2.1 Parameters and Variables	212
A.2.2 Rider Algorithm - <i>riderAlgo(Group g)</i>	214
A.2.3 Searching Strategies - <i>searchForTaxis(Group g)</i>	215
A.3 Taxi Algorithm and Data	216
A.3.1 Taxi Algorithm	217
A.3.2 <i>precheck(Group g, Taxi t, Array< Groups > A)</i>	222
A.3.3 <i>bestRoute(Group a, myRoute)</i>	226
A.3.4 Shift Start and End	229
A.3.5 Electric Vehicles Charging	232

	Page
A.4 Speed Estimation	234
B Supplementary Analysis for Studying Ride Sharing in Shared Autonomous Vehicles and Traditional Taxi Cabs	238
B.1 Sensitivity of Busy Start Percent	238
B.2 Explanation of Spatial Coverage Changes	238
B.3 Statistical Significance Tests	240
C Scenarios Generated by Simulation Optimization	243
D Scenarios from Mixture Experiments	274
E Regression models for Mixture Experiments	288
F Genetic Algorithm to Site EV Charging Stations	302
F.1 Genetic Algorithm - Details	302
F.1.1 Generate initial solutions (parents)	302
F.1.2 Filter initial candidate set	307
F.1.3 Genetic operators	309
G Progression of Objective Function in Genetic Algorithm	315
H Sensitivity Analysis	319
H.1 Budget Sensitivity	319
H.2 Sensitivity of C_N for $C_u = 1$	320

LIST OF TABLES

Table	Page
2.1 Sample data from NYC-TLC with pick up (PU) and drop off (DO) time and locations in latitudes (Lat) and longitudes (Long), and the number of passengers traveled together (Group Size)	23
2.2 Settings for the <i>regLimit</i> , <i>shareLimit</i> and <i>exitTime</i> parameters . . .	33
3.1 Simulation scenarios to understand ride sharing using shared autonomous vehicles and traditional taxi cabs	47
3.2 Scenarios and their parameter sets that can provide service within 2% of the base scenario, during the evening peak. We will from here on refer to these scenarios by the label.	48
4.1 Simulation scenarios to study ride sharing adoption in mixed autonomous vehicles and traditional vehicle fleets of varying capacities	62
4.2 The new designs that were used to generate the response surface. Each cell is a vector of 2 numbers which correspond to the low, and high settings for that variable.	71
4.3 Significant terms in the least squares regression	73
4.4 Regression model for waiting time $RMSE = 8.81s$ and $R^2 = 0.97$. . .	80
4.5 Cluster centers obtained using k-means clustering on <i>percentSharing</i> and <i>percentAV</i> from the comparable scenarios in Table 4.1	82
4.6 Sample taxi status table for a scenario with 3000 taxis. Each row sums to the total number of taxis in the system, in this case 3000	85
5.1 Simulation scenarios for sharing preference	99
5.2 Simulation scenarios to study heterogeneous sharing preferences in ride sharing	103
5.3 Regression model for served PC. $RMSE = 0.057$ and $R^2 = 0.86$	104

Table	Page
6.1 Gaps in recent literature in EV charging station siting. Station Counts refers to whether the number of charging stations were decided by the model, Charging Ports refers to whether the number of ports at each EV charging charging stations were decided by the model, Queuing at Station refers to whether the model considered that EVs would wait idle at a charging station if currently occupied, Multiperiod refers to whether the literature sited charging stations for multiple periods in time.	128
6.2 Sample demand data generated from the agent-based model	135
6.3 Scenarios for which EV charging stations are sited	147
A.1 Boundary points defining the regions of NYC	208
A.2 Speeds on NYC roads (m/s)	237
C.1 All runs for case study in Section 4.4	243
D.1 Mixture design for case study in Section 5.3	274
E.6 Significant terms in the least squares regression	288
E.1 Regression model for percent shared rides $RMSE = 0.055$ and $R^2 = 0.92$	290
E.10 Significant terms in the least squares regression	290
E.2 Regression model for average taxi capacity $RMSE = 0.095$ and $R^2 = 0.91$	293
E.3 Regression model for served fraction for rider type 1 $RMSE = 0.039$ and $R^2 = 0.96$	294
E.4 Regression model for Served fraction for rider type 2 $RMSE = 0.020$ and $R^2 = 0.96$	295
E.5 Regression model for served fraction for rider type 3 $RMSE = 0.016$ and $R^2 = 0.97$	296
E.7 Regression model for served fraction for rider type 5 $RMSE = 0.015$ and $R^2 = 0.60$	297
E.8 Regression model for waiting time for rider type 1 $RMSE = 13.4$ and $R^2 = 0.95$	298
E.9 Regression model for waiting time for rider type 2 $RMSE = 24.17$ and $R^2 = 0.99$	299

Table	Page
E.11 Regression model for waiting time for rider type 4 $RMSE = 8.38$ and $R^2 = 0.99$	300
E.12 Regression model for waiting time for rider type 5 $RMSE = 19.09$ and $R^2 = 0.90$	301

LIST OF FIGURES

Figure	Page
1.1 Dynamic point-to-point ride sharing - the three rides shown can be combined to a single ride by a car with a capacity of 4 with a small amount of delay in arrival for each rider. (PU- Pick Up, DO - Drop Off) . . .	6
1.2 Research overview	20
2.1 Model overview. Matching colors in the flowchart represent an event / decision taking place simultaneously by the rider and the taxi. The circular blocks labeled A, B, C, D are used as "connectors" to imply a transfer of flow to their corresponding connector	24
2.2 The relation between Algorithms 1-7 in the agent-based model	28
2.3 Time stamps recorded for each rider group in the system. e is the time that the rider group enters in the system; t_r is the time that the taxi responds to the rider group to commit a pickup; t_{p1} is the time that the rider group is picked up; and t_{d1} is the time that the rider group is dropped off at its destination.	30
2.4 NYC taxi demand represented as pickups. (a) is a temporal histogram for the pick ups every 15 minutes, (b) shows the spatial density of taxi pick ups in NYC. (Grid resolution is $0.005^\circ \times 0.005^\circ$, roughly equivalent to $0.5km \times 0.5km$). (c) Histogram for the number of passengers in a rider group. (a) and (b) plots use the data from the date 8/24/2014. (c) uses data from the year of 2014.	42
2.5 Average number of taxis in operation (on shift) every minute (NYC DOT, 2014)	44

Figure	Page
3.1 The average fraction of served rider groups (the ratio of rider groups served by the taxis to the total ride groups) in the morning (7:01am-3:00pm) and evening (5:01pm-12:00am) peak periods with different sharing and vehicle type scenarios. The light-red band indicates a service level within 2% of the base scenario	49
3.2 Change in (a) waiting time (TR_W) and (b) ride time (TR_{Ride}) under different ride sharing participation and fleet size scenarios	50
3.3 (a) Average increase in trip distance for the riders. (b) Average fractional increase in distance for the riders. (c) The average fraction of individual utilized trip deviation relative to the individual acceptable level, presented as a violin plot. (A violin plot is two vertical density plots attached together at their bases. The vertical bar shows the range of the values while the horizontal width shows the density of the points at that value).	51
3.4 Sharing performance: (a) Change in occupancy in different scenarios with similar service levels; (b) The actual percentage of rider groups that participated in sharing; (c) Number of rider groups shared a ride together varied with service levels	53
3.5 (a) Total distance traveled with respect to the base scenario (Scenario B) for different scenarios. (b) The fraction of occupied distance (distance traveled to deliver passengers) for all scenarios	55
3.6 Spatial distribution change in terms of service level. (a) The fraction change in service level in Scenario A compared to Scenario B, calculated as $\frac{Service_B - Service_A}{Service_B}$. (b) The fraction change in service level in Scenario B compared to Scenario A3, calculated as $\frac{Service_B - Service_{A3}}{Service_B}$ (c) The fraction change in service level in Scenario B compared to Scenario S, calculated as $\frac{Service_B - Service_S}{Service_B}$ (d) The fraction change in service level in Scenario A compared to Scenario A3, calculated as $\frac{Service_A - Service_{A3}}{Service_A}$. (e) The fraction change in service level in Scenario A compared to Scenario S, calculated as $\frac{Service_A - Service_S}{Service_A}$. Blue color indicates Scenario B or S has better service level than Scenario A, while red color indicates Scenario A provides better service. Grid resolution is $0.005^\circ \times 0.005^\circ$, roughly equivalent to $0.5km \times 0.5km$	57

Figure	Page
3.7 Change in response time (TR_R) under different ride sharing participation and fleet size scenarios	58
4.1 Flowchart of meta-model based simulation optimization method to find comparable scenarios	69
4.2 The input parameters of all the scenarios that have equal service level. The color refers to the Capacity of the taxis (2 - red), (4 - blue) and (6 - pink)	76
4.3 (a)Errors in in-sample prediction (a) Cross validation errors (b) The partial dependencies of fleet size with Capacity (panels), sharing participation (x-axis) and fraction of AVs (y-axis), the contours are spaced at 1000 taxis.	78
4.4 Raster plot of (a) average fractional increase in travel distance for riders (b) Waiting time for riders	79
4.5 Partial dependency plots of waiting time (s) from the linear model (table 4.4) with (a) Fleet Size and Capacity (b) sharing participation and capacity (c) AV adoption and Capacity (d) $dtMode$ and percent sharing. In (d) the waiting time is represented by the colour of the contours in the plot of $dtMode$ and percent sharing.	81
4.6 CO ₂ per day for each scenario with fleet size. The $taxiCapacity$ is represented by the shape of the point and the color refers to the cluster centers obtained using k-means clustering on $percentSharing$ and $percentAV$. The coordinates of the cluster centers are in Table 4.5	82
4.7 Raster plot of (a) Average occupancy for riders (b) Average number of shares for riders	84
4.8 Scatter plot matrix of the pivot coordinates. The colours represent the cluster that the scenario belongs to based on the taxi status data (Table 4.6)	89
4.9 Scatter plot matrix of input data. The colours represent the cluster that the scenario belongs to based on the taxi status data (Table 4.6)	90

Figure	Page
4.10 Composition plot for the hierarchical clustering on the pivot coordinates of the central taxi status composition (Figure 4.8), in original coordinates colored by <i>percentSharing</i> . The y axis is the fraction of the taxis in each status (x-axis) for each level of <i>taxiCapacity</i> (column) and each cluster (row). Each path line of the composition plot represents a single scenario	91
4.11 Location and direction of taxis that can accept shared rides	94
5.1 Surface plots of the predicted service level using the model represented by 5.3 at different compositions of rider types at <i>fleetSize</i> = 6000 and <i>dtMode</i> = 1	105
5.2 Effect plots of the predicted percent shared rides using the model represented by Table E.1 at different compositions of rider types at (a) <i>fleetSize</i> = 4000 and <i>dtMode</i> = 0 (b) <i>fleetSize</i> = 6000 and <i>dtMode</i> = 1. I have plotted the points with large standard error of prediction ($se > 0.015$) using hollow circles. Each line on the effect plot belonging to rider type i indicates the predicted response if the amount of that rider type increased/decreased by a fraction of Δ and the rest of the mixture components decreased/increased in the same proportion as they were in the reference mixture $\bar{X} = \{0.2, 0.2, 0.2, 0.2, 0.2\}$	107
5.3 Effect plots of the predicted occupancy using the model represented by Table E.2 at different compositions of rider types at (a) <i>fleetSize</i> = 4000 and <i>dtMode</i> = 0 (b) <i>fleetSize</i> = 6000 and <i>dtMode</i> = 1. I have plotted the points with large standard error of prediction ($se > 0.005$) using hollow circles. Each line on the effect plot belonging to rider type i indicates the predicted response if the amount of that rider type increased/decreased by a fraction of Δ and the rest of the mixture components decreased/increased in the same proportion as they were in the reference mixture $\bar{X} = \{0.2, 0.2, 0.2, 0.2, 0.2\}$	108
5.4 rider type compositions for scenarios with service level between 0.85 to 0.86 ranked in decreasing order of	111
5.5 Box plots of (a) fraction of riders served and (b) Waiting time for the five rider types	112

Figure	Page
5.6 2D bar plots for the change in the (a) served fraction (b) waiting time (color) for a rider type on the x axis with a change in fraction of the rider type on the y axis between the values indicated by the panel labels when <i>fleetSize</i> = 6000 and <i>dtMode</i> = 1	113
5.7 2D bar plots for the change in the (a) served fraction (b) waiting time (color) for a rider type on the x axis with a change in fraction of the rider type on the y axis between the values indicated by the panel labels when <i>fleetSize</i> = 4000 and <i>dtMode</i> = 0	114
6.1 Adding a new charging station can decrease the total travel distance for charging. The blue dots indicate the charging demand (e.g., vehicles with the need for charging), the green chargers indicate the locations of the existing charging stations, the yellow charger indicates the additional charging station, and the dotted lines indicate the route to the closest charging station for each charging demand.	119
6.2 Flowchart of the method. The red dashed arrows represent an input and the blue dashed arrows represent an output. The colors in the flowchart represent the different parts of the method. green: demand generation (Section 6.3.2), orange: genetic algorithm (Section 6.3.3.1); yellow: objective function evaluation (Section 6.3.4)	131
6.3 Illustration of steps used in the objective function evaluation. The 3 charging stations each have 1 charging port. The dots represent the current locations of the demands. The color of the dot represents the demands <i>State</i> - Red = 1 (Entered the simulation and moving towards charging station); Light Yellow = 2 (queued at charging station); Yellow = 3 (charging); Green = Exiting. The direction of the black arrows progressing time steps. The boxes above the charging stations are the queuing areas for the charging stations.	145
6.4 Spatial visualization of the charging demands for each scenario. The contours denote the demand distribution for that scenario and are superimposed on a map of New York City bounded by the coordinates (-74.1, 40.55) and (-73.7,40.95). The demand concentrated in the Manhattan area.	149
6.5 Count of unrestricted charging demand for every 15 minutes with different EV adoption scenarios: (a) 10%, (b) 40%, (c) 70%, and (d) 100% . . .	150

Figure	Page
6.6 Future case : Best charging station siting plan as determined by the genetic algorithm. The red dots indicate the site and the numbers indicate the total number of charging ports at that station. The purple charging station labels indicate new stations built for that EV adoption level, while the yellow labels indicate station which exist from the previous epoch .	152
6.7 Present case: Best charging station siting plan as determined by the genetic algorithm. The red dots indicate the site and the numbers indicate the total number of charging ports at that station. The purple charging station labels indicate new stations built for that EV adoption level, while the yellow labels indicate station which exist from the previous epoch .	153
6.8 Mixed case: Best charging station siting plan as determined by the genetic algorithm. The red dots indicate the site and the numbers indicate the total number of charging ports at that station. The purple charging station labels indicate new stations built for that EV adoption level, while the yellow labels indicate station which exist from the previous epoch .	154
6.9 Present day charging station sites in NYC	155
6.10 (a) Objective function values reached in different adoption scenarios, (b) Scaled objective function value based on the total number of EVs, (c) Scaled travel time, and (d) Scaled waiting time. The scaled objective function value is the ratio of the objective function value to the number of EV taxis in that scenario. The scaled objective function can be interpreted as the average time additional time that an EV taxi wastes in moving to and waiting at a charging station	156
6.11 Comparison between the demands simulated based on the optimal charging station siting and the estimated unrestricted demands in terms of the number of vehicles charging at the charging station over time. The solid lines represent Case F, Case P, and Case M, while the dotted lines represent the unrestricted scenarios. (a)10%, (b)40%, (c) 70%, and (d) 100% EV adoption.	158
6.12 Count of vehicles queuing at the charging station. The solid lines represent Case F, Case P, Case M and the dotted lines represent the unrestricted scenario (a)10% (b)40% (c) 70% (d) 100%	159

Figure	Page
6.13 (a) The current power consumption profile of NYC (New York Independent System Operator, 2019), and (b) The power consumption by charging EVs for 100% EV adoption in the evaluated cases (color). The solid lines represent Case F, Case P, Case M and the dotted lines represent the unrestricted scenario (a)10% (b)40% (c) 70% (d) 100%	160
6.14 Service level for (a) rider demand peak (6:00pm to 12:00am) (b) charging demand peak (1:00 am to 7:00 am); The solid lines represent Case F, Case P, Case M and the dotted lines represent the unrestricted scenario (a)10% (b)40% (c) 70% (d) 100%	162
6.15 Change in service levels over the city between 100% EV adoption and 0% EV adoption for the entire day. The color in each cell corresponds to the change in service level, which is measured as the ratio of the change in the number of riders served by the two scenarios in that cell to the total number of riders in that cell	163
6.16 The distance travelled by (a) all the taxis (b) Only the electric taxis with increasing EV adoption	165
6.17 CO ₂ emissions in Tonnes for increasing EV adoption	165
A.1 Graphical illustration of the <i>PreCheck</i> algorithm - T : The current location of the taxi, P: The pick up point of the potential share, and D: The drop off point of the potential share. The yellow rectangle represents the largest bounding box formed by the points on the original trip chain and the current location of the taxi. The gray box is the bounding box formed with the drop off location of the potential share and the current location of the taxi. To pass the <i>preCheck</i> , the drop off point of the potential share should either lie in the yellow box or form a gray bounding box that covers the yellow box.	223
A.2 Percent of taxis which start their shift at different times of the day (NYC DOT, 2014)	232
A.3 Speeds on NYC roads as estimated from the data set	235

Figure	Page
A.4 (a) Reason for speed estimation, the green line represents the non-sharing route for the taxi in green color, while the red line represents the non-sharing route for the taxi in red color. The black dotted line represents the shared trip. While the speed of the road segments covered by the original non-sharing trips can be estimated using the historical trip data, we do not have information for the speed of the additional roads traveled due to the deviated sharing trip (highlighted in the yellow rectangle). (b) Illustration of the speed estimation process. Each taxi drops equidistant points along the path it travels. The average speed of the region (shaded in black) and the road (in yellow) is set as the average speed of all the points in that region or on that road.	236
B.1 Sensitivity of busy start percent	239
B.2 Radius of gyration	240
B.3 ANOVA to test for the statistical significance of the results with the Scenario and the random simulation Seeds (Seed) for (a) waiting time and riding time (b) distance travelled and deviation tolerance difference.	242
F.1 Steps to generate initial population for $nRand = 2$ and $reps = 3$: cluster the demand (dots) into $j = 1, 2, 3$ centers (stars). The color represents the associated cluster center; Snap the cluster centers to potential charging station (triangle) locations with probability proportional to inverse of distance; Assign charging stations at the spots snapped in and generate the chromosome with all other potential charging station locations set to 0 and those selected set to the number of ports at the station randomly. The number of ports in the station are set randomly with probability proportional to the number of demands associated with that cluster.	305

Figure	Page
F.2 Operations used in our genetic algorithm. The chromosomes here have a target budget of $B = 20$ for illustration: (a) Rank Selection: The $n_{best} = 4$ highlighted chromosomes are selected by the rank selection algorithm; (b) Crossover: The yellow genes from the first chromosome are interchanged with the green in the second; (c) Mutation: Two instances of mutation where the genes in yellow are switched; (d) Perturbation: The highlighted genes are randomly moved to a new location (e) Fission: The one highlighted gene is split into 2 (f) Fusion: The two highlighted genes are merged into one; (g) Correction: The non zero genes are edited to meet the budget requirements $B = 20$ while preserving the number of sites	308
G.1 Objective function progression for case F	316
G.2 Objective function progression for case P	317
G.3 Objective function progression for case M	318
H.1 Sensitivity of the (a) objective function (b) scaled objective function (c) waiting time (d) scaled waiting time (e) traveling time (f) scaled travel time for cases which defer by budget as per Section H.1 (color and line-type)	321
H.2 Sensitivity of the (a) objective function (b) scaled objective function (c) waiting time (d) scaled waiting time (e) traveling time (f) scaled travel time for cases which defer by C_N as per Section H.2 (color and line-type)	322

ABBREVIATIONS

EV	Electric Vehicles
AV	Autonomous Vehicles
RS	Ride Sharing
PP-SAEV model	Parameterized Preference-based Shared Autonomous Vehicle model
NYC	New York City
GHG	Greenhouse gas
CO ₂	Carbon Dioxide
US	United States
NHTS	National Household Travel Survey
ICV	internal combustion engine vehicles
VMT	Vehicle miles travelled
NYCTLC	New York City Taxi and Limousine Commission
PU	Pick up
DO	Drop off
SAV	Shared Autonomous Vehicles
SAEV	Shared Autonomous Electric Vehicle
non-AV	Non-autonomous Vehicle/ Traditional vehicle/ Shifted vehicle
ABM	Agent-based model
B	Base Scenario 13,500 traditional taxis and 0% sharing participation

A	5,500 autonomous taxis and 100% sharing participation
A2	6,000 autonomous taxis and 75% sharing participation
A3	7,000 autonomous taxis and 25% sharing participation
S	8,000 traditional taxis and 100% sharing participation

ABSTRACT

Lokhandwala, Mustafa Ph.D., Purdue University, August 2019. ASSESSING THE ENVIRONMENTAL IMPACTS OF SHARED AUTONOMOUS ELECTRIC VEHICLE SYSTEMS WITH VARYING ADOPTION LEVELS USING AGENT-BASED MODELS. Major Professor: Dr. Hua Cai.

In recent years, there has been considerable growth in the adoption and technology development of electric vehicles (EV), autonomous vehicles (AV), and ride sharing (RS). These technologies have the potential to improve transportation sustainability. Many studies have evaluated the environmental impacts of these technologies but the existing literature has three major gaps: (1) the adoption of these three technologies need to be evaluated considering their impact on each other, (2) many existing models do not evaluate systems on a common ground, and (3) the heterogeneous preferences of riders towards these emerging technologies are not fully incorporated. To address these gaps, this work studies and quantifies the environmental and efficiency gains that can be gained through these emerging transportation technologies by developing a Parameterized Preference-based Shared Autonomous Electric Vehicle (PP-SAEV) agent-based model. The model is then applied to a case study of New York City (NYC) taxis to evaluate the system performance with increasing AV, EV, and RS adoption.

The outputs from the PP-SAEV model show that replacing taxi cabs in NYC with AVs along with RS potentially can reduce CO₂ emissions by 866 metric Tones per day and increase average vehicle occupancy from 1.2 to 3 persons in vehicles with

passenger seating capacity of 4. A prediction model based on the PP-SAEV output recommends that 6000 vehicles are needed to maintain the current level of service with 100% AV and RS adoption using capacity 4 taxis. Taxi fleets with capacity 4 with high RS and low AV adoption are also found to have the least CO₂ emissions. Because the heterogeneous sharing preferences of riders have shown as the major limiting factor to ride sharing, these heterogeneous sharing preferences are further modelled. The results show that high service levels are achieved when all the riders are open to sharing, and the maximum service level is reached when 30% of riders will only accept shared rides and 70% of the riders are either indifferent to sharing or prefer to use ride sharing over riding alone. Additionally, the service level and waiting time of riders that are inflexible (will accept only shared or non-shared rides) are greatly impacted by varying mix of riders with different sharing preference. Finally, an optimization model was built to site charging stations in a system with continually increasing EV adoption. Using the best charging station locations, transforming a fleet of autonomous or traditional vehicles to electric vehicles does not significantly change the system service level. The results show that increasing the EV adoption in fleets with 100% RS and AV adoption reduced the daily CO₂ emissions by about 861 Tones and transforming a fleet of traditional taxi cabs to electric taxi cabs reduced the daily CO₂ emissions by 1100 Tones.

In summary, this dissertation evaluates the potential growth of autonomous vehicles, ride sharing, and electric vehicles in systems where riders may have heterogeneous sharing preferences, from a system performance's perspective and assesses the environmental impacts. The developed model and the insights gained from this study can inform policy makers to develop sustainable transportation systems incorporating the emerging transportation technologies.

1. INTRODUCTION

Urban road transportation systems, despite being one of the most important infrastructures in the modern world, often bear high levels of congestion and pollution. In 2017, the transportation sector accounted for 29% of all greenhouse gas (GHG) emissions in the United States (U.S.), which makes it the largest contributor of carbon emissions (U.S. DOE, 2019). Individual vehicle systems, such as taxis and personal vehicles, make up a large percent of on-road traffic. However, they are highly inefficient in operation, often running with less than maximum capacity. According to the recent National Household Travel Survey (NHTS) data (FHWA, 2011), over 90% of work-related trips were made using a personal vehicle¹. Additionally, FHWA (2016) found that the average occupancy of these personal vehicles was 1.6, showing that these systems have very low efficiency.

There are different methods of reducing the GHG emissions from vehicle transportation systems. Current solutions have focused on solving the problems at the vehicle level such as 1) improving the fuel economy by improving engine efficiency (Farrington and Rugh, 2000; Greene and Plotkin, 2011), 2) the use of alternative fuel such as biofuel and electricity in cities with appropriate power generation mix (Egbue and Long, 2012; Cai et al., 2016; Hawkins et al., 2013; FHWA, 2016), 3) the use of alternative individual mobility services such as bike sharing systems (Romero et al., 2012; Luo et al., 2019) or car sharing systems (Chen and Kockelman, 2016a;

¹A vehicle that is meant for point to point transit, unlike mass transit systems (bus, metro, monorail, etc.)

Galland et al., 2014), and 4) introducing connected autonomous vehicles that are predicted to bring about positive changes in vehicle ownership, usage and driving patterns (Taiebat et al., 2018). Other solutions involve system-level changes, for example 1) optimizing the allocation of existing taxi systems by presenting analytical solutions such as large scale spatio-temporal visualizations (Jianqin et al., 2015) or suggesting better taxi dispatching rules (Maciejewski and Bischoff, 2015) and demand forecasting (Xu et al., 2015; Salanova et al., 2014; Kim et al., 2011), or 2) ride sharing using either personal vehicles (like Uber, Lyft) or using shared taxis. In recent years, with the growth and acceptance of the sharing economy and information and communication technologies (ICT), ride sharing has emerged as a viable low cost avenue to reduce the transportation system's energy use and emissions. Among these solutions, there has been significant effort in introducing ride sharing (RS) (Fortune, 2016), electric vehicles (EV) (Tesla, 2019), and autonomous vehicles (AV)² (Waymo, 2019) to aid individual mobility.

The implementation of one of these technologies (RS, EV and AV) can have an impact on the performance of the other. For example, an advantage of using AVs in a RS system is that they could potentially reduce fleet sizes since they do not need to shift in and out of operation like traditional vehicles (non-AV) since human riders get fatigued and require breaks. Because AVs are available all day and do not require drivers, they could help reduce costs of fleet operation (e.g. labour wages for drivers). EVs have a large potential to reduce environmental emissions (Cai and Xu, 2013) in areas where electrical energy comes from clean sources (such as wind, or geothermal), and hence the adoption of EVs should be promoted. AVs may be able to accelerate the adoption of EVs, because without having drivers, the electric AVs will

²In this dissertation, autonomous vehicles refer to vehicles with level 5 autonomous driving (fully automated vehicles with no human intervention required)

not suffer from common operational inefficiencies that human driven EVs may have (for example range anxiety). Additionally, AVs can be made to follow a well defined charging schedule so that the loss of on-road vehicles due to charging can be limited. On the other hand, the introduction of EVs in a ride sharing system will potentially affect the systems performance parameters negatively. EVs require a long recharging time (approximately 30 minutes for 80% charge for a vehicle with a 35kWh battery and with a charger delivering 50kW electric power) as compared to gasoline vehicles which would require less than 15 minutes to fill their gas tanks (NYCTL (2013) observed that taxis spent about 6.5 minutes refueling per shift). EV charging stations are not as widespread as gas stations, and to sustain an increased EV adoption, widespread availability of EV charging stations is necessary. For a ride sharing system, potentially the limited range and long charging time may restrict EVs from accepting new ride shares, if at the completion of a ride the EV's state of charge is low. Additionally, the long recharge time and the sparsity of charging stations may cause many taxi drivers to search for new riders while charging at a charging station thus affecting the locations at which those taxis would be available.

Because RS, AV, and EVs could be adopted individually or together, it is important to discuss the relevant studies in RS individually, then discuss studies in which AVs and EVs are discussed as a part of a RS system. Therefore, the literature review is organized as follows: Section 1.1 discusses the concept of ride sharing and relevant literature in the field. Section 1.2 focuses on studies model ride sharing using AVs and non-AVs, while Section 1.3 discusses relevant literature pertaining to the synergy between ride sharing, electric vehicles, and autonomous vehicles. Section 1.4 discuss the gaps in the literature. Finally, Section 1.5 outlines the questions that this dissertation aims to answer and the plan for the research.

1.1 Ride Sharing

Unlike the other solutions proposed to reduce vehicular emissions, ride sharing does not require any special physical infrastructure for implementation and has been tested and implemented in various platforms, such as UberPool and Lyft Line. In this study, ride sharing is defined as sharing a vehicle with multiple groups of riders³ along a fully or partially overlapped route, serving the travel demands of multiple groups of riders in the shared vehicle (for example, a shared taxi). This is different from car sharing, for which a car owner or a fleet owner allows others to use their car when it is not in use (for example, ZipCar, car2go, UberX). Uber claims that about 20% of its rides globally are shared rides using UberPool (Fortune, 2016). While the social and environmental impacts of these ride service apps are still in debate (e.g., increasing the total vehicle miles traveled and the total number of cars on the road, competing with taxi drivers for jobs and public transit for riders) (NYC DOT, 2016b), increasing vehicle occupancy rate in private and public vehicles (e.g., taxis) through ride sharing still offers great opportunities in improving the transportation sector's efficiency.

Many earlier studies (Barth and Todd, 1999; Galland et al., 2014) have focused on the traditional ride sharing (i.e., car pooling), for which the ride sharing is pre-arranged (e.g., with friends, family members, and colleagues) and often has the same trip origins and/or destinations. For example, Caulfield (2009) analyzed one day's commute trip data (reported as part of a Census survey) in Dublin, Ireland and found that 4% of the respondents ride-share to work. They estimated that this ride sharing

³I consider a group of riders to be made up of one or more persons riding from and to the same origin and destination using a single request.

reduced 12,674 t of CO₂ emissions annually. Hong et al. (2017) proposed a clustering algorithm on GPS trace data to match trips and select routes for carpooling.

In recent years, enabled by the development of information and communication technologies, dynamic ride sharing has received increasing attentions. Dynamic ride sharing allows shared rides to form in short notice and among strangers who do not know each other's trip itinerary. The higher flexibility of dynamic ride sharing offers additional opportunity to maximize sharing benefits and improve system efficiency. Figure 1.1 presents an example of dynamic ride sharing where three rides are being combined into one single shared ride.

Group ride systems, as discussed in Qian et al. (2017), are a hybrid between dynamic ride sharing and static ride sharing. While the riders and their routes are not known ahead of time, the group ride system asks riders to walk to a common meeting location and ride together. Qian et al. (2017) tested the system using taxi trip data from 30-minute periods during peak and off-peak hours in three cities and concluded that this type of ride sharing can reduce vehicle VMT by over 47%. However, being different from the door-to-door service provided by traditional taxis, group ride requires the riders to walk to and from the taxi pick-up and drop-off locations, reducing the convenience of taking taxis. Additionally, the system can only match taxi trips that are of approximately equal length, and hence reduce opportunities for sharing.

In a ride sharing system it is important to find good matches for riders and drivers, and these problems are often difficult to solve. Hence, Section 1.1.1 discusses studies that focus on algorithms to match riders to taxis in a ride sharing system. Whether riders choose to use a RS system is also an important factor that affects the systems efficiency. The factors affecting this ride sharing choice are studied in literature and these studies are discussed in Section 1.1.2.

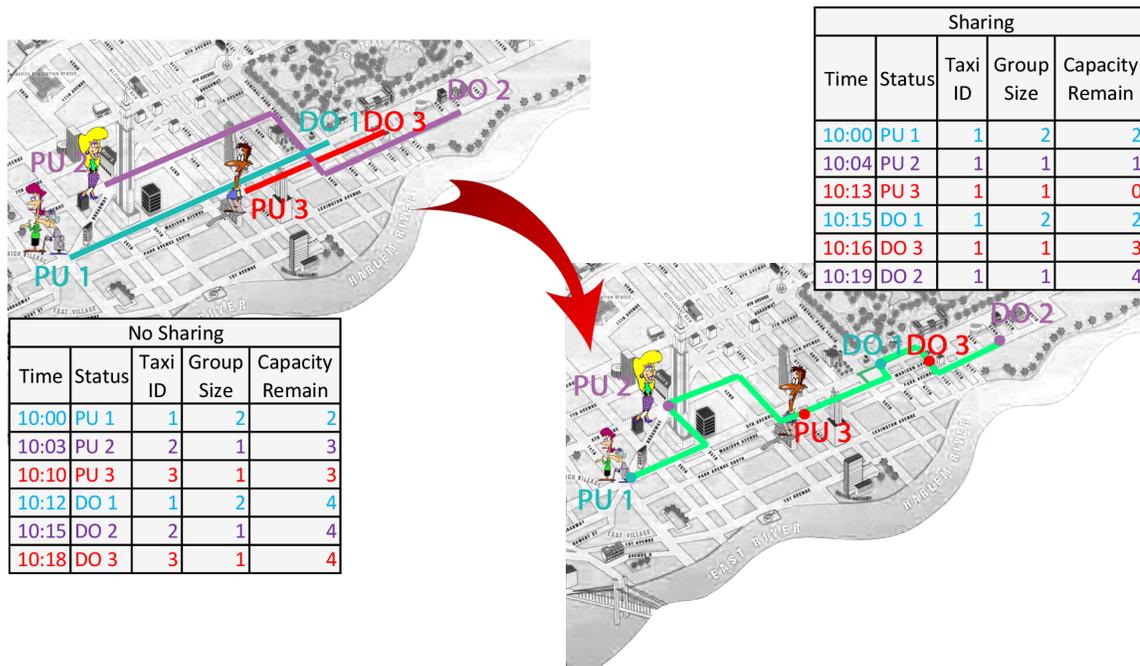


Figure 1.1. Dynamic point-to-point ride sharing - the three rides shown can be combined to a single ride by a car with a capacity of 4 with a small amount of delay in arrival for each rider. (PU- Pick Up, DO - Drop Off)

1.1.1 Ride Matching

In a ride sharing system, it is critical to match the appropriate riders to form the shared ride. Therefore, many researchers focus on developing algorithms for static⁴ ride matching. In particular, Kleiner et al. (2011) proposed an auction mechanism to match rides between two parties and tested its performance using the map of

⁴Static ride sharing assumes knowledge of all the trips ahead of time and proposes the most efficient overall match. Since in actual ride sharing systems, the demand is seldom known in advance, this assumption is a major limitation of static ride matching

Freiburg, Germany with simulated rides randomly sampled from a uniform distribution. Agatz et al. (2011) compared the optimization-based approach with a simple rule-based greedy matching algorithm using travel data from Atlanta, Georgia and concluded that optimization methods have better system performance in matching rides and reducing total system vehicle miles traveled (VMT) compared to greedy matching. However, to simplify the analysis, these studies limited the number of rides that can be shared at a time to be two (i.e., maximally, two passengers can share a vehicle).

More recent research has proposed more flexible models to optimize passenger-vehicle matching and vehicle routing on the fly (dynamic ride matching), considering the vehicle capacity and the number of passengers traveling together (Lin et al., 2012; Santos and Xavier, 2015). Some recent work (Li et al., 2016a) proposed an algorithm to match rides for the purpose of ride sharing as well as parcel delivery. Hosni et al. (2014); Mahmoudi and Zhou (2016); Masoud and Jayakrishnan (2017) have made significant methodological advancements to the matching of shared rides using optimization techniques. While these studies have focused on providing efficient solutions to the dynamic ride sharing problem with given requests and vehicle instances, they have not commented on the city-scale impacts of implementing these systems.

In summary, most of the research on ride matching techniques makes simplifications by either limiting the number of shared rides (Agatz et al., 2011; Kleiner et al., 2011), or testing the models on simulated data and small network sizes (Lin et al., 2012; Santos and Xavier, 2015; Hosni et al., 2014; Mahmoudi and Zhou, 2016; Masoud and Jayakrishnan, 2017) and are hence unable to provide real world inferences. Additionally, since static ride sharing assumes that all routes are known ahead of time, these matching models cannot be used in a dynamic ride sharing setting where riders' routes are known only at the time that they wish to be picked up.

1.1.2 Heterogenous Rider Preference in Ride Sharing

Ride sharing involves multiple riders sharing a vehicle together for potential cost and environmental savings, while potentially enduring a longer ride. Hence it is important to understand how different riders prioritize the cost and environmental savings over a longer ride and having to share their trip with other people. Existing literature that focuses on understanding behavior of riders in carpooling and ride sharing commonly distributes surveys and then analyzes the responses of these surveys. In order to provide a deeper literature review, in addition to ride sharing studies, I also consider carpooling⁵ in this literature review because ride sharing and carpooling share many common factors (such as safety, increased travel distance and time) which could limit their adoption by riders. Most studies on carpooling (Shaheen et al., 2016; Gheorghiu and Delhomme, 2018; Delhomme and Gheorghiu, 2016) and ride sharing behaviors (Zhang and Zhang, 2018; Amirkiaee and Evangelopoulos, 2018; Nielsen et al., 2015; Neoh et al., 2018; Krueger et al., 2016; Alemi et al., 2018) have identified that common demographic factors such as age, education, gender, income level, and logistical factors (e.g, time), nearness to public transit, and cost have an effect on the probability of using these services. Since most of these studies have different metropolitan areas of focus, many of them also identified unique features of the city that either aid or inhibit ride sharing adoption. For example, researchers used surveys in San Francisco Bay Area (Shaheen et al., 2016) and France (Gheorghiu and Delhomme, 2018; Delhomme and Gheorghiu, 2016) to study the reasons for which people used carpooling and found that most carpoolers were motivated by potential time or monetary savings, the convenience of not needing to drive, and the proximity

⁵which is the pre-arranged matching of a driver and passengers when they are going along the same route - for example, carpooling to work

to metro transit. Shaheen et al. (2016) also noted that the Bay Area carpooling system is well established, and riders were in general served by the systems reliability, which could be a factor that influences ride sharing adoption. Krueger et al. (2016) noted that in Australia, the purpose of the trip would be an important factor in the respondents' reason to use AVs but their survey could not identify any trends in the reasons to use AVs with dynamic ride sharing. Neoh et al. (2018); Nazari et al. (2018) examined commuters in the UK and US and found that the presence of school going children, intra-day commutes for work, irregular work schedule, and the general convenience and flexibility of having access to transportation were also factors that determined whether an individual uses ride sharing or not. However, even though these studies indicate that riders may have heterogeneous preferences towards ride sharing, the existing literature (Alonso-Mora et al., 2017; Fagnant and Kockelman, 2016; Ma et al., 2015; Simonetto et al., 2019) do not incorporate these heterogeneous preferences in their models, and instead assign fixed delays and inflexible ride selection rules to riders. These gaps are discussed in detail in Section 1.2.

1.2 Ride Sharing using AVs and non-AVs

To study ride sharing in autonomous vehicles and in non-autonomous vehicles for real world case studies, researchers commonly use agent-based models. ABMs consist of individual entities (agents) that have their own parameters and actions and have been used in the field of transportation for many purposes, including travel time estimation (Chen and Rakha, 2016), disaster relief logistics (Wang et al., 2016), and choice models (Zou et al., 2016) which are used to incorporate individual preferences with regards to transportation modes. To study dynamic ride sharing, researchers

developed agent-based models to simulate such systems (Nourinejad and Roorda, 2016; Fagnant and Kockelman, 2014; Chen et al., 2016a). However, these models are mostly based on simplified system setups, not considering the real-world road infrastructure and the actual travel demands. Nourinejad and Roorda (2016) implemented three cases of ride sharing system on a small test network containing 24 nodes and 76 edges, and Fagnant and Kockelman (2014); Chen et al. (2016a) considered ride sharing in a simplified gridded city.

Some studies use agent-based models to make inferences on ride sharing systems which use AVs. For example, Brownell and Kornhauser (2014); Ma et al. (2015); Fagnant and Kockelman (2016) focused on the economics of ride sharing and showed significant potential monetary savings in New Jersey, Beijing, and Austin, respectively. Ma et al. (2015) also showed that using ride sharing, 2.2 million kg of carbon dioxide can be saved every year in Beijing, while Fagnant and Kockelman (2016) showed that ride sharing has the potential to replace eleven private cars. Martinez et al. (2015); d'Orey and Ferreira (2014) studied shared taxis in Lisbon and Porto, respectively, to infer system level benefits. Both papers (Martinez et al., 2015; d'Orey and Ferreira, 2014) showed that taxi sharing can help passengers reduce travel costs with an increase in total transit time. Santi et al. (2014) introduced the concept of share-ability networks and proposed a mathematical model to quantify the benefits of ride sharing using AVs. They analyzed the taxi trip data in New York City and concluded that ride sharing can reduce cumulative trip length by 40% or more. However, their model also constrained the sharing to be between two riders, ignoring the potential benefits from a more flexible system. Additionally, they assumed that the tolerance level for trip delay is identical for all riders, ignoring the individual heterogeneous tolerance and needs in the real world. Alonso-Mora et al. (2017) extended the methodology used by Santi et al. (2014) by allowing for shares with more than

two riders and showed that the percent of riders served by the system is improved with the increased amount of fixed delay in travel time that is accepted by the riders using small and large capacity autonomous vehicles. Wang et al. (2018) used a simulation model of the city of Singapore to show that SAVs could increase the number of requests served by about 20% - 25%. Yu et al. (2017) used ride sharing trip data to estimate the impact of SAVs in the city of Beijing and inferred that SAVs are able to reduce 46.2 thousand tonnes of CO₂ per year. Mourad et al. (2019) has published a comprehensive review on ride sharing models that have been used for inference.

While ride sharing using AVs (future) is well studied, there are three major gaps. First, many studies in existing literature do not compare these ride sharing systems to present day non-AV systems, nor do they take into account fleets having mixture of AVs and non-AVs. A few studies in ride sharing literature, such as, Brownell and Kornhauser (2014) and Fagnant and Kockelman (2016) performed analysis on taxis and simulated the system for less than three hours, and hence did not need to consider whether the vehicles in the model were autonomous or needed to follow shifting schedules. Others such as Alonso-Mora et al. (2017); Santi et al. (2014); Ma et al. (2015) considered the vehicles serving the passengers as AVs and did not consider non-AVs. Since currently we do not have any level 5 AV in operation, and the full scale implementation of AVs is not expected for a long time, it is important to understand how an increasing adoption of AVs will impact a ride sharing system. Such transitional fleets will have mixtures of AVs and non-AVs operating in the system at the same time.

Additionally, all of the studies in the literature, discussed in this section, that study SAVs using agent-based models discuss the variation of output parameters (for example, reduction in VMT, increase in waiting time, increase in number of riders served) only describe how the outputs will change will varied inputs without.

While understanding the response of the output parameters with respect to the input parameters is important, it is also important to compare those scenarios that offer similar levels of service (number of riders served by the system) to compare the systems on a fair ground. For example, Alonso-Mora et al. (2017) has established relationships between the percent of served requests (output) and the maximum set waiting time for the rider (input). Additionally, they have plotted other output parameters such as the mean travel delay, mean waiting time, and the percent of shared riders against the same input parameter. While this analysis gives us insight into the ride sharing system that they studied, the scenarios are not comparable because some scenarios serve more riders and would cost more. An analysis between scenarios on these plots would not be comparable. For example, scenarios that allow higher allowable delay for the riders would result in a higher average delay (decreased quality of service), while increasing the fleet size or the capacity of the vehicles would be associated with an increased cost to the fleet operators. One way to evaluate these systems would be to find all scenarios that have the same service levels or cost. Fagnant and Kockelman (2014) and Chen et al. (2016a) found comparable scenarios by forcing all riders to get served in their scenarios, but varied a small number of parameters in isolation of each other. While their scenarios were comparable, these papers were unable to quantify the impact of varying these parameters since the parameters in these scenarios are set independently from each other. Boesch et al. (2016) plotted the service level against the AV adoption for different levels of demand and inferred the size of AV fleet required to serve 85% of the demand against AV adoption from the same plot. However, since their study was a simulation, it is difficult to make such inferences without actually running the simulation experiments, because simulations are unlikely to produce deterministic results at each model run. Simulation models that consider varying adoption of RS, and AV along with varying

rider heterogeneity, could potentially have a large number of input parameters. Varying these parameters to find comparable scenarios becomes challenging, especially if each simulation takes a long time to run. Hence, to find comparable scenarios, new methods will be required.

Second, ride sharing has an efficiency limit that has not been investigated in recent literature. In a ride sharing system, the efficiency of the system can be measured by studying the average occupancy of the vehicles during the simulation. Alonso-Mora et al. (2017) predicted that ride sharing can increase the average occupancy of taxis in New York City (NYC) from 1.2 to approximately 3 persons per vehicle during the peak hours of the day. Ma et al. (2015) estimated that 11.4% of the taxi seats were occupied through the day in Beijing if ride sharing is implemented. Even though ride sharing has potential to bring improvements in transportation systems, recent literature shows that ride sharing is not able to reach maximum occupancy of vehicles for an entire day. The benefits of understanding this efficiency limit could guide system planners on designing incentives to improve this efficiency limit.

Third, the heterogeneous preferences of riders towards RS has received very limited consideration in SAV modeling literature. Ma et al. (2015); Alonso-Mora et al. (2017); Fagnant and Kockelman (2016); Wang et al. (2018) did not consider that riders may have heterogeneous strategies in searching for new rides. Ma et al. (2015); Alonso-Mora et al. (2017); Wang et al. (2018); Simonetto et al. (2019) modeled all riders in the system to be able to share rides and every rider in the system could be matched to the earliest available shared or non-shared ride and whether the match met the time and monetary constraints of the rider. In reality, not all riders may search for shared rides in the same way, as discussed in Section 1.1.2. Recently, Kamel et al. (2019) considered these preferences in a simulation study in Paris, for which they randomly generated a population of riders from demographic data, and

assigned travel modes (walk, car, SAV, bike, and public transport) for these riders based on a travel scoring function and found that when sharing preferences were taken into account, the percent of people who chose SAVs reduced by 1.5%. However, their model and inferences were focused only on a single scenario, and hence their inferences were limited to that specific scenario.

Even for those riders who accept shared rides, their tolerance to accepting a longer route may be different. A rider who has a time sensitive ride (for example, going to the airport, or reaching on time to work) may have a very small tolerance to deviating from their route due to ride sharing as compared to a more flexible rider (for example a cost sensitive tourist, or a person travelling to the super market). However, this deviation heterogeneity has also received limited consideration in existing SAV system literature. For example, Alonso-Mora et al. (2017); Ma et al. (2015); Simonetto et al. (2019) considered that each rider would allow a fixed delay in their travel time, Wang et al. (2018) considered several fixed constraints to limit the maximum waiting time, maximum arrival and departure delay, and the minimum taxi fare reduced in their model, while Fagnant and Kockelman (2016) considered that riders could encounter a delay which was within a fixed percentage of their travel distance. However, as discussed in Section 1.1.2, literature has identified that common demographic factors like age, education, gender, income level, and logistical factors such as time, nearness to public transit, and cost have an effect on the probability of riders using ride sharing services. This suggests that all passengers may not have the same fixed parameters, which are being used in existing literature. Additional research is needed to incorporate these heterogeneous preferences in simulation models that are used to understand the system performance parameters to provide better understanding of ride sharing systems and help design appropriate policies.

In summary, the main gaps in the research on ride sharing systems using AVs and non-AVs are: 1) most existing studies do not compare the key performance parameters of ride sharing systems across scenarios is not done on a fair ground (comparatively), 2) The efficiency limitations of ride sharing are not well studied, 3) heterogeneous sharing preferences are not considered for the riders in a SAV system.

1.3 Electric Vehicles and Integration with SAVs

While there is a lot of literature on EVs and EV systems (Richardson, 2013; Yilmaz and Krein, 2012; Ko et al., 2017), the major limitation the integration of EVs in an SAV system, is the long charging time of EVs and limited availability of charging infrastructure. Hence, the literature in this section is reviewed from two main perspectives, 1) siting charging stations for charging infrastructure development and, 2) understanding how the adoption of EVs impact the performance of SAV systems.

It is important to find optimal sites for EV charging stations because EVs need a long time to charge as compared to internal combustion engine vehicles (ICV) hence some studies develop methods to do so. Researchers have worked on understanding EV charging behavior and developed a methodology to optimally locate EV charging stations. For example, Cai et al. (2014a); Shahraki et al. (2015); Yang et al. (2017a) used the GPS trajectory data to select sites for optimal charging locations. While Cai et al. (2014a) used the trip data to identify parking events and site the EV charging stations accordingly, Shahraki et al. (2015) used an optimization model to resite the 40 charging stations in Beijing thus improving the electrification rate by 37%. He et al. (2018); Arslan and Karaşan (2016); He et al. (2015); Davidov and Pantoš (2017) focused on developing the better algorithms for the optimization of

EV charging station locations. The literature on charging station siting is discussed in detail in Chapter 6.

Some literature (Chen et al., 2016a; Chen and Kockelman, 2016b; Bauer et al., 2018) has studied the impact of EVs in SAV systems. Chen et al. (2016a) used a simplistic agent-based model to infer that a 200-mile battery ranged shared Autonomous Electric Vehicle (SAEV) could replace approximately 5.5 privately owned vehicles. However, the model was tested on a simulated network using simulated data and cannot draw significant inferences on system performance parameters. Chen and Kockelman (2016b) suggested that if SAEVs are priced between \$0.75 to \$1 per mile, a 80 - mile ranged electric vehicle can generate significant revenues for the operator and the fleet owner. Bauer et al. (2018) estimated that for Manhattan, NY the costs would be lowest for EVs with battery ranges between 50-90miles and 66 chargers per square mile of 11kW each, and an SAEV fleet could reduce GHG emissions by 73%. However, the previous studies (Chen et al., 2016a; Chen and Kockelman, 2016b; Bauer et al., 2018) do not compare the differences in siting charging stations for a non autonomous vehicle - ride sharing (AV-RS) system with that for a AV-RS system and study the system performance parameters as the EV adoption increases. Since AVs, RS and EVs are each emerging technologies that are growing at variable rates, it is important to understand the synergies between these systems to inform policy makers and city planners on best practices regarding the growth of these technologies.

Hence to summarize, there is a lack of research in the field that considers ride sharing, autonomous vehicles, and electric vehicles together and compares the results against the existing scenario (no ride sharing, gasoline vehicles and traditionally driven taxis). Research is also needed in siting EV charging stations in a SAEV scenario considering the uptake of these technologies.

1.4 Research Gaps

In summary, while existing literature makes great advancement in understanding ride sharing using optimization models and agent-based models, there are several gaps that this dissertation aims to address:

1. Existing literature does not study fleets that contain a mixture of EV and gasoline vehicles, AVs and non-AVs (non autonomous vehicles or traditional vehicles) in a system where each rider has their own preference of sharing a ride or not. Such future transportation systems have a large number of parameter settings and current literature often fix some of the parameters to extreme values (for example, 100% AV or EV adoption).
2. Literature on ride sharing commonly use simulations to find trends between the input parameters and the output parameters, which help make inferences on how changing a particular input (say ride sharing adoption) could impact the system performance (for example number of riders served and waiting time). However, these studies do not evaluate scenarios on the same ground, and hence limit the understanding of system outputs with respect changing multiple input parameters at the same time. Methods are needed to find comparable scenarios in future transportation systems.
3. Existing research has not explored reasons behind ride sharing being unable to fully occupy all the seats in the vehicle. While existing literature discuss the benefits of ride sharing (that include a reduction of fleet size with moderate increase in ride time, environmental benefits, potential monetary savings), each of these studies have also noted that, even with ride sharing, most taxis are unable to reach their maximum capacity. By studying the reasons for these effi-

ciency limitations, we can better plan ride sharing systems by either deploying ride sharing in areas where there would be higher participation, or assigning high capacity vehicles in places where capacity is a limiting factor.

4. EV charging station siting literature also has multiple gaps (siting charging stations for multiple levels of EV adoption, and the consideration of queuing of EVs at charging stations, and inflexible charging rules for EVs) which are discussed in detail in Chapter 6. The existing SAEV system studies do not consider potential differences in siting charging stations, and system performance parameters for present day non AV-RS systems and future AV-RS systems for increasing EV adoption.

1.5 Research Structure

To address the gaps discussed in Section 1.4, this dissertation (pictorially represented in Figure 1.2) develops a Parameterized and Preference-based Shared Autonomous Electric Vehicle (PP-SAEV) model to simulate ride sharing in systems which have a mix of AV and non-AVs, EVs and ICVs, and riders can have heterogeneous sharing preferences. The model is parameterized so that several simulation scenarios can be run with varying AV, EV, and RS adoption, fleet sizes, taxi capacities, and heterogeneous preferences. By running the PP-SAEV model with different parameter settings, the dissertation aims to answer the below research questions. In each study the performance of the systems are evaluated based on service level, waiting time, riding time, percent of shared rides, average occupancy, and GHG emissions.

1. How does increasing ride sharing adoption affect systems where the vehicles used are AVs and non-AVs?
2. How can we size taxi fleets and evaluate system performance comparatively (providing the same service level) in scenarios where taxi systems may have different levels of AV and RS adoption?
3. What are the reasons that limit the ability of ride sharing to increase the efficiency of a taxi system to 100%?
4. How would rider sharing preferences change the quality of service in a SAEV system?
5. In a shared electric autonomous vehicle scenario, how do we better site electric charging stations and what would be the optimal capacity of those charging stations under continually increasing demand?
6. How does increasing EV adoption impact autonomous vehicle - ride sharing (AV-RS) systems and non AV-RS systems in service level, and how would this growing EV taxi fleet impact the electric power grid.

The rest of this dissertation is structured as follows. Chapter 2 presents the PP-SAEV model and describes the case study which is used to to illustrate this PP-SAEV model. Then, I analyze the performance of ride sharing in an autonomous vehicle system and a traditional (non AV) system by finding comparable scenarios in Chapter 3, to answer question 1. Chapter 4 introduces a meta-model based simulation optimization method to find comparable scenarios to analyze fleets that can have a mixture of AVs and non-AVs, with varying capacities, and varying sharing participation tolerances to answer question 2. Chapter 4 also analyzes the efficiency limit

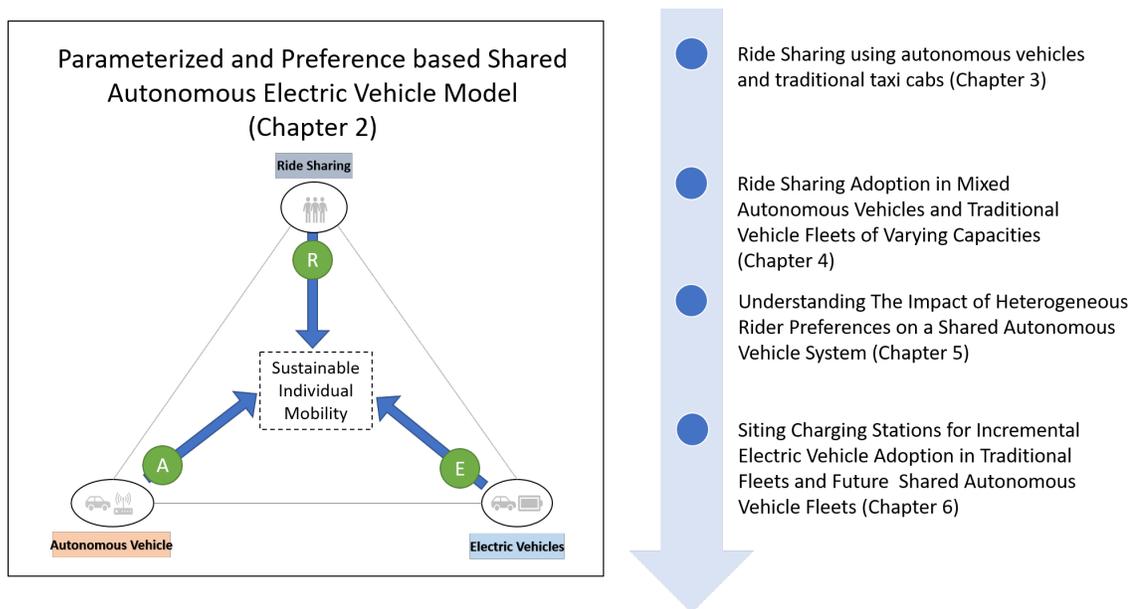


Figure 1.2. Research overview

of ride sharing using compositional data analysis and suggests policies to overcome this sharing limitation to answer question 3. Chapter 5 studies the effect of heterogeneous rider searching preferences on system performance using mixture experiments and hence addresses question 4. Chapter 6 builds a model to site charging stations for a growing EV fleet to address question 5 and analyzes systems with fleets that have traditional technologies (non AV-RS) and fleets that have future technologies (AV-RS) to address and addresses question 6. Finally, in Chapter 7, I summarize the findings and contributions of this dissertation and discuss directions for future research.

2. PARAMETERIZED PREFERENCE-BASED SHARED AUTONOMOUS ELECTRIC VEHICLE (PP-SAEV) MODEL

To study the synergies between ride sharing and other technological advancements in vehicles, namely autonomous vehicles (AV) and electric vehicles (EV), this dissertation uses agent-based models. Agent-based modeling allows the modeler to represent a dynamic ride sharing system with heterogeneous rider and taxi preferences, and also to look into finer performance parameters of the ride sharing system. The proposed Parameterized Preference-based Shared Autonomous Electric Vehicles (PP-SAEV) model simulates ride sharing in systems which have a mix of AV and non-AVs, EVs and ICVs, and riders can have heterogeneous sharing preferences. The model is parameterized so that several simulation scenarios can be run with varying AV, EV, and RS adoption, fleet sizes, taxi capacities, and heterogeneous preferences. Taxi data from the New York Taxi and Limousine Commission (NYC DOT, 2016b) from the year 2014 is then used to build a case study to answer the questions in Section 1.5.

The agent-based simulation model (PP-SAEV) described in this chapter accommodates the heterogeneous behaviors of the riders, the taxis, and the charging stations by modeling taxis and riders as individual entities (agents) that each have their own parameters and preferences. The model generates data for each agent at several time-steps through the simulation, using which it is possible to understand spatial

and temporal variability of performance indicators of the system. The system as well as each agent (riders, taxis and charging stations) in the model have their own parameters which can be adjusted to construct different scenarios.

2.1 System Level Parameters

The system level parameters that can be set in the model are listed below:

- *fleetSize* : The number of taxis that are present in the system
- *riderTypeDistribution* : A discrete distribution that can take values 1-5 (defined in Section 2.2 and detailed in Algorithm 2) from which the *Type* of each rider will be randomly drawn.
- *EVPercent* : The percent of taxis that are electric vehicles
- *AVPercent* : The percent of taxis that are Autonomous
- *taxiCapacityDistribution* : A discrete distribution that can take values from 1 to 10 from which the capacity of the taxi (*taxiCapacity*) is drawn.
- *b* : The constants of the charging utility functions as mentioned in Section A.3.5. The charging utility function determines whether a taxi moves towards a charging station or not. The utility function considers the current charge level and the taxis closeness to a charging station.
- *outShiftTime* : the off-shift time after a taxi ends a shift set to 3 hours
- *dtMode*¹ : The mode of the distribution of the deviation tolerance *deviationTolerance* of the riders. In order to model the heterogeneity of rider-groups, each of them

¹Since we do not have prior information of the additional distance a rider is willing to travel in order to enable ride sharing, we randomly set these *deviationTolerance* for each rider

have a unique distance deviation that they allow so as to enable them to share rides.

Table 2.1.

Sample data from NYC-TLC with pick up (PU) and drop off (DO) time and locations in latitudes (Lat) and longitudes (Long), and the number of passengers traveled together (Group Size)

PU Time	DO Time	PU Lat	PU Long	DO Lat	DO Long	Group Size
06/12/2014 0:00:00	06/12/2014 0:13:24	-73.937	40.797	-73.964	40.760	1
06/12/2014 0:00:00	06/12/2014 0:05:00	-73.993	40.719	-74.001	40.717	3
06/12/2014 0:00:00	06/12/2014 0:04:00	-73.983	40.771	-73.984	40.765	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

I run the PP-SAEV for different scenarios by varying these system level parameters. The rider parameters are taken from a database similar to that described by Table 2.1. The individual agents themselves (taxis, riders, and charging stations) also have parameters that are set according to these system level parameters. The

from a triangular distribution with minimum value = 0, maximum value = 1, and mode $dtMode$. The $deviationTolerance$ of the rider-group is calculated as $deviationTolerance = \frac{\text{Acceptable maximum trip distance after sharing} - \text{Original Trip Distance}}{\text{Original Trip Distance}}$. We have defined the distance that a rider group allows for deviating from their original path to be proportional to the unshared trip distance, because it is unreasonable to expect a rider group whose trip is short (e.g., 0.5 miles) to accept a large change in route (e.g., 2 miles).

rider parameters are their sharing preference (defined in Section 2.2), their pick up and drop off location, and time. The taxis parameters are their capacity, and whether or not they are autonomous vehicles or electric vehicles. The charging stations parameters are their location, and number of charging ports (capacity). These are set from an external database.

2.2 Agent Behaviors

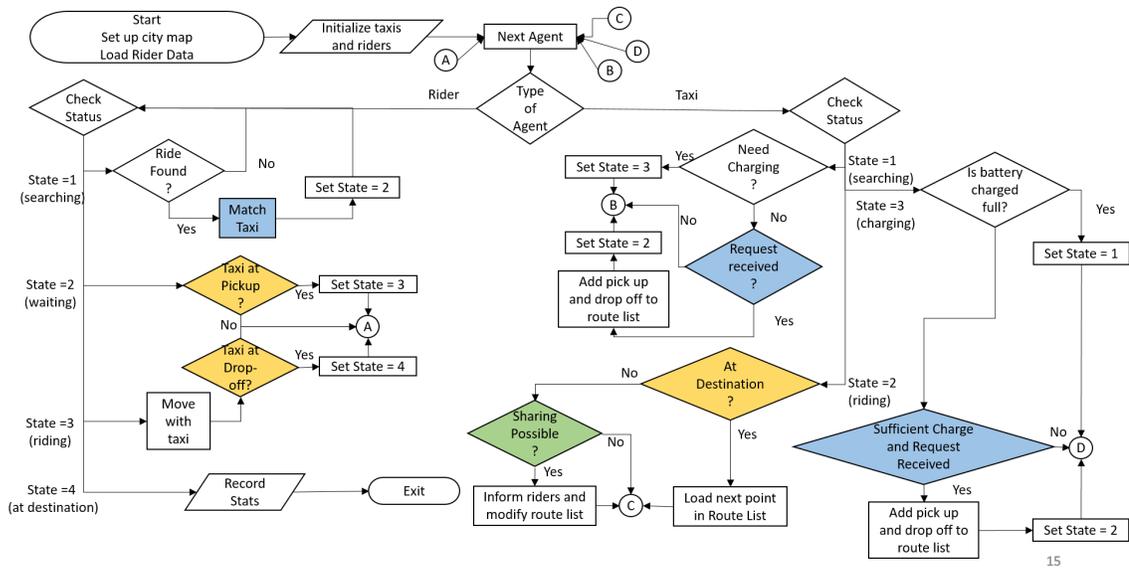


Figure 2.1. Model overview. Matching colors in the flowchart represent an event / decision taking place simultaneously by the rider and the taxi. The circular blocks labeled A, B, C, D are used as "connectors" to imply a transfer of flow to their corresponding connector

The actions of the agents (riders and taxis) are described simplistically in Algorithm 1-7. A simplified overview of the model can be seen in Figure 2.1. A more

detailed pseudo-code that rigorously describes these actions is presented in Appendix A.

The execution of the algorithm starts with the initialization. The map of the city-of-interest is loaded into the model. The taxis are then loaded into the model all at once. Each taxi at this time decides whether or not it is an EV, AV, or both, based on the system level predetermined EV and AV percentages (*EVPercent* and *AVPercent*). The rider-groups enter into the model as per their pick up time and at the pick up location. Each rider-group decides their sharing strategy while entering the model. Each agent then executes their own actions at every time step. The actions that the rider-groups perform are detailed in Algorithm 1. The rider-groups search for rides as per their searching preference which is detailed in Algorithm 2. If they receive a response from a taxi, they wait until the taxi arrives to the pick up location, then ride with the taxi to the drop off location, and then exit the system. The search preferences of the riders as detailed in Algorithm 2 divides the riders into five rider-group types as detailed below:

- *riderType* = 1 (Non-sharing only): This rider type represents a class of riders who do not want to share at all. The reasons that riders may not want to share could be either that they are in a hurry and cannot accept deviation from route or they may not want to get into a taxi with another passenger due to social reasons. If a rider of type 1 does not find a shared ride in *exitTime* seconds, it exits the system unserved.
- *riderType* = 2 (non-sharing preferred) : This rider will first search for non shared rides until time *regLimit* and if a non-shared ride is not found they will seek a shared ride until *exitTime*. We use this rider type to represent the class

of riders who may not want to share due to social reasons, but due to lack of availability of a regular ride are forced to use a shared ride.

- $riderType = 3$ (indifferent) : These types of riders are indifferent to sharing, and weigh the advantages (potentially lower cost, better for the environment) and disadvantages (potentially longer travel distance, safety concerns) of sharing equally. These riders stay in the system searching for shared and non shared rides until $exitTime$
- $riderType = 4$ (sharing preferred) : These riders first search for a shared ride until $shareLimit$ before searching for non shared rides until $exitTime$. These represent cost sensitive travellers who would prefer to share a ride for potentially lower costs. If however, shared rides are unavailable, they will seek a non-shared ride.
- $riderType = 5$ (sharing only) : These riders are extremely cost sensitive and would rather not travel by taxi if sharing was not available. We can assume that these riders would use another means of reaching their destination (for example, public transit) if a shared ride was not available in $exitTime$ seconds.

The taxis execute their actions as per Algorithm 3. If the taxis are traditional taxi cabs (not autonomous), they will shift in and out of the model as per the rules defined in Algorithm 5 based on their shift schedules. To model the shift changes for non-autonomous taxis, at the initialization of the model, we determine the number of taxis that are in-shift (n_s) based on the availability ratio at the model start time (normally midnight) and day (weekday or weekend), according to Figure 2.5. From all taxis, n_s of them are then randomly selected to be in-shift while the rest to be

out-shift. Then, the Algorithm 5 is run every 15 minutes to make shift changes. Only taxis that are in-shift are allowed to participate in serving rider groups.

If the taxis are electric taxis they will follow rules as per Algorithm 6. The electric taxis go towards the charger if one of the two conditions are met: 1) The battery energy level is below the minimum operating level as defined by *minSOC* or 2) the charging utility is positive. The charging utility is a score that considers battery remaining and closeness to a charger (Section A.1). Intuitively, a taxi that is currently searching for a rider, and is close to a charging station and with moderate battery energy remaining (say 50%) may want to charge its battery while not serving any riders. This would help the taxi remain in service (have enough battery energy to serve customers) during periods of high demand.

When the taxis initially enter into the model, they search for rider-groups. The taxis are matched to rider-groups by using Algorithm 4. Once a match is found, the taxi adds the rider-group's pick up and drop off point to its *routeList* (a list of points that a taxi visits) and proceeds to serving the rider-group.

On its route, the taxi will try to search for sharing rider-groups using Algorithm 4. There are cases when a taxi will stop searching for shares, and the model tracks these reasons that a share would be stopped using Algorithm 7. If a valid share is found, the new pick up and drop off point is added to the route list, and then the taxi follows this modified route list. Once the taxi has visited all the points on the *routeList*, it is designated as idle and starts searching for its first ride again. The relations between Algorithms 1-7 are illustrated in Figure 2.2.

The ride matching as detailed in Algorithm 4 is a two-step process. The model uses a *preCheck* algorithm (Details in Section A.3.2) to filter out rides that would be unlikely sharing candidates. Those rides that pass through this *preCheck* are

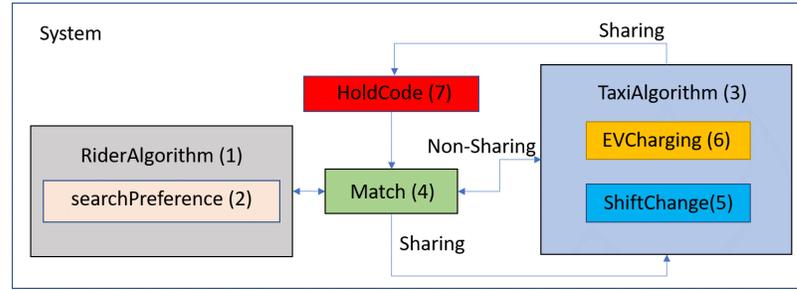


Figure 2.2. The relation between Algorithms 1-7 in the agent-based model

evaluated using the *bestRoute* algorithm (Details in Section A.3.3) where the ride that minimizes inconvenience to passengers is selected.

2.3 Model Outputs

The outputs of the model include:

- The number of rider-groups leaving the system without being served. A rider-group is left unserved if the total time in system for the rider group exceeds their acceptable waiting time.
- Time stamps for each rider at every status change, the information of which we can use to calculate the following: (Figure 2.3)
 - Time taken for a taxi to respond to each rider group’s request and commit a pickup, TR_R
 - Time that a rider group waited for a taxi, TR_w
 - Time the rider group spent riding, TR_{Ride}

- Time between the arrival of the rider in the system till its departure from the system after reaching its destination, TR_{SYS}
- The number of rider-groups and total passengers in a taxi at any time
- The distance traveled by each taxi with and without passengers
- The reasons each taxi stopped searching for shares, which are:
 1. The sharing tolerance limit was reached
 2. The taxi was at its maximum capacity
 3. The taxi needed to shift out
 4. The taxi needed charging
- The state of charge of each electric taxi at all times
- Charging statistics which include the charging station utilization, times when each taxi started charging and stopped charging

Using the information from these outputs, we can infer:

- The number of vehicles that can be reduced in the system under different scenarios compared to the base scenario (traditional gasoline vehicle with shifts and no sharing)
- The degradation in quality of service for the riders in terms of the additional distance traveled, additional time taken to reach destinations, as well as the increase in waiting time for the riders.
- The reduction of distance traveled by the taxis due to ride sharing and consequently the reduction of greenhouse gas emissions from the system. I use

4.17×10^{-4} metric tons CO₂-eq/VMT estimated by EPA (2017) to convert the changes in VMT under different scenarios to emission reductions.

- The distributions of the number of taxis that reached the sharing limit for different reasons, which would help us infer the efficiency limitations of ride sharing for each scenario.
- The charging station utilization which would help to plan for charging station infrastructure for the city. This would also give me the variation of the amount of energy used for charging stations with time of the day (load profiles). This would help us determine the amount of energy used by the charging stations.

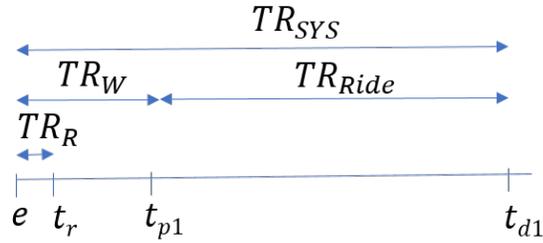


Figure 2.3. Time stamps recorded for each rider group in the system. e is the time that the rider group enters in the system; t_r is the time that the taxi responds to the rider group to commit a pickup; t_{p1} is the time that the rider group is picked up; and t_{d1} is the time that the rider group is dropped off at its destination.

2.4 Model Assumptions

The key assumptions of the PP-SAEV are :

1. All rider-group in NYC can be classified into the five rider types as defined in Section 2.2.
2. A rider group is only eligible to share a ride if the rider group allow a distance overage of at-least 100m. Similarly, the maximum distance a rider can deviate has been capped at 10,000m. These numbers were tested for sensitivity and I found that the model output parameters did not change significantly by allowing less than 100m or more than 10,000m of sharing.
3. The refueling time for ICEVs is considered negligible.
4. Taxis can be a mix of EV Traditional, EV Autonomous, Gasoline Autonomous, and Gasoline Traditional.

Variables used in Algorithms 1-7 are defined below:

- Rider Group (Algorithms 1, 2, and 4)
 - *tis* : The time that the taxi has spent in the system at the current evaluation
 - *riderGroup* : set of riders currently in the model
 - *regLimit* : Time until which the rider group of *riderType* = 1, 2, 4 will search for non-shared rides. The *regLimit* was set to 5 minutes, as we assumed that this was a reasonable amount of time that riders would search for taxis².
 - *shareLimit* : Time until which the rider group of *riderType* = 2, 4, 5 will search for shared rides. In our model, we assumed that riders *riderType* =

²We tested the value of 5 minutes for sensitivity and found no significant difference in the model outputs for values of *regLimit* more than 5 minutes

2, 4 that shared a ride would spend a time proportional to their deviation tolerance in searching for shared rides, and $riderType = 5$ would spend 10 minutes searching for shared rides, since we assume that riders who search for only shared rides would allow for additional time to search for an appropriate ride share.

- *exitTime* : Time by which a rider group exits the system unserved if no ride is found. For $riderType = 1, 2, 4, 5$ $exitTime = shareLimit + regLimit$. For $riderType = 3$, $exitTime = (1 + deviationTolerance) \times 5$ (same as the other flexible rider types 2 and 4). The settings for the *regLimit*, *shareLimit* and *exitTime* parameters are summarized in Table 2.2.
 - *State* : the state of the rider. Can be set to 1) Searching 2) Waiting for pick up 3) Riding with taxi 4) Reached destination
 - *riderType* : the rider sharing preference as defined earlier in this section
 - *deviationTolerance* : percent by which rider will tolerate deviation from path
- Taxi (Algorithms 3, 4, 5, 6, and 7)
 - *Taxis* : Collection of taxis in the system
 - *State* : the state of the taxi. It can be set to 0) Charging 1) Searching for first ride 2) Riding and searching for shares
 - *searchSOC* : The *SOC* at which the taxi is allowed to start searching for passengers when at *State* = 0
 - *SOC* : State of charge. $SOC = \frac{\text{the distance that the taxi can drive untill next charge}}{\text{The distance the taxi can travel in a single full charge}}$

- *requestList* : the list of riders that have requested a ride
- *holdCode*[.] : A set of 4 boolean flags that indicate the reason for which the taxi stops sharing. If any flag is set to *TRUE* the taxi stops searching for new shares. The flags are set according to rules in Algorithm 7
- *bestMatch* : the rider that was the best match for the taxi as per algorithm 4

Table 2.2.
Settings for the *regLimit*, *shareLimit* and *exitTime* parameters

<i>riderType</i>	<i>regLimit</i>	<i>shareLimit</i>	<i>exitTime</i>
1	5	0	5
2	5	$deviationTolerance \times 5$	$(1 + deviationTolerance) \times 5$
3	NA	NA	$(1 + deviationTolerance) \times 5$
4	5	$deviationTolerance \times 5$	$(1 + deviationTolerance) \times 5$
5	0	10	10

Algorithm 1 Rider group algorithm

```

1: procedure RIDERALGORITHM(riderGroup)
2:   for all  $r \in \textit{riderGroup}$  do
3:     if  $State = 1$  then ▷ searching for taxis
4:       Execute Algorithm 2
5:     if  $State = 2$  then ▷ Waiting for Pick Up
6:       if taxi is at pick up then
7:          $State \leftarrow 3$ ;
8:       if  $State = 3$  then ▷ Riding with taxi
9:         if taxi is at drop off location then
10:           $State \leftarrow 4$ ;
11:        else
12:          Move with taxi
13:          if Algorithm 4 finds a share then
14:            Update deviation tolerance
15:          if  $State = 4$  then ▷ Exit the system (either served or unserved)
16:            Record stats and exit the system

```

Algorithm 2 Rider search preference

```

1: procedure SEARCHPREFERENCE(riderGroup  $r_c$ )
2:   if  $riderType = 1$  or  $riderType = 3$  or ( $riderType = 2$  and  $tis < regLimit$ )
   or ( $riderType = 4$  and  $tis \geq shareLimit$ ) then
3:     Broadcast non-sharing requests
4:   if  $riderType = 5$  or  $riderType = 3$  or ( $riderType = 2$  and  $tis \geq regLimit$ )
   or ( $riderType = 4$  and  $tis < shareLimit$ ) then
5:     Broadcast sharing requests
6:   else
7:     Set  $State = 4$  ▷ Exit system unserved

```

Algorithm 3 Taxi algorithm

```

1: procedure TAXIALGORITHM(Taxis)
2:   for all  $x_c \in Taxis$  do
3:     if State = 0 then ▷ Charging
4:       Check battery level and call Algorithm 6
5:     if State = 1 then ▷ Searching
6:       if battery needs charging then ▷ Algorithm 6
7:          $State \leftarrow 0$ 
8:       else
9:         receive requests from nearby non-sharing riders and add request
           to requestList
10:        Call Algorithm 4
11:       if a match is found then
12:         if taxi can serve rider with available battery energy then
13:           add pick-up and drop-off point to route list
14:           Proceed to the first point on the route list
15:           Set  $State \leftarrow 2$ 
16:       if State = 2 then ▷ in service
17:         if reached destination then Load / Unload rider group as per type of
           destination (pick up / drop off)
18:         if there are more points to visit in route list then
19:           Move to next point in route list
20:         else
21:           Set  $State \leftarrow 1$ 
22:       else
23:         if Taxi can share ( $holdCode = \{F, F, F, F\}$ ) then ▷ Algorithm 7
24:           search for share requests and add request to requestList
25:         Call Algorithm 4

```

Algorithm 4 Matching algorithm

```

1: procedure MATCH(Taxi  $x_c$ , List requestList)
2:   remove all  $req \in requestList$  where  $req.State \neq 1$   $\triangleright$  Each rider group sends
   requests to multiple taxis and could have got matched before  $x_c$  evaluates their
   request
3:    $bestMatch \leftarrow NULL$ 
4:   for all  $req$  in requestList do
5:     if  $x_c.State = 2$  and  $holdCode = 0$  then  $\triangleright$  Evaluate Algorithm 7 to set
     holdCode  $\triangleright$  Evaluating a sharing request
6:       Use preCheck to eliminate infeasible shares  $\triangleright$  (Section A.3.2)
7:       if  $req$  passes preCheck then
8:         run bestRoute algorithm to insert  $req$  into the route list of  $x_c$  in
         the most optimal way while validating deviation tolerances for all riders involved
         in the share  $\triangleright$  (Section A.3.3)
9:         if bestRoute algorithm does not return a valid route then
10:           match is not found
11:         else
12:           match is found
13:           if  $req$  is better than  $bestMatch$  then  $\triangleright$  A better match is
           defined as a one that has a better score based on Equation A.1
14:             Set  $bestMatch \leftarrow req$ 
15:         else  $\triangleright$  Evaluating a non-sharing request
16:           if rider  $req$  is the closest rider to taxi  $x_c$  in requestList then  $\triangleright$  Taxi
           prefers the closest rider group
17:             Add  $req$  pick-up and drop-off point to routeList
18:             Add  $req$  to list of possible rides

```

Algorithm 5 Shift change

```

1: procedure SHIFTCHANGE(Taxis)
2:   Randomly select  $n_i$  taxis to be off-shift and the rest of the taxis to begin
   shifts
3:   for all  $x_c \in Taxis$  that are in-shift do
4:     if  $x_c$  is in-shift then           ▷ in-shift taxis are searching for passengers
5:       if time of in-shift > 8 hours then
6:         if State = 2 then               ▷ still delivering riders
7:           Stop accepting new shares           ▷ holdCode = 4
8:           Set to be off-shift after the last drop off
9:         else
10:          Set to be off-shift                 ▷ Not searching for new rides

```

Algorithm 6 EV charging

```

1: procedure EVCHARGING(Taxi  $x_c$ )
2:   if  $x_c.State = 1$  then                                     ▷ Searching for rides
3:     if  $x_c.SOC < minSOC$  then                               ▷  $minSOC$  is the minimum  $SOC$  with
       which a taxi can continue to search for new passengers
4:       Set  $State \leftarrow 0$ 
5:     else
6:       if there is a charging station close by and the battery needs charge
       then
7:         search for charger and move towards charging station
8:       if taxi is at charging station then
9:         update  $x_c.SOC$  as per time spent at charging station
10:      if  $x_c.SOC = 1$  then
11:        release the charger
12:   if  $x_c.State = 0$  then
13:     if  $x_c.SOC$  is greater than  $searchSOC$  then
14:        $State \leftarrow 1$ 

```

Algorithm 7 Detect sharing stoppages

```

1: procedure HOLDCODE(Taxi  $x_c$ )
2:    $holdCode \leftarrow \{F, F, F, F\}$  ▷ F: False, T:True
3:   if  $x_c.Status = 2$  then ▷ Searching for shared rides
4:     if any riders in curRiders cannot share then
5:        $holdCode[1] \leftarrow T$  ▷ rider tolerance reached
6:     if taxi is at max capacity then
7:        $holdCode[2] \leftarrow T$ 
8:     if taxi has low battery level then
9:        $holdCode[3] \leftarrow T$ 
10:    if taxi needs to get off shift then
11:       $holdCode[4] \leftarrow T$ 

```

2.5 Data and Exploratory Analysis

The PP-SAEV model is applied to a case study of NYC taxis to quantify the impacts of ride sharing at the city-scale. Although my analysis is based on one city, it is notable that the model and framework is applicable to any city if similar data is available. We choose NYC as our case study due to the following reasons:

- **Data availability:** NYC DOT (2016b) has published an extensive and highly detailed database which allows us to validate the agent-based model at a micro level.
- **Potential impacts:** The great demands of taxi rides in NYC indicate potential significant saving opportunities. The average number of daily trips by taxis in NYC is 485,000 (NYC DOT, 2014). During the evening peak hour, on average, there are over 8,000 pickups within every 15 minutes (Figure 2.4(a)).
- **Spatial sharability:** Trips in NYC are highly concentrated (e.g., over 90% of the taxi pickups are in the Manhattan region)(Figure 2.4(b)). The high number of taxi rides along with the high spatial and temporal concentration of rides make ride sharing a great transportation alternatives.
- **High ratio of single-rider trips:** Over 65% of all trips in NYC are single-person trips (Figure 2.4(c)), which leaves a large amount of unused capacity in the vehicles. This unused capacity can be filled by shared trips.

A sample of the data used in this study is shown in Table 2.1. The data we used are the green and yellow taxi trip data from the year 2014 (NYC DOT, 2016b). We chose data from 2014 because, at that time, ride sharing applications such as Uber and Lyft had not been widely adopted to impact taxi ride demands (NYC DOT,

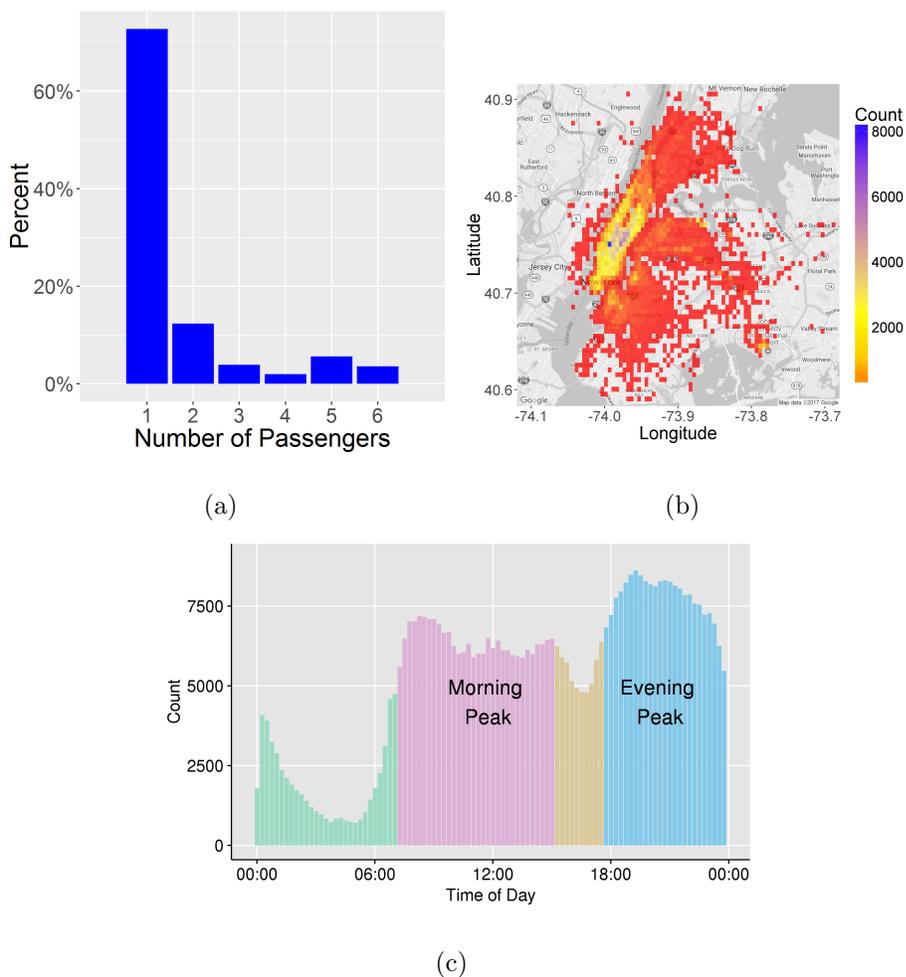


Figure 2.4. NYC taxi demand represented as pickups. (a) is a temporal histogram for the pick ups every 15 minutes, (b) shows the spatial density of taxi pick ups in NYC. (Grid resolution is $0.005^\circ \times 0.005^\circ$, roughly equivalent to $0.5km \times 0.5km$). (c) Histogram for the number of passengers in a rider group. (a) and (b) plots use the data from the date 8/24/2014. (c) uses data from the year of 2014.

2016a). Hence the trip data from 2014 is more representative of the total demands

in the city than the most recent data. The NYC-TLC records each pick up and drop off for all yellow and green taxis registered in the city. The green taxis are not allowed to pick up passengers below West 110th Street and East 96th Street, or at the two NYC airports (NYC DOT, 2014), while the yellow taxis do not have such restrictions. The reason for having this distinction between the Green and Yellow taxis was to have more taxis available in the suburban region of the city (NYC DOT, 2014). The data recorded by the NYC-TLC is the trip pick up time and location (in longitude and latitude), drop off time and location (in longitude and latitude), and group size (number of people traveling together).

The travel demands in NYC have been divided into four phases as shown in Figure 2.4(a) based on pick up time. While data is available for all phases and the models are run for the entire day, and for multiple days, the discussions are mainly focussed on the peak demand periods (the evening peak from 5:01pm to midnight), because peak demand periods can benefit more from system efficiency gain through ride sharing.

2.6 Modeling Taxi Shifts (for traditional taxi)

Unlike the current taxis that are temporarily unavailable during the off-shift periods, autonomous vehicles can operate 24×7 . Figure 2.5 shows the service valleys of NYC taxis. These valleys can be eliminated by autonomous taxis. In order to model traditional taxis, we modeled the change of shifts for the traditional taxis to have shift schedules that are similar to the existing NYC taxi operation schedules (Figure 2.5). The shifting schedules are modeled in Algorithm 5

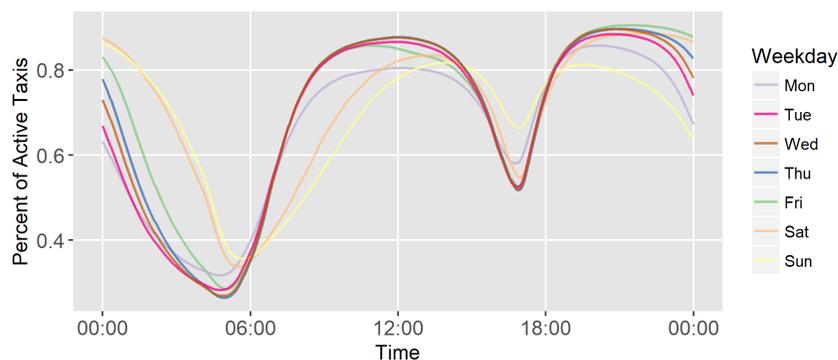


Figure 2.5. Average number of taxis in operation (on shift) every minute (NYC DOT, 2014)

2.7 Summary

An agent-based model was built to study systems with varying adoptions of RS, EV, AV wherein riders can have heterogeneous searching preferences. By varying the parameters of this PP-SAEV model, new scenarios of the system can be generated and output parameters can be studied. The demand to the PP-SAEV is the NYC TLC taxi data which identifies the pick-up time and location and the drop-off locations of the riders. Subsequent chapters will describe different studies that I have performed by varying the parameters of the PP-SAEV. By studying the output from different parameter variations of the PP-SAEV model, I am able to provide insights for policy makers and system designers about SAEV systems.

3. RIDE SHARING USING SHARED AUTONOMOUS VEHICLES AND TRADITIONAL TAXI CABS

The results in this chapter have been published in a journal paper: Lokhandwala, M., & Cai, H. (2018). Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC. *Transportation Research Part C: Emerging Technologies*, 97, 45-60.

3.1 Introduction

Ride sharing and autonomous vehicles as discussed in Chapter 1 are emerging mobility systems. While the implementation of ride sharing has already begun (for example UberPool and Lyft Line), fully autonomous vehicles are not expected for the next 10 years (McKinsey & Company, 2018). Hence, it is important to understand how an increasing adoption of ride sharing impacts both present day non-AV fleets and future AV fleets ride sharing using autonomous vehicles and using traditional taxi cabs in order to understand the changes that autonomous vehicles will bring. This chapter uses the PP-SAEV model developed in Chapter 2 to understand the impact of increasing ride sharing participation¹ in a fleet of AVs and non-AVs. Section 3.2 describes the scenario settings that are used for the PP-SAEV model. Section 3.3

¹In this chapter the only two rider types (defined in Section 2.2) used are rider type 1 (non sharing) and rider type 5 (sharing preferred). Sharing participation is defined as the ratio of rider type 4 to the total number of riders in the model expressed as a percent. Hence $percentSharing = \frac{Rider\ type\ 4}{Rider\ type\ 4 + Rider\ type\ 1}$. The impacts of the other rider types 2,3,5 are discussed in detail in Chapter 5

discusses the results from the case study of NYC in terms of the service level, waiting time, riding time, percent of shared rides, GHG emissions and potential changes in the the service level spatially as a result of ride sharing. The conclusions of this chapter are discussed in Section 3.4

3.2 Simulation Scenarios

In order to study the system of ride sharing, several scenarios of the PP-SAEV described in Chapter 2 were run to analyze the impacts of adopting ride sharing and autonomous driving. The purpose of running the simulation scenarios was to study the change in the performance parameters of the ride sharing system with the RS adoption and the fleet size for AVs and for non-AVs. I used the demand from May 7th, 2014 from the yellow and green NYC-TLC dataset NYC DOT (2016b). I varied the percent of people who were willing to ride share with others (ride sharing participation) among the values {0%, 25%, 50%,75%, 100%}, and the fleet size among the values {3000,4000, 5000, 5500, 6000, 7000, 8000} for each of the autonomous vehicle case (percent of AV's = 100%) and the traditional (shifted) vehicles (percent of AV's = 0%) scenario. The percent of EV's was set to 0 for this study. These setting as well as the settings of other parameters are detailed in Table 3.1. In order to compare the results against the existing NYC taxi system, a base scenario with 13,500 traditional taxis running in shifts (13,500 is approximately the number of yellow taxis currently in operation NYC DOT (2014)) was constructed. Yellow taxis are the focus in this study because the areas they serve (e.g., Manhattan) have higher trip density and can potentially benefit from ride sharing more.

Table 3.1.

Simulation scenarios to understand ride sharing using shared autonomous vehicles and traditional taxi cabs

Parameter	Setting
<i>percentSharing</i>	{0%, 25%, 50%,75%, 100%}
<i>fleetSize</i>	{3000,4000, 5000, 5500, 6000, 7000, 8000, 13500}
<i>AVPercent</i>	{0%,100%}
<i>EVPercent</i>	0
<i>taxiCapacity</i>	4
<i>dtMode</i>	0.5

3.3 Results

The scenarios mentioned in Section 3.2 were evaluated to infer city level statistics such as service levels², fleet reduction, waiting time, resource utilization, distance traveled by the taxis and the riders, and spatial service level change to capture both the advantages and disadvantages of increasing RS adoption.

3.3.1 Fleet Reduction

By better utilizing the available space in each vehicle, ride sharing can help reduce the fleet size needed to serve the same demand. I consider a reduced fleet with ride sharing as having the same service level as the existing system, if it can serve the

²The service level is defined as the number of rider groups that were served by the system. (Rider groups may leave the system unserved if they could not be matched with a taxi within five minutes. This represents the situations that people give up and seek alternative transportation options after waiting for too long.)

same number of rider groups as the base scenario (percent of rider groups transported from their pick up points to the drop off points as compared to the scenario with 13,500 taxis without sharing). Figure 3.1 shows that a fleet of 5,500 autonomous vehicles is sufficient to serve the demands during the morning peak period without sharing. However, to satisfy the demand of the evening peak period at the same level as the base scenario (labeled as “B” in Figure 6), 5,500 autonomous vehicles with 100% ride sharing participation (labeled as scenario “A” in Figure 6) is needed. This service level can also be achieved by scenarios with other parameter sets as described in Table 3.2.

Table 3.2.

Scenarios and their parameter sets that can provide service within 2% of the base scenario, during the evening peak. We will from here on refer to these scenarios by the label.

Scenario Label	Parameter Set
B	Base Scenario 13,500 traditional taxis and 0% sharing participation
A	5,500 autonomous taxis and 100% sharing participation
A2	6,000 autonomous taxis and 75% sharing participation
A3	7,000 autonomous taxis and 25% sharing participation
S	8,000 traditional taxis and 100% sharing participation

It is notable that, with the same ride sharing participation, a fleet of 5,500 autonomous vehicles has similar service level as a fleet of 8,000 traditional taxis, indicating that autonomous driving is roughly equivalent to adding 2,500 traditional

taxis to the system. It can also be seen that shared autonomous taxis (Scenario A) has better service level than all the other equivalent scenarios in the morning peak. The main reason for this is that for the B and S scenario the number of taxis available during the morning peak is less than the number of taxis in the evening peak (Figure 2.5).

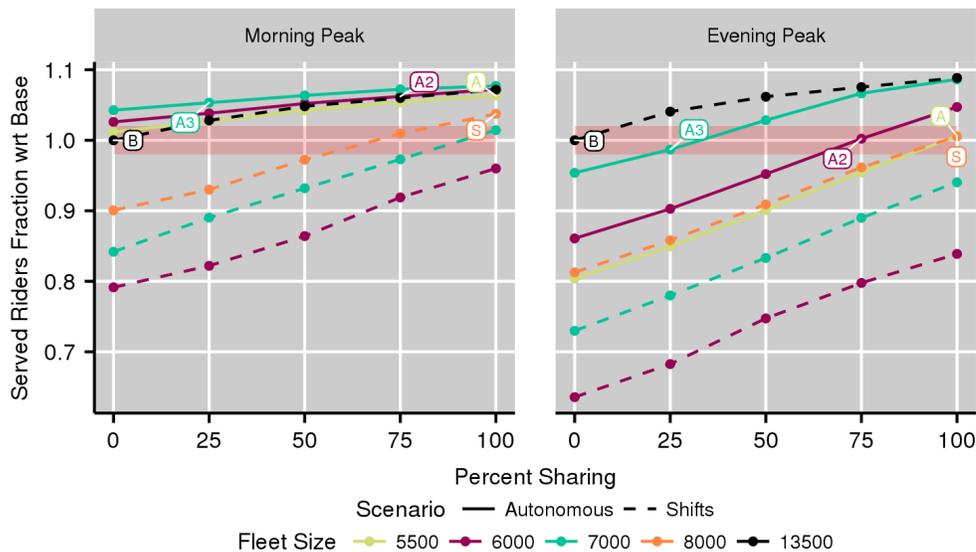


Figure 3.1. The average fraction of served rider groups (the ratio of rider groups served by the taxis to the total ride groups) in the morning (7:01am-3:00pm) and evening (5:01pm-12:00am) peak periods with different sharing and vehicle type scenarios. The light-red band indicates a service level within 2% of the base scenario

A potential concern with ride sharing and a reduced fleet is the increased waiting time and ride time for riders. However, my results (Figure 3.2(a)) show that, with a fleet of 5,500 autonomous taxis (Scenario A), the average waiting time for the served passengers (TR_W) only increases by less than two minutes compared to the base

scenario. The other sharing scenarios also have similar waiting time increase. One reason for the increased wait time is that the rider groups are spending additional time searching for a shared ride when an unoccupied taxi may be more readily available. The average ride time, on the other hand, increases with more ride sharing participation. In Figure 3.2(b), we can see that scenarios with higher sharing participation (A, S, A2) have a higher ride time, approximately 10 minutes longer on average.

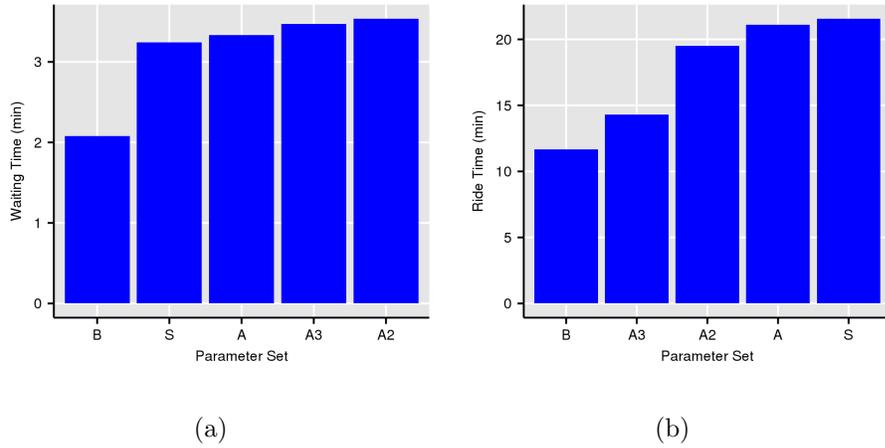


Figure 3.2. Change in (a) waiting time (TR_W) and (b) ride time (TR_{Ride}) under different ride sharing participation and fleet size scenarios

Figure 3.3(a) shows that the extra distance traveled by shared riders increases only with the percent sharing and does not change significantly with the fleet size. This is due to the fact that higher sharing participation increases the number of rider groups that shared the ride together (Figure 3.4(c)). Having more rider groups sharing a ride increases the required deviations and the extra distance. However, this increase is, on average, less than 33% of the original trip distance (Figure 3.3(b)).

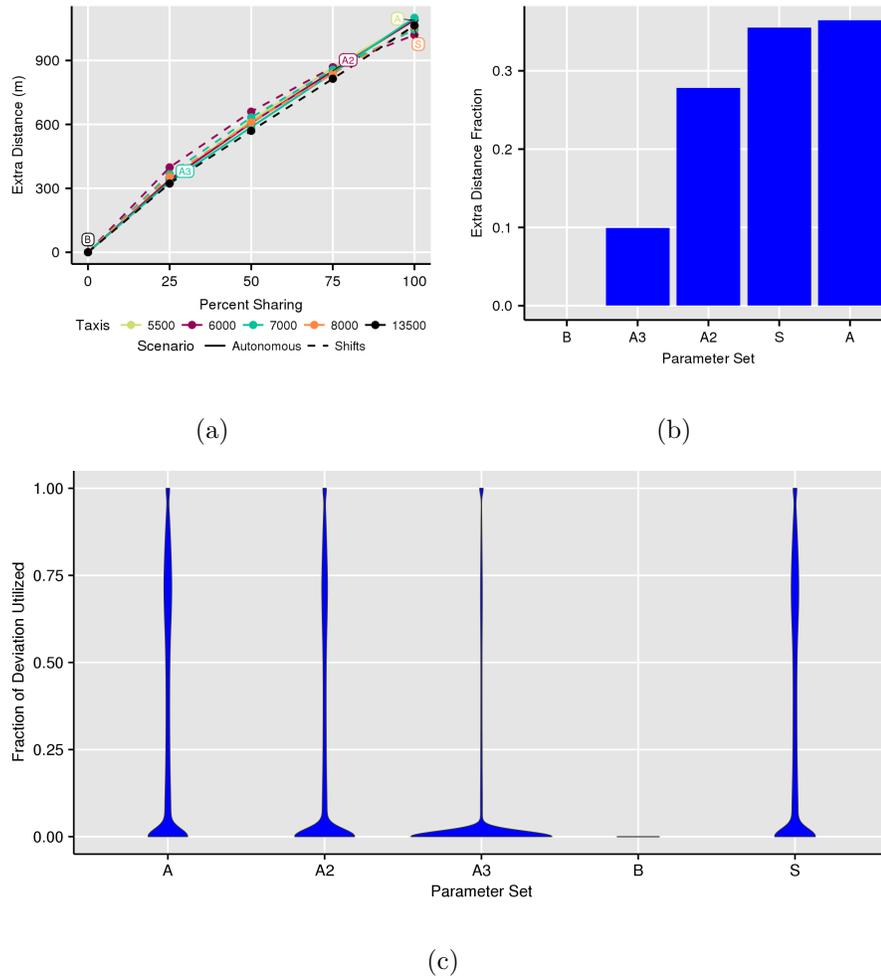


Figure 3.3. (a) Average increase in trip distance for the riders. (b) Average fractional increase in distance for the riders. (c) The average fraction of individual utilized trip deviation relative to the individual acceptable level, presented as a violin plot. (A violin plot is two vertical density plots attached together at their bases. The vertical bar shows the range of the values while the horizontal width shows the density of the points at that value).

Compared to the individual trip deviation tolerance, Figure 3.3(c) shows that very few people utilize the full tolerance level in the sharing. About 40% of the riders in scenarios S, A, and A2 only used less than 25% of the tolerable trip deviation. In scenario A3, where the sharing participation is lower (25%), the utilized tolerance is even less (over 70% of the riders only used less than 10% of their trip deviation tolerance).

3.3.2 Increased Resource Utilization

The vehicle occupancy (calculated as the average number of passengers in a taxi) is a measure of the utilization efficiency of the taxis in the system. Higher occupancy indicates better system efficiency. Figure 3.4(a) indicates that, on average, the vehicle occupancy increases with higher participation of ride sharing. With the 4-seat vehicle capacity modeled in this study, the occupancy can increase from 1.2 (Scenario B) to 3 (Scenarios S and A). The average number of groups in a vehicle indicates the average number of shares taking place. This value is 1 without sharing (Scenario B) and increases to 2.5 per vehicle (Scenarios S and A) as a result of sharing (Figure 3.4(c)). This results show that studies which constrain the sharing to be only between two rider groups are limiting the sharing participation to be at the 50-75% level Kleiner et al. (2011); Agatz et al. (2011). The results show that the actual percentage of sharing participation is lower than the percentage of rider groups that are willing to share. In scenarios that all riders are willing to share, the actual percentage of rides that are shared is only about 80% (Figure 3.4(b)).

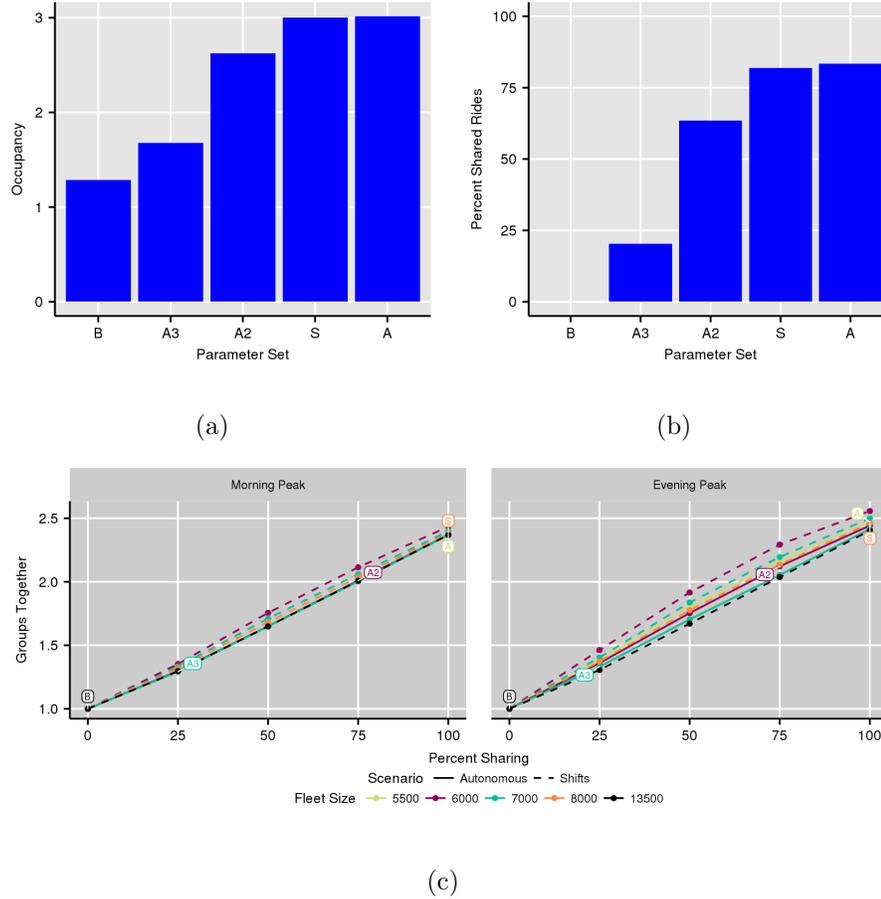


Figure 3.4. Sharing performance: (a) Change in occupancy in different scenarios with similar service levels; (b) The actual percentage of rider groups that participated in sharing; (c) Number of rider groups shared a ride together varied with service levels

3.3.3 Environmental Benefits

My simulations have shown that the total distance traveled by the taxis reduces as the ride sharing participation increases. Also, when compared to the base scenario,

we see a reduction of approximately 2.8×10^6 km in total per day (which is 55% of the distance traveled by the taxis) in scenario A, with 5,500 autonomous taxis and 100% ride sharing participation (Figure 3.5(a)). This reduction in total travel distance translates to approximately lowering CO₂ emissions by 725 metric tonnes per day. Compared the 40% trip reduction in NYC from ride sharing estimated by Santi et al. (2014), our value is higher because we did not constrain sharing to be only formed between two groups. We note though, that this emission reduction is computed purely on the basis of total distance traveled. However, as estimated by Wadud et al. (2016), autonomous vehicles may help achieve, on an average, a net of 10-15% energy consumption saving due to potential changes in driving patterns such as platooning, smoother driving, crash avoidance mechanisms. Another paper, Gawron et al. (2018), has used life cycle assessment to estimate that introducing connected automotive vehicles could reduce energy consumption by 9% due to these driving pattern changes. If we consider this additional 9% reduction, for scenario A, the overall reduction in CO₂ emissions will be 802 metric tonnes per day. On the other hand, for ride sharing with traditional taxis, even though the number of vehicles in Scenario S is higher, the total travel distance is lower than Scenario A (approximately 45% of the base scenario B). This results in a reduction of 3.42×10^6 km or a reduction of approximately 866 metric tones per day of CO₂ emissions.

The percentage of distance for which the taxi is occupied can be studied to gauge the efficiency of the system from an environmental perspective. Figure 3.5(b) shows that the percent of occupied distance traveled by the traditional taxis (with shifts) increases with higher sharing participation (from scenario B and S). For the autonomous taxis, the fraction of distance for which the taxi remains occupied stays relative stable regardless of the level of sharing participation. This tells us that, even though the taxi is serving more customers, it will be traveling less to do so.

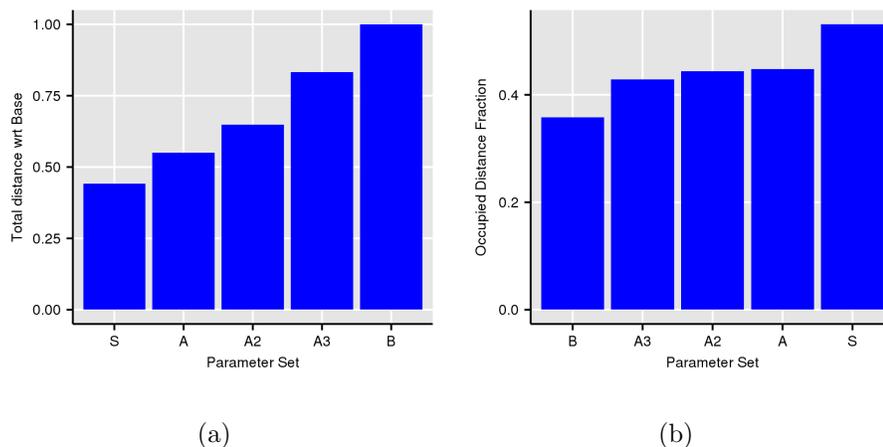


Figure 3.5. (a) Total distance traveled with respect to the base scenario (Scenario B) for different scenarios. (b) The fraction of occupied distance (distance traveled to deliver passengers) for all scenarios

3.3.4 Spatial coverage change

In order to evaluate the impact of ride sharing on the spatial distribution of service levels, I compared scenarios that have nearly equivalent service levels (within 5% difference in served riders) and analyzed the fractional change in service level in different regions. Figure 3.6(e) shows the fractional difference between the base scenario (Scenario B) and Scenario A with 5,500 autonomous vehicles and 100% sharing. We can see that using conventional taxi cabs without sharing as opposed to the SAVs has a positive effect in the suburban region (shown as the purple and blue cells in Figure 3.6(e)), but has a negative effect (shown as the red cells in Figure 3.6(e)) in the regions where the demand is the most dense (Manhattan and, more significantly, Times Square). To identify whether this service coverage change is due to autonomous driving or sharing, I further compared scenarios B and A3 to

evaluate the impact of autonomous driving with no/low sharing and scenarios B and S (Figure 3.6(b)) to evaluate the impact of sharing with traditional taxis. In both cases, I observed similar spatial service coverage change. On the other hand, when I compare scenarios A and S (Figure 3.6(d)) or scenarios A and A3 (Figure 3.6(b)), the spatial service coverage is quite similar. These results show that both ride sharing and autonomous vehicles will cause taxis to focus more on areas with higher demands. While better serving the demands in the regions with more demand, the reduced fleet decreases the service level in the suburban regions (more in Section SIC). To remedy this disproportionate change in service, appropriate policies would be needed to insure service in the suburban regions. Such policies could include providing price incentives or restricting a portion of the fleet to the suburban regions (similar to the way NYC currently distinguishes between Green and Yellow taxi cabs as mentioned in NYC DOT (2014)).

3.3.5 Changes in efficiency of matching

The response time (TR_R) represents the efficiency of matching. Our results show that TR_R is lower in the scenarios where the riders are homogeneous (all sharing or all non-sharing) but higher in scenarios with a mix of sharing and non-sharing riders (Figure 3.7). The reason for this is that our model assumes that all riders who are willing to share will first search for a shared ride. So in scenarios with mixed rider types (some rider groups are willing to share and some are not), it is possible that a sharing taxi is close to a rider group that is not willing to share or an occupied non-sharing taxi is close to a rider group that is willing to share. In these situations, the matching cannot be formed. As a result, the time required to find a match increases

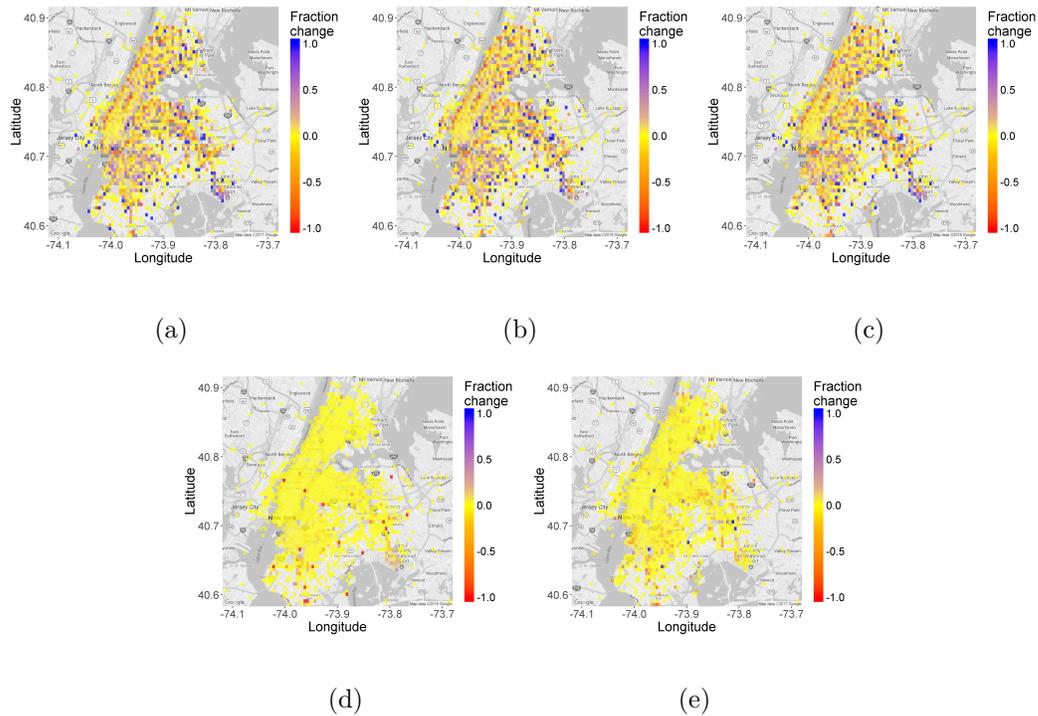


Figure 3.6. Spatial distribution change in terms of service level. (a) The fraction change in service level in Scenario A compared to Scenario B, calculated as $\frac{Service_B - Service_A}{Service_B}$. (b) The fraction change in service level in Scenario B compared to Scenario A3, calculated as $\frac{Service_B - Service_{A3}}{Service_B}$. (c) The fraction change in service level in Scenario B compared to Scenario S, calculated as $\frac{Service_B - Service_S}{Service_B}$. (d) The fraction change in service level in Scenario A compared to Scenario A3, calculated as $\frac{Service_A - Service_{A3}}{Service_A}$. (e) The fraction change in service level in Scenario A compared to Scenario S, calculated as $\frac{Service_A - Service_S}{Service_A}$. Blue color indicates Scenario B or S has better service level than Scenario A, while red color indicates Scenario A provides better service. Grid resolution is $0.005^\circ \times 0.005^\circ$, roughly equivalent to $0.5km \times 0.5km$.

in the scenarios with mixed rider types, indicating a lower efficiency of matching. However, the delay is less than one minute.

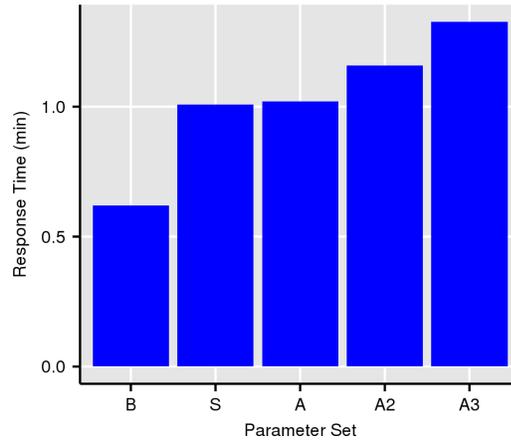


Figure 3.7. Change in response time (TR_R) under different ride sharing participation and fleet size scenarios

3.4 Conclusion

The PP-SAEV model was used to determine the impact of increasing RS adoption in AVs and non-AV fleets. The insights gained from this study are: 1) while maintaining the same service level, ride sharing with autonomous vehicles can potentially decrease the fleet size by up to 59% without significant waiting time increase or additional travel distance; 2) the benefit of ride sharing is significant with increased occupancy rate (from 1.2 to 3), decreased total travel distance (up to 55%), and reduced carbon emissions (725 metric tonnes per day). However, since the occupancy rate does not go beyond 3 for a 4 seat vehicle, we can also conclude that ride sharing has an efficiency limit ; 3) constraining the sharing to be only between two groups limits the sharing participation to be at the 50-75% level and underestimates the potential benefits; and 4) ride sharing may reduce the service level in the suburban

areas, which will require complementary policies or incentives to help balance service in different regions.

4. RIDE SHARING ADOPTION IN MIXED AUTONOMOUS VEHICLES AND TRADITIONAL VEHICLE FLEETS OF VARYING CAPACITIES

4.1 Introduction

Chapter 3, considered taxi fleets which had either fully autonomous vehicles or fully non-autonomous vehicles, and found scenarios that had the same percent of riders served across multiple scenarios at 5 discrete levels of ride sharing. Then, various performance parameters of the system (for example waiting time, environmental impacts, ride time, etc) were compared for the scenarios that had the same service level. However, in reality, the adoption levels of both AV and RS will increase gradually (continuously). Thus as these adoption levels increase gradually, systems where the vehicle fleets have a mix of AV and non-AV will exist. Additionally, since we have no data on how heterogeneous the riders in the system are, we need to test the system for sensitivity of such sharing heterogeneity. However, studying such systems using models such as the PP-SAEV model would require large number of parameters to be varied, and hence a large number of scenarios to be run. Hence, methods are required to find comparable scenarios for such high dimensional systems.

The work in this chapter uses meta-model based simulation optimization to extend the methodology of finding scenarios that are comparable to each other for scenarios that can have, fleets with a mixture of autonomous vehicles and shifted vehicles, the taxis can be set to have different capacity, and the distribution of the

deviation tolerances of the riders could be different. These comparable scenarios will then be used to build a model to predict fleet size, understand other system performance parameters of the system (like waiting time, and environmental impacts) and understand the limitations of ride sharing.

In the previous chapter, the parameters of the agent-based model developed in Chapter 2 were set as per Table 3.1. The *fleetSize*, and *percentSharing* were varied in discrete steps only, and the values of *AVPercent* were set as either 0 and 1. This limited the inference to only systems that had either completely autonomous, or completely traditional taxi fleets. Additionally, the value of *taxiCapacity* was set to 4 and the value of *dtMode* was set to 0.5, further limiting our ability to understanding the effect of varying these parameters. Since the number of parameters varied in Chapter 3 were small, with two continuous parameters (ride sharing and fleet size) and one discrete parameter (AV adoption), I was able to find five comparable scenarios by running 50 combinations of inputs. In cases that the model has more parameters, the parameter space is too large for comprehensively evaluating all the possible scenarios. In order to extend the analysis presented in Chapter 3, to include all possible parameter settings that can help us understand how different adoption levels of RS and AV can influence system performance when taxis fleets are allowed to have different maximum capacities, in this paper we used meta-model based simulation optimization to find scenarios that had the same target service level. These inferences can be used by the modeler to infer the important parameters that affect the output that we look to optimize. Then we will use the scenarios that we have identified to have the same service level to make predictions on the correct fleet size for a variety of systems, and also make inferences on the environmental impacts of different possible systems. In this study, in order to understand the performance of fleets with a mix of AVs and non-AVs, with varying fleet sizes of different capacities

and sharing participation, I varied the parameters of the agent-based model as per Table 4.1.

Table 4.1.
Simulation scenarios to study ride sharing adoption in mixed autonomous vehicles and traditional vehicle fleets of varying capacities

Parameter	Setting
<i>percentSharing</i>	{0%, ..., 100%}
<i>fleetSize</i>	{5000, ..., 15000}.
<i>AVPercent</i>	{0%, ..., 100%}
<i>taxiCapacity</i>	{2, 4, 6}
<i>dtMode</i>	{0, ..., 1}

In Table 4.1, there are 4 parameters (*percentSharing*, *fleetSize*, *AVPercent*, *dtMode*) that are varied as continuous variables, and one parameter that is varied as a discrete variable (*taxiCapacity*). If these variables were to be discretized and run simulation scenarios like the study in Chapter 3, there would be $10^4 \times 3 = 30000$ different scenarios for the PP-SAEV (assuming each continuous variable is discretized in 10 steps). Additionally, it would difficult to know if any of these scenarios actually was within the range of the target service level (0.82-0.86), thus making it essential to run each of the 30000 scenarios to find those that met the service targets. Since it is not computationally feasible to run that many scenarios (each scenario run takes 3-7 days to complete), a method is needed to run only those scenarios that I suspect will lie within the target service range. In order to do this, a meta-model based methodology was used to initially fit a model which describes the relationship between the output (service level) and input parameters (Table 4.1, and then make

predictions to find which parameter setting would give a service level that meets the target.

The rest of this chapter is divided into 4 sections, Section 4.2 discusses relevant literature that pertains to work done to find comparable scenarios in agent-based models, Section 4.3 discusses the experimental design methodology that was used to run the ABM with appropriate parameter settings, and Section 4.4 uses the methodology on a case study and discusses fleet sizing, environmental impacts, and ride sharing limitations for that case study. the conclusions of this study are outlined in Section 4.5.

4.2 Literature Review

The use of agent-based models to study systems with complex interactions have increased in recent years owing to the rapid increase in processing power (agent-based models are computationally intensive). Agent-based models have been used in the fields such as public policy, social science, economics, biology, and military (Heath et al., 2009) and have also been used to simulate optimum decision making in these fields (Barbati et al., 2012). The ability to include transportation networks, consider individual demands, include parameters that define agent heterogeneity, enable inter-agent communication, and build protocols for finding the optimum value of a certain objective for the agents has made agent-based modeling a great tool for researchers in the field of transportation. Bazzan and Klügl (2014) has conducted an extensive review of papers that have used agent-based models in transportation research, showing the wide use of agent-based models to simulate traffic flow, routing, and collaborative driving etc.

In the field of ride sharing and autonomous vehicles as discussed in Section 1.1, 1.2 and 1.3, ABMs have been used by many studies (Mourad et al., 2019; Taiebat et al., 2018) to understand system effects. For example, Alonso-Mora et al. (2017); Ma et al. (2015); Qian et al. (2017); Fagnant and Kockelman (2014); Fagnant et al. (2015); Chen et al. (2016a); Simonetto et al. (2019); Boesch et al. (2016); Zhu and Kornhauser (2017); Farhan and Chen (2018); Loeb et al. (2018); Bauer et al. (2018) built agent-based models to analyze emerging transportation technologies like ride sharing, autonomous vehicles and electric vehicles. They studied the effect of varying their model input parameters (which consisted of the maximum allowed ride sharing delay and the vehicle capacity) on the output parameters (service levels, waiting time, and travel delay experienced).

However, these papers only evaluated the models performance on a maximum of 3 dimensions. For example, Qian et al. (2017) studied the effect of the provided discount, the day of the week, and the time of the day on the reduction of total trip mileage. Ma et al. (2015) plotted several performance parameters like satisfaction rate, ride sharing rate, average saving rate against the demand increase, and the money to time rate for different ride matching schemes. Simonetto et al. (2019) considered different demand rates, fleetsizes. Loeb et al. (2018); Bauer et al. (2018) simulated a system of shared autonomous electric vehicles (SAEV) using ABMs. Loeb et al. (2018) measured the charging performance as a function of vehicle range, and charging rate and Bauer et al. (2018) measured cost of service, and the minimum taxi fleet required while varying the EV charging station power rating, the EV range. However, in modern and evolving transportation systems with fleets that have mixed compositions, often three dimensions are not sufficient to capture all the potential changes.

Alternatively, some of the studies evaluated the performance of SAV systems for multiple parameters by modeling a limited number of specific scenario settings. Farhan and Chen (2018); Chen et al. (2016a) used an ABM to simulate SAEVs and simulated 8 parameter sets and 5 parameter sets respectively. Fagnant and Kockelman (2014) built an SAV model and ran 25 different scenarios with various parameter settings. However, simulating a limited number of scenarios based on specific parameter settings do not give an idea of how an output changes with any of the input parameters. Second, even though these papers evaluated multiple scenarios and multiple performance indicators, there is a need to compare these scenarios against one another on a fair ground as detailed in Section 1.2. For example, (Alonso-Mora et al., 2017) built an agent-based model to study ride sharing using shared autonomous vehicles of varying capacities. They tested the effect of varying the maximum allowed time for a shared trip and showed its impact on the percent of served requests, waiting time, travel time, and percent of shared rides. However, the scenarios that were presented in this work were not equivalent because they did not have the same level of service.

Metamodels, involve estimating the relationship between the simulation output and the parameter settings and then using optimization techniques to find the optimal parameter setting and have been used in many simulation studies (Box and Wilson, 1951; Amaran et al., 2016; Ankenman et al., 2010; Barton, 2009; Chow and Regan, 2014; Ekström et al., 2016; Osorio and Bierlaire, 2013) and can be used to identify comparable scenarios in the PP-SAEV model. There are two main advantages of using meta-models to finding sets of comparable scenarios. First, the method does not require any hyper-parameters to be set in order to be used unlike level set approximation approaches (Huang and Zabinsky, 2013) or metaheurstics (Mishra, 2018; Joines et al., 2002; Ding et al., 2005). Second, at each simulation optimiza-

tion stage, we can infer the important parameter settings easily by observing the meta-model (which relates the inputs and outputs using regression). This can help us guide the simulation optimization.

4.3 Method

The objective is to find multiple combinations of input parameters (X), which, when applied to the PP-SAEV model, produce an output (Y) that is within range ($[Y_l, Y_u]$) of a target value (Y_T). We will assume that we have access to N starting points, which could be either solutions that satisfy this target value or scenarios that would be able to enhance our understanding of the system. While we prefer solutions that meet our target output since starting with these solutions would enhance our knowledge of the surface in the region of interest, we also consider points that do not meet the target in the initial set because they could help improve the overall initial estimation of the response surface. The initial solutions can be obtained by domain expert knowledge, or by varying a subset of the parameters while keeping the others fixed. Then, we can use the following method to find the target output value:

1. Begin with N starting points. Add those solutions which meet the target value to a set of accepted solutions A .
2. Generate N designs using the N starting solutions as center points. Each design will have R runs, each with different input parameters as generated by the design. The designs can be generated using any of the techniques available (for example, factorial designs (Fisher, 1935; Box and Hunter, 1961b,a), central composite design (Box and Wilson, 1992), and Box-Benkhen design (Box and Behnken, 1960) etc.) The value R is chosen based on the simulation budget.

3. Run the ABM with the parameter settings ($X = (X_c, X_d)$ where X_c are the settings for the continuous parameters, and X_d are the parameter settings for the discrete parameters) generated in step 2 and obtain the outputs (Y). Add these $R \times N$ model runs to *Scenarios*.
4. Build a meta-model $f(X)$ which approximates the response Y as a function of the system parameters $X \in \text{Scenarios}$. The model that is chosen should have good predictive accuracy and can be chosen from a variety of meta-model classes like linear models, non-linear models like GLMs, spline based models, and neural networks (Barton, 2009).
5. Use an optimization algorithm to find M solutions that would have the output within the target range. The optimization model can be in the form of:

$$\text{Maximize} \quad \sum_{d=1}^D \sum_{m=1}^{|M|} \min(\|\bar{X}_{c,a} - \bar{X}_{c,m}\|; X_a \in A, X_{d,m}, X_{d,a}) \quad (4.1)$$

$$\text{s.t.} \quad Y_l \leq f(X_m) \leq Y_u \forall X_m \in M \quad (4.2)$$

$$X \in \mathcal{X} \quad (4.3)$$

where $\bar{X}_{c,a}$ and $\bar{X}_{c,m}$ are the normalized values of $X_{c,a}$ and $X_{c,m}$ respectively, $X_m = (X_{c,m}, X_{d,m})$, $X_a = (X_{c,a}, X_{d,a})$, D is the number of unique settings that the discrete parameters can take and $X_{d,a}, X_{d,m}$ is the d -th setting for the discrete parameters¹. Y_l and Y_u represents an acceptable lower and upper bound of our target service level, respectively. We find scenarios that fall within a band of our target response because it would be extremely challenging to find scenarios that exactly meet our target output. Here \mathcal{X} is the set of all possible

¹If there are V_D discrete parameters, $D = \prod_{i=1}^{V_D} S_i$ where S_i is the number of unique values discrete variable i can take

values the parameters X can take. The objective function (Equation 4.1) maximizes the total minimum euclidean distance between the normalized values of the new proposed M solutions and the existing accepted solutions A . We use this objective function, so that at each iteration, we are evaluating solutions in parts of the solution space that is least explored by the model. I choose to normalize the variables so that a particular parameter is not weighted higher than any other parameter. The value M is chosen based on our simulation budget and the predictive accuracy of our meta-model.

6. Run the simulation model with the input parameters of the M solutions to verify the output. Set $Scenarios \leftarrow Scenarios \cup M$ The set of inputs $x_m; m \in 1 \dots M$ with outputs within our target range ($[Y_l, Y_u]$) are added to our verified set of solutions (A). We can compare the other key performance parameters of these accepted solutions to give us a like-to-like comparison.
7. If budget permits and if more information (better model or more comparable scenarios) is needed, proceed to step 8 else go to step 10
8. If predictive accuracy of meta model is sufficient ($\alpha\%$ of the M solutions are in the set A) go to 4
9. Select N new starting points from the set of M simulation parameters which were not a part of the accepted solution set A . We select solutions that were not a part of set A so that we can improve the estimation of the metamodel in regions where our estimates were inaccurate. Go to step 2.
10. Else Stop

A pictorial representation of this method is included in Figure 4.1.

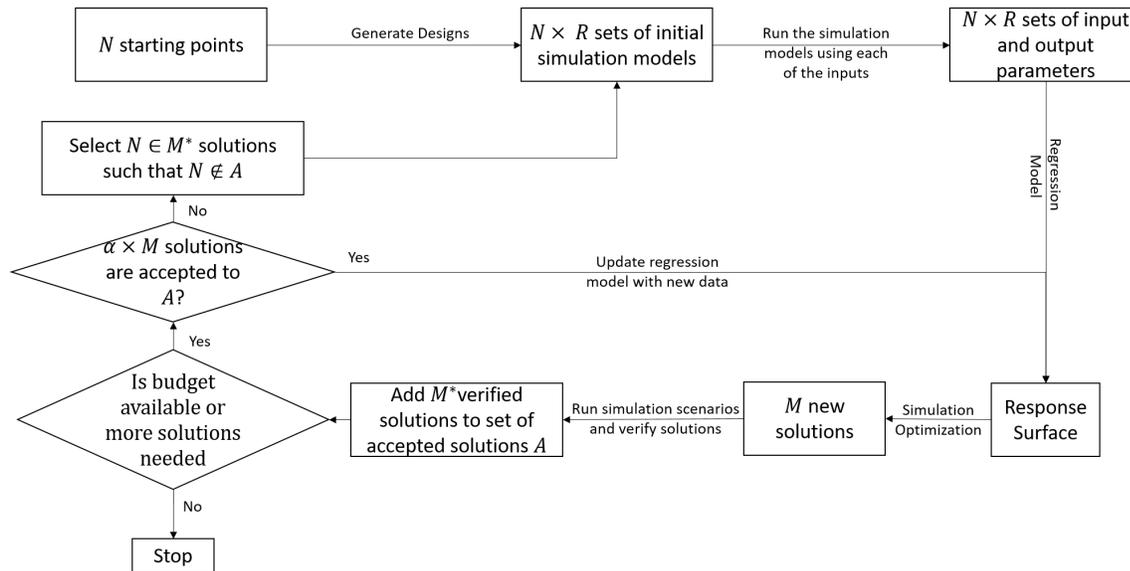


Figure 4.1. Flowchart of meta-model based simulation optimization method to find comparable scenarios

In the next section we will apply this methodology to the PP-SAEV model to demonstrate the use of generating a set of solutions with a target response to make decisions on fleet sizing and inferences on system performance parameters like CO_2 emissions and waiting time for riders.

4.4 Case Study Results

In order to understand the impact of an increasing rate of adoption of AVs and ride sharing, as well as evaluate systems with different mix of rider sharing preferences, I use the PP-SAEV model built in Chapter 2 with the range of parameters mentioned in Table 4.1. In Chapter 3, I identified 5 scenarios, which had a service

level equal to the service level of the base scenario (table 3.2). In order to find similar equivalent scenarios, I used the method described in Section 4.3. First, I selected $N = 5$ starting solutions (table 3.2) which I knew have the same service level as the target service level of 0.86. I was running the models on Purdue's Snyder computing cluster on which I had access to 5 nodes with 20 cores each. This meant that at a time I was able to run up to 100 scenarios of PP-SAEV with different parameter settings. So, for each starting point, and each capacity level, I created a half 2^4 fractional factorial design. The total number of simulation scenarios that were run in Step 1 were $2^4/2 \times 3 \times 5 = 120$. Even though this number is greater than 100 (the number of scenarios that I could run at a time), the scenarios that have a lower sharing participation get completed much quicker than those with high sharing percentage. Hence, in order to optimize the computing resources available to me, I always ran a little more scenarios in a step than what my budget allowed me. The parameter settings for the $1/2$ factorial design are given in table 4.2

Then, I used a linear regression model with 2 factor interactions to model the relationship between the fraction of riders served by the system (*ServedPC*) during the evening peak hours (6:00pm to midnight as per figure 2.4(c)) and the input parameters in table 4.2 for these 120 scenario runs. Then, using the optimization model represented in equation 4.1 - 4.3, I generated the next set of 125 scenarios. Since it is difficult to solve the optimization model (the constraints and the objective function are non linear), I used an approximate method to select a set of comparable scenarios.

1. I discretized the continuous input parameters (*fleetSize* in steps of 100 and *PercentAV*, *PercentSharing*, *dtMode* in steps of 0.025) and enumerated all possible combinations of these parameters

Table 4.2.

The new designs that were used to generate the response surface. Each cell is a vector of 2 numbers which correspond to the low, and high settings for that variable.

Derivative Scenario	<i>fleetSize</i>	<i>PercentAV</i>	<i>PercentSharing</i>	<i>dtMode</i>	<i>taxiCapacity</i>
A	{5000, 6000}	{0.8, 1}	{0.8, 1}	{0.25,0.75}	{2,4,6}
B	{11500, 13500}	{0, 0.2}	{0, 0.2}	{0.25,0.75}	{2,4,6}
S	{7500, 8500}	{0.8, 1}	{0.8, 1}	{0.25,0.75}	{2,4,6}
A2	{6500, 7500}	{0.8, 1}	{0.6, 0.9}	{0.25,0.75}	{2,4,6}
A3	{5500, 6500}	{0.8, 1}	{0.1, 0.4}	{0.25,0.75}	{2,4,6}

2. I predicted the response for each of these combinations of inputs
3. I filtered these combinations of inputs to keep only those combinations that had a predicted response within my target service band (potential next simulation runs T).
4. I used the following greedy algorithm to select the top N solutions based on the distance metric (equation 4.1) from the set of accepted solutions A :

- (a) Calculate the distance metric for all points in set T from all points in set A
- (b) Filter set T by removing all scenarios with a distance metric of less than 0.02 units
- (c) Set $Temp = A$
- (d) Repeat for $\text{length}(T)$ repetitions :
 - (e) Select the scenario in set T with the largest distance metric and add it to set $Temp$
 - (f) Calculate the distance metric $Dist = ||\bar{X}_{c,a} - \bar{X}_{c,t}||$ for all points $t \in Temp$
 - (g) End For
- (h) Select the top N from set T solutions with largest $Dist$ calculated in step 4f

While this algorithm is not guaranteed to give a globally optimum set of parameters to run, for the purpose of this study, the solutions generated by this algorithm are sufficient to give us the inferences that we need.

I followed metamodel based algorithm presented in Section 4.3 for 5 iterations (with a total of 620 simulation scenarios) and obtained 518 combinations of input parameter settings which were within my target service range of 0.84 to 0.88. At the end of the 5th iteration, the minimum distance between any two solutions is 0.02 units. The linear regression model from the 620 simulation runs is shown in Table 4.3

Table 4.3.: Significant terms in the least squares regression

Term	Coef	SE Coef	T-Value	P-Value
<i>(Intercept)</i>	$-2.01e - 02$	$2.61e - 02$	$-7.70e - 01$	$4.42e - 01$
<i>fleetSize</i>	$1.47e - 04$	$3.84e - 06$	$3.83e + 01$	$< 2e - 16$
<i>percentSharing</i>	$1.12e - 01$	$1.21e - 02$	$9.30e + 00$	$< 2e - 16$
<i>PercentAV</i>	$2.74e - 01$	$1.71e - 02$	$1.60e + 01$	$< 2e - 16$
<i>taxiCapacity4</i>	$-6.56e - 02$	$1.40e - 02$	$-4.68e + 00$	$3.80e - 06$
<i>taxiCapacity6</i>	$-1.52e - 01$	$1.46e - 02$	$-1.04e + 01$	$< 2e - 16$
<i>dtMode</i>	$1.93e - 04$	$9.71e - 03$	$2.00e - 02$	$9.84e - 01$
<i>I(fleetSize²)</i>	$-5.96e - 09$	$1.50e - 10$	$-3.97e + 01$	$< 2e - 16$
<i>I(PercentAV²)</i>	$-1.85e - 02$	$5.70e - 03$	$-3.25e + 00$	$1.25e - 03$
<i>I(dtMode²)</i>	$7.74e - 03$	$5.14e - 03$	$1.50e + 00$	$1.33e - 01$
<i>fleetSize</i> :	$-7.97e - 06$	$9.65e - 07$	$-8.26e + 00$	$1.52e - 15$
<i>percentSharing</i>				
<i>fleetSize</i> :	$-1.88e - 05$	$1.12e - 06$	$-1.67e + 01$	$< 2e - 16$
<i>PercentAV</i>				
<i>fleetSize</i> :	$8.44e - 06$	$1.16e - 06$	$7.30e + 00$	$1.26e - 12$
<i>taxiCapacity4</i>				
<i>fleetSize</i> :	$1.87e - 05$	$1.25e - 06$	$1.49e + 01$	$< 2e - 16$
<i>taxiCapacity6</i>				
<i>fleetSize : dtMode</i>	$-1.06e - 06$	$6.79e - 07$	$-1.56e + 00$	$1.20e - 01$
<i>percentSharing</i> :	$-1.84e - 02$	$4.66e - 03$	$-3.95e + 00$	$9.04e - 05$
<i>PercentAV</i>				

continued on next page

Table 4.3.: *continued*

Term	Coef	SE Coef	T-Value	P-Value
<i>percentSharing</i> <i>taxiCapacity4</i>	$3.94e - 02$	$4.45e - 03$	$8.85e + 00$	$< 2e - 16$
<i>percentSharing</i> <i>taxiCapacity6</i>	$5.72e - 02$	$4.60e - 03$	$1.24e + 01$	$< 2e - 16$
<i>percentSharing</i> <i>dtMode</i>	$1.10e - 02$	$3.74e - 03$	$2.94e + 00$	$3.49e - 03$
<i>PercentAV</i> <i>taxiCapacity4</i>	$2.68e - 02$	$5.67e - 03$	$4.73e + 00$	$2.93e - 06$
<i>PercentAV</i> <i>taxiCapacity6</i>	$5.45e - 02$	$5.82e - 03$	$9.37e + 00$	$< 2e - 16$
<i>taxiCapacity4</i> <i>dtMode</i>	$3.67e - 03$	$3.18e - 03$	$1.15e + 00$	$2.49e - 01$
<i>taxiCapacity6</i> <i>dtMode</i>	$7.60e - 03$	$3.61e - 03$	$2.11e + 00$	$3.56e - 02$

Figure 4.2 shows the parameters of all of the scenarios that have the same service level. In the plot of *fleetSize* with *PercentAV* and *PercentSharing*, we see that for higher capacities, the fleet size required is much lower. Also, we can see that scenarios with capacities 6 form a lower boundary for the scenarios that are equivalent to each other at $(fleetSize, PercentAV, PercentSharing) = (6000, 1, 1)$. There seems to be no apparent upper bound as the scenarios with capacity 2 are limited by their low capacity, and no increase in fleet size, AV percent, or sharing participation seems to have any impact on the percent of riders served after a certain value. We also see

from figure 4.2 and table 4.3 that the variable *dtMode* does not have a strong impact on the percent of riders served as compared to the other variables.

The following sub sections will more formally explore the relations between the input parameters, and the other output performance indicators of the system.

4.4.1 Fleet Sizing

One of the important decisions while considering transportation systems such as the one described in this dissertation is the number of vehicles that will be required to maintain the current service level. By estimating the fleet size, a transportation planner can also estimate the budget that would be required, considering that other input parameters (*percentSharing*, *percentAV*, *taxiCapacity*, *dtMode*) are fixed. In order to estimate the fleetsize required I built seperate regression models for each level of *taxiCapacity*. For *taxiCapacity* = 2, I used Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), for *taxiCapacity* = 3, 4 I used a log transformed linear model. The prediction model for *taxiCapacity* = 2 had an $R^2 = 0.32$, for *taxiCapacity* = 4 had an $R^2 = 0.94$, and for *taxiCapacity* = 6 had an $R^2 = 0.97$. The reason that the *taxiCapacity* = 2 model has a low goodness of fit (and large errors - figure 4.3(a)) is that there are potentially many solutions which have the same *percentSharing*, *percentAV* and *dtMode*, and different fleet size (figure 4.2), however, the model would only predict one value at a time. Figure 4.3(c) are the partial dependence plots for the fleet size models. Partial dependence plots illustrate the predicted change in the response variable for a marginal change in the predictor variable (Brandon Greenwell, 2016). Using Figure 4.3(c), we can infer the fleet size change that would be needed as the AV and ride sharing adoption increases. From the contours of *taxiCapacity* = 4, we can infer that in order to reduce the fleet size

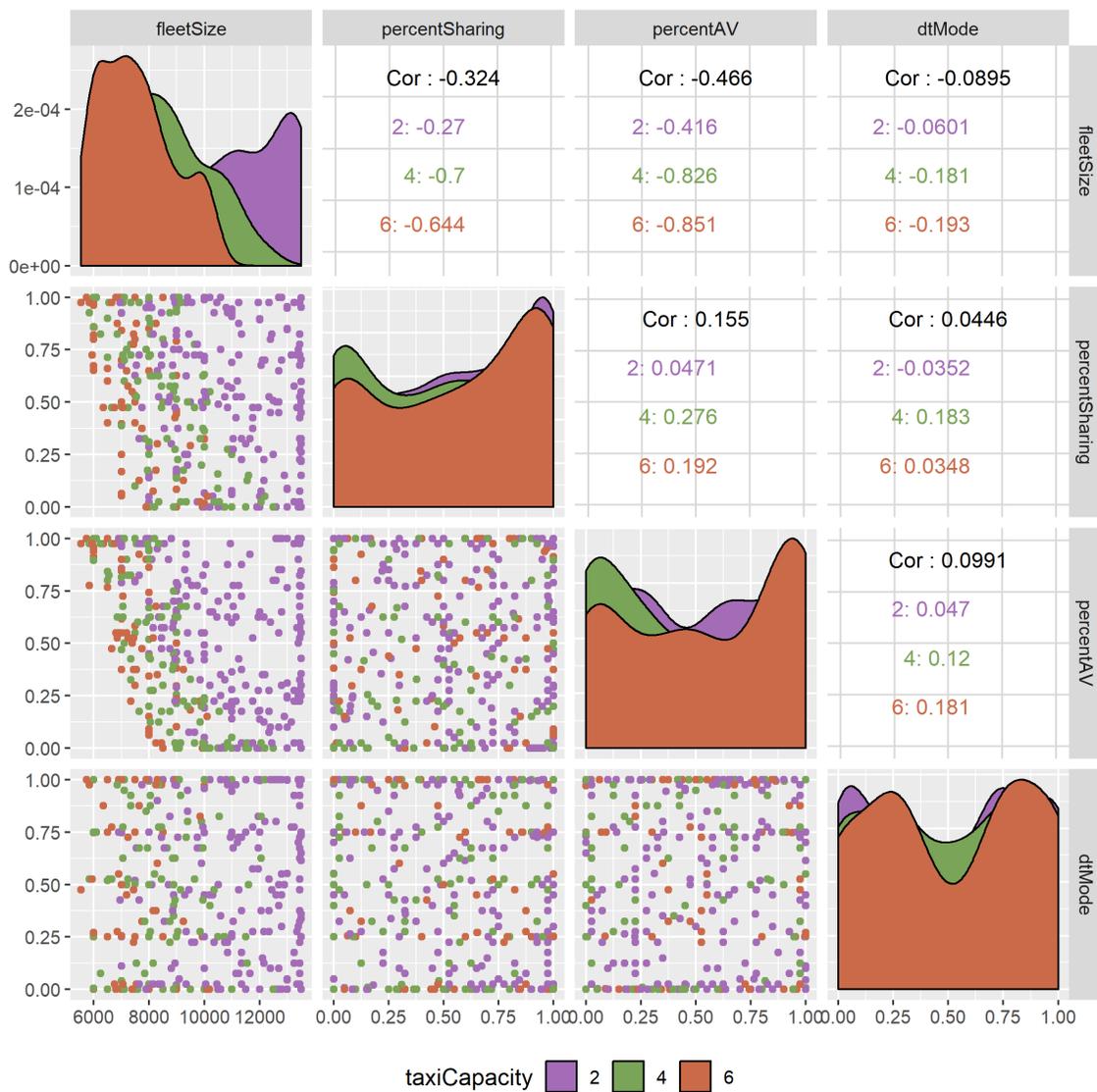


Figure 4.2. The input parameters of all the scenarios that have equal service level. The color refers to the Capacity of the taxis (2 - red), (4 - blue) and (6 - pink)

by 1000 vehicles, we would need to either increase the sharing participation by 30% or increase the AV adoption by 25%. For a similar reduction in vehicle fleet with $taxiCapacity = 6$ we would need to increase sharing participation by approximately 50% or AV adoption by 37%. However, it is noteworthy that for the same level of sharing participation and AV adoption, the fleet size that is required for a scenario with vehicles that have a capacity of 6 is less by approximately 500 taxis, as compared to scenarios with capacity 4.

4.4.2 Analysis of Key Output Parameters

While evaluating the sustainability and usefulness of transportation systems, system operators often have to consider multiple performance indicators such as the inconvenience cause to the riders (increased waiting time or increased travel distance) and the environmental impact.

4.4.2.1 Waiting time and Travel distance increased

When ride sharing systems are concerned, the major consideration for riders is the increased travel distance that would result due to the vehicle deviating from its route. Additionally, since the number of vehicles in ride sharing systems would be less (figure 4.3(c)) there is a possibility that riders would have to wait more time for service.

We can see from figure 4.4, the increase in travel distance for riders is mostly dependent on increasing sharing participation (an increase of 20% in sharing participation results in an average increase in travel distance by approximately 10%). In the case of waiting time though, the trend is not as clear as the increase in travel

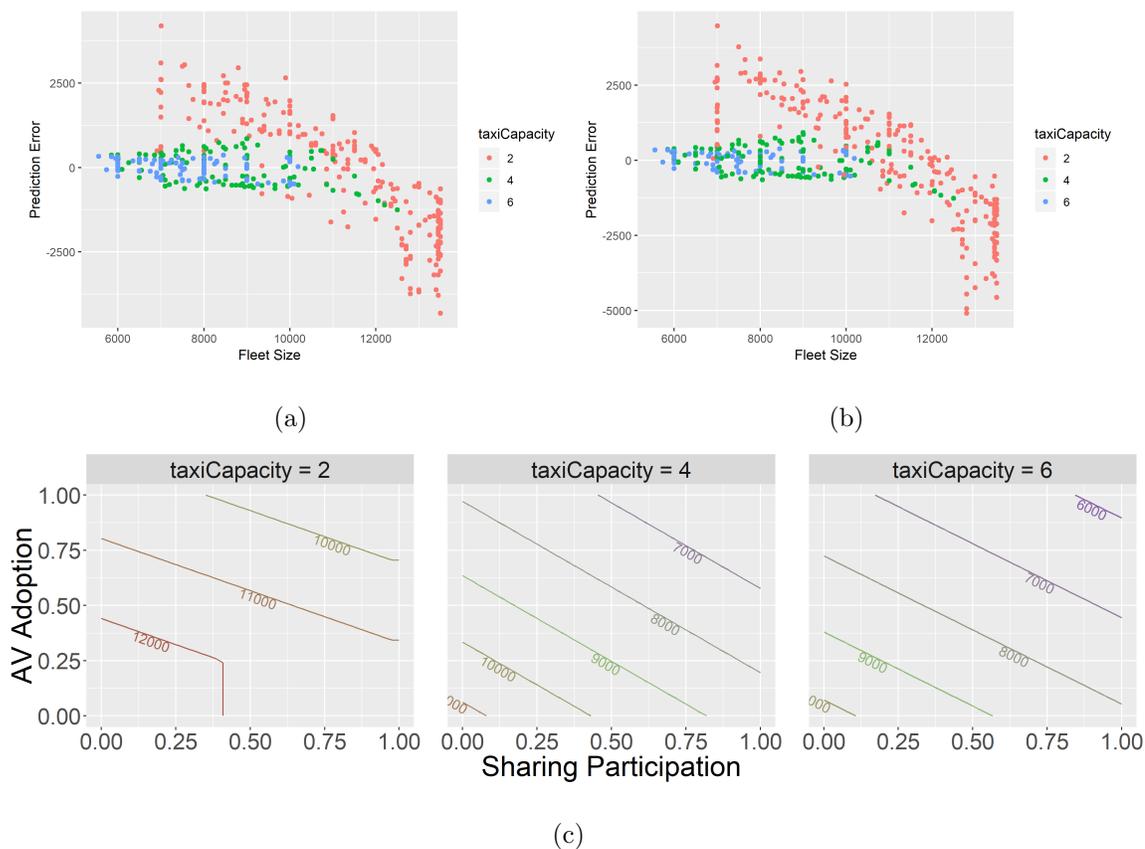


Figure 4.3. (a) Errors in in-sample prediction (a) Cross validation errors (b) The partial dependencies of fleet size with Capacity (panels), sharing participation (x-axis) and fraction of AVs (y-axis), the contours are spaced at 1000 taxis.

distance. Hence, in order to better understand the relation of the waiting time on the input parameters I built a linear regression model of waiting time against all the input parameters. The model (table 4.4) has an R^2 of 0.97, with an out of sample RMSE of 8.807s. Using the partial dependence plots (Figure 4.5) we can further analyze this model.

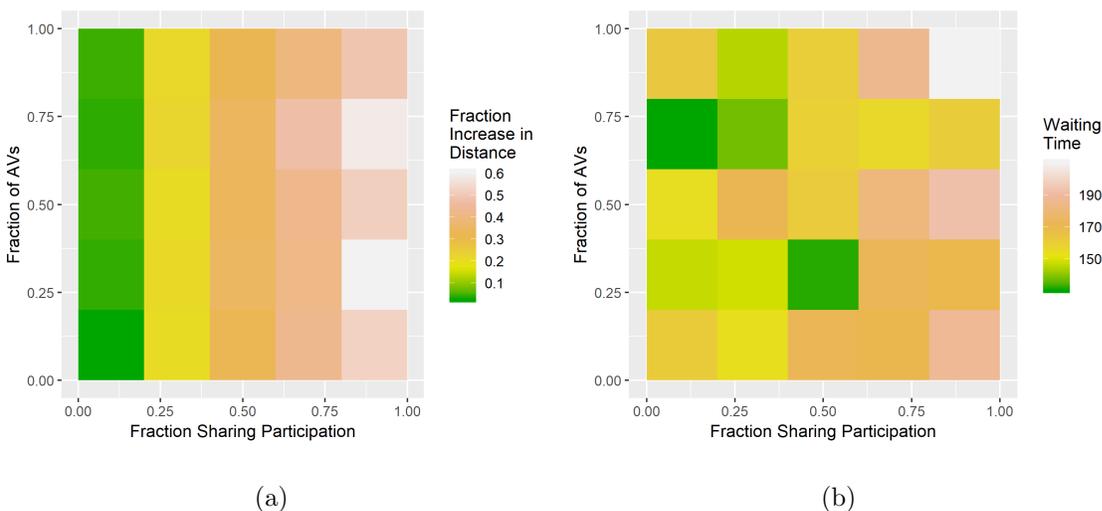


Figure 4.4. Raster plot of (a) average fractional increase in travel distance for riders (b) Waiting time for riders

From figure 4.5(a) and 4.5(b) we see that, an increase in fleet size by 1000 taxis or an increase in AV adoption by 25% can reduce the average system waiting time by approximately 30s in systems with taxi fleets with capacity 4 and 6. In systems with capacity 2 fleets, an increase in 1000 taxis or an increase in AV adoption by 37% reduces the average waiting time by approximately 20s. The reason for this difference between the capacity 2 fleets and the capacity 4/6 fleets is due to the small capacity vehicles being unable to accept certain riders due to their restricted capacity.

The partial dependence of waiting time on the percent sharing though is interesting. From figure 4.5(c) we see that for capacity 2 taxis the waiting time increases for increase in the sharing participation since it is difficult to find shared rides when the capacity is low and riders would need to wait more to find a ride. In cases where

Table 4.4.
Regression model for waiting time $RMSE = 8.81s$ and $R^2 = 0.97$

term	β	Std. Error	t value	p value
<i>(Intercept)</i>	277.5	9.402	29.509	< 2e-16
<i>fleetSize</i>	-0.01253	0.0007462	-16.786	< 2e-16
<i>percentSharing</i>	54.13	11.87	4.558	6.52e-8
<i>percentAV</i>	-45.02	3.061	-14.71	< 2e-16
<i>taxiCapacity4</i>	215.9	20.27	10.653	< 2e-16
<i>taxiCapacity6</i>	215.3	38.92	5.534	5.12e-10
<i>dtMode</i>	33.92	6.239	5.436	8.59e-10
<i>fleetSize : percentSharing</i>	-0.005281	0.0009148	-5.773	1.38e-10
<i>fleetSize : taxiCapacity4</i>	-0.01682	0.001803	-9.33	< 2e-16
<i>fleetSize : taxiCapacity6</i>	-0.01602	0.003847	-4.164	3.7e-7
<i>fleetSize : dtMode</i>	-0.00313	0.0005625	-5.564	4.34e-10
<i>percentSharing : percentAV</i>	6.158	4.102	1.501	0.134
<i>percentSharing : taxiCapacity4</i>	-39.31	5.642	-6.967	1.05e-14
<i>percentSharing : taxiCapacity6</i>	-14.53	8.232	-1.765	0.0782
<i>percentSharing : dtMode</i>	15.12	3.499	4.322	1.87e-6
<i>percentAV : taxiCapacity4</i>	-53.15	5.871	-9.052	< 2e-16
<i>percentAV : taxiCapacity6</i>	-43.33	10.53	-4.113	4.58e-8

the capacity is 4, the waiting time decreases with increase in the percent sharing, as more riders are able to find shared rides when the capacity is 4. However, when the capacity is 6, the percent sharing has no effect on the waiting time, since in

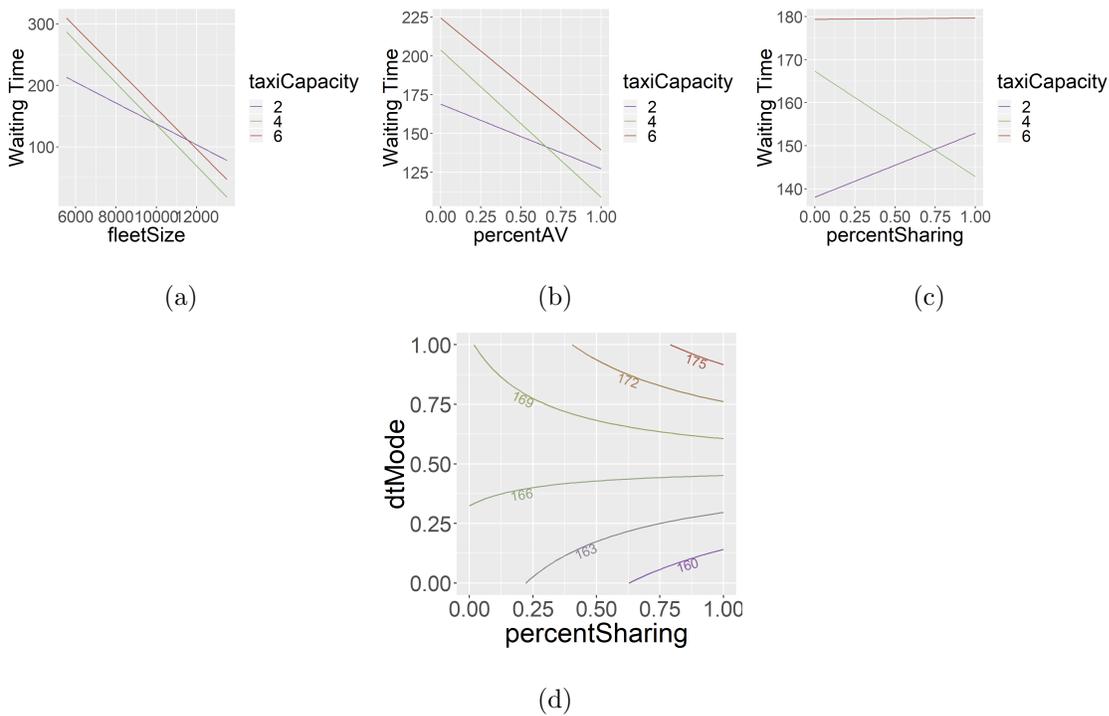


Figure 4.5. Partial dependency plots of waiting time (s) from the linear model (table 4.4) with (a) Fleet Size and Capacity (b) sharing participation and capacity (c) AV adoption and Capacity (d) $dtMode$ and percent sharing. In (d) the waiting time is represented by the colour of the contours in the plot of $dtMode$ and percent sharing.

scenarios where the capacity is 6, the fleet size is generally low. As indicated by the model (table 4.4), fleet size is the most important predictor of waiting time since a larger fleet means there will be more available taxis at a time. Additionally, the $dtMode$ parameter has an important role in determining the average waiting time (figure 4.5(d)). The waiting time increases with increase in the $dtMode$ parameter. For low values of percent sharing the $dtMode$ parameter has a smaller impact than at higher sharing adoption since, the $dtMode$ parameter is only meaningful for riders

who participate in sharing, and when the sharing participation is low the *dtMode* parameter becomes irrelevant.

Table 4.5.

Cluster centers obtained using k-means clustering on *percentSharing* and *percentAV* from the comparable scenarios in Table 4.1

	<i>percentSharing</i>	<i>percentAV</i>
1	0.1184431	0.1497469
2	0.8719863	0.8690239
3	0.7854922	0.2014679
4	0.2458002	0.8088082

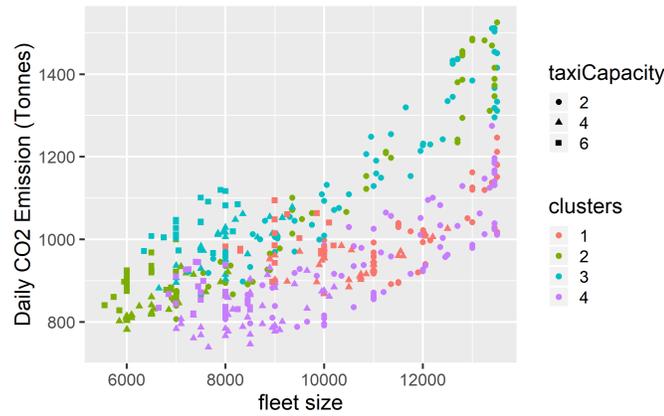


Figure 4.6. CO₂ per day for each scenario with fleet size. The *taxiCapacity* is represented by the shape of the point and the color refers to the cluster centers obtained using k-means clustering on *percentSharing* and *percentAV*. The coordinates of the cluster centers are in Table 4.5

4.4.2.2 CO₂ Emissions

Ride sharing and autonomous vehicles have been associated with lower carbon emissions (Brownell and Kornhauser, 2014; Taiebat et al., 2018; Chen and Kockelman, 2016a; Fagnant and Kockelman, 2015). In order to quantify these environmental impacts, I multiplied the distance travelled by all the vehicles in each equivalent scenario by the CO₂ emitted per mile. I used separate the CO₂ emitted per mile factors for *taxiCapacity* 2,4, and 6 vehicles (same as the Chevrolet Spark (271 g CO₂ per mile), Toyota Corolla (286 g CO₂ per mile) and the Chevrolet Equinox (315 g CO₂ per mile) from EPA (2018)). Also, for the AVs in the model, the total daily emission was reduced by 9% as per the findings of Gawron et al. (2018). Figure 4.6 plots the CO₂ emitted per day for different scenarios against the fleet size. In order to better understand the impact of the input parameters on the environment, I clustered these scenarios by *percentSharing* and *percentAV* (Table 4.5 are the cluster centers). We can see that the most scenarios with the least tailpipe emissions have high percent sharing, low AV adoption and a capacity of 4 since these scenarios have both a low fleet size and low CO₂ emitted per mile. The scenarios where the *taxiCapacity* = 2 has the highest range in the total daily emissions, and mainly depend upon the number of vehicles in the system (since the total distance travelled increases with the number of vehicles). Even though the CO₂ emitted per mile for capacity 2 vehicles is low, since these scenarios have a relatively large number of vehicles and hence a higher distance travelled, they have higher daily CO₂ emissions as compared to capacity 4 vehicles. In contrast, scenarios with capacity =6 have higher emissions per mile but lesser distance travelled and the lowest daily emission for the scenarios with capacity 6 are approximately the same as that with capacity

4. So, if scenarios that have lower carbon emissions are desired, we should prefer fleets of capacity 4 with lower AV adoption and higher ride sharing participation.

4.4.3 Studying the efficiency limitations of ride sharing

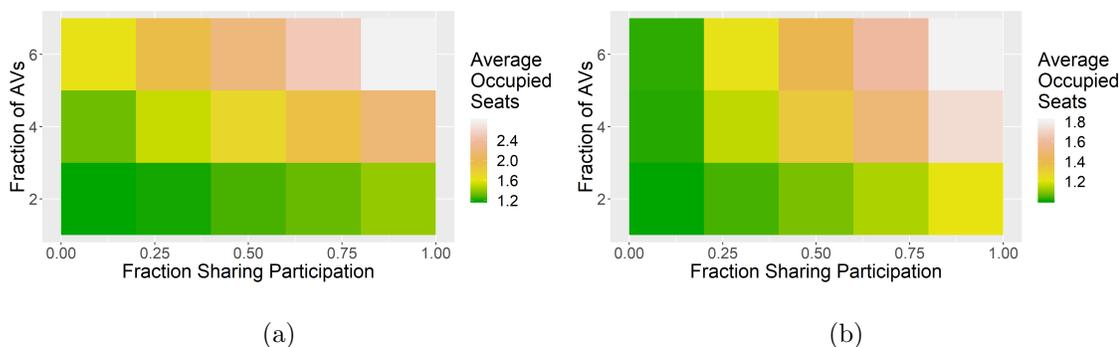


Figure 4.7. Raster plot of (a) Average occupancy for riders (b) Average number of shares for riders

Figure 4.6 shows us that increasing ride sharing in taxi systems has the potential to reduce daily CO₂ emissions, by reducing the total distance travelled by the vehicles in a day. Taxi sharing is able to make these reductions in total travel distance by pooling rides that are along the same path together. We can argue that if a larger number of rides are pooled in the same vehicle together, we can further reduce the travel distance or serve more riders by travelling the same distance. However, pooling more rides together may not always be possible for various reasons (for example, capacity limitations of vehicles, sharing tolerance of riders, and not having enough riders sharing a similar route). Using the comparable models developed in Section 4.4, we can plot the changes in capacity for the taxis in the system.

From Figure 4.7(a) we can see that in scenarios where the taxi capacity is set to 4 or 6, the taxis are unable to reach the maximum capacity available to them using ride sharing, even though increasing ride sharing permits vehicles to have more riders. In order to investigate the cause of taxis being unable to reach the maximum capacity, I studied the status of the taxis from each scenario that I ran of the PP-SAEV, and recorded events at which the taxis are prevented from sharing by recording the sharing stoppages using algorithm 7 in Chapter 2. The different statuses that the taxis can be in are :

Table 4.6.

Sample taxi status table for a scenario with 3000 taxis. Each row sums to the total number of taxis in the system, in this case 3000

Time (s)	Outs- hift	Idle	Shar- ing	LTole- rance	LCapa- city	LShift- Out	LTol- Cap	LShift- Cap	LTol- Shift	LAll3
10	1500	200	1000	286	10	0	4	0	0	0
20	1450	230	988	318	12	1	1	0	0	0
30	1455	220	1100	211	11	0	2	0	1	0
40	1523	231	945	286	8	0	6	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- Outshift: The taxi is not accepting any new passengers as it is not in shift (meaningful for non-autonomous vehicles). Shifts are decided as per the shifting schedules presented in Section 2.6.

- Idle: Not currently serving any customers and willing to accept their first ride. (Status = 1 in Algorithm 3)
- Sharing: Currently serving passengers and are accepting additional passengers to share rides. This status occurs only when all the riders in the taxi can accept shared rides, their deviation tolerance limit has not been reached, the taxis are not at maximum capacity, and the taxis are not scheduled to go out of shift at the termination of the current ride. (*holdCode* = {*F, F, F, F*} in Algorithm 7)
- LTolerance: Currently serving passengers and not accepting additional passengers to share rides since not all the riders in the taxi can accept shared rides since the riders deviation tolerance limit has been reached either due to sharing or if they were non-sharing riders. (*holdCode* = {*T, F, F, F*} in Algorithm 7)
- LCapacity: Currently serving passengers and not accepting additional passengers to share rides since the taxi is at maximum capacity. (*holdCode* = {*F, T, F, F*} in Algorithm 7)
- LShiftOut: Currently serving passengers and not accepting additional passengers to share rides since the taxi is scheduled to shift out at the termination of the current ride. (*holdCode* = {*F, F, F, T*} in Algorithm 7)
- LTolCap: Currently serving passengers and not accepting additional passengers to share rides since not all the riders in the taxi can accept shared rides since the riders deviation tolerance limit has been reached and the taxi is at maximum capacity. Ideally, in order to use the maximum capacity of the taxi each time we would prefer that all of the taxis would eventually reach this status while serving riders. (*holdCode* = {*T, T, F, F*} in Algorithm 7)

- LShiftCap: Currently serving passengers and not accepting additional passengers to share rides since the taxi is at maximum capacity and at the termination of the ride the taxi is scheduled to go out of shift. ($holdCode = \{F, T, F, T\}$ in Algorithm 7)
- LTolShift: Currently serving passengers and not accepting additional passengers to share rides since not all the riders in the taxi can accept shared rides since the riders deviation tolerance limit has been reached and at the termination of the ride the taxi is scheduled to go out of shift. ($holdCode = \{F, T, F, T\}$ in Algorithm 7)
- LAll3: Currently serving passengers and not accepting additional passengers to share rides since not all the riders in the taxi can accept shared rides since the riders deviation tolerance limit has been reached, the maximum capacity of the taxi has been reached, and at the termination of the ride the taxi is scheduled to go out of shift. ($holdCode = \{T, T, F, T\}$ in Algorithm 7)

In order to build a data set of the actions of the taxis at a given time, I collected such status updates from each taxi at 10 second intervals in a data set similar to Table 4.6. Then, since the data in the table reflects a composition where in the counts of each taxi in a particular status is recorded, (and the total of each row is the number of taxis in the system), I used compositional data analysis techniques (Filzmoser, 2018), to first find the mean composition of the taxi status in every simulation during the peak hours of the day, and then I clustered the data by first transforming these compositions to their pivot coordinates using Equation 4.4 (Filzmoser, 2018) and then using hierarchical clustering on the data to identify groups of scenarios which have similar number of taxis in each state on an average during the evening peak demand period (Figure 2.4(a)).

$$z_p = \sqrt{\frac{D-p}{D-p-1}} \ln \frac{x_p}{\sqrt[D-p]{\prod_{k=p+1}^D x_k}} \quad (4.4)$$

In Equation 4.4 z_p is the $p \in \{1 \dots D-1\}$ pivot coordinate, D is the number of components in the compositional data, and x_i is the i -th component of the compositional data. This pivot coordinate transformation is needed since compositional data exists in a space defined by the D dimensional simplex S^D (also known as Aitchison geometry) defined by its coordinates, and the transformation given by Equation 4.4 maps these compositional coordinates from S^D to \mathbb{R}^{D-1} . In the S^D space, standard distance measures (for example L_2 norm) do not represent the distance between the two points, and data analysis techniques (such as clustering) would not work as desired. However, by mapping the compositional data to \mathbb{R}^{D-1} we can use data analysis techniques like clustering to interpret our data. Figure 4.8 and 4.9 shows the cluster assignment using hierarchical clustering in pivot coordinates and in terms of the input parameters respectively. Using Figure 4.9, we see that Cluster 1 consists of scenarios that have 100% AV adoption. As a result of having 100% AV adoption from the scatter plot of *fleetSize* with *percentSharing* we see that these scenarios give the lowest number of vehicles that are needed to achieve the same service level of 0.86. Cluster 2 mostly consists of scenarios with a high ride sharing adoption (above 40%), and in the case where *taxiCapacity* = 2, all scenarios with ride sharing adoption greater than 0. Cluster 3 contains all scenarios with no ride sharing. Cluster 4 contains all scenarios with 100% ride sharing adoption and *taxiCapacity* = 2. Finally, cluster 5 consists of those scenarios with low ride sharing adoption and AV adoption less than 0.6. Further analysis of the taxi states within these clusters can help us understand how we can increase the efficiency of the system in each scenario.

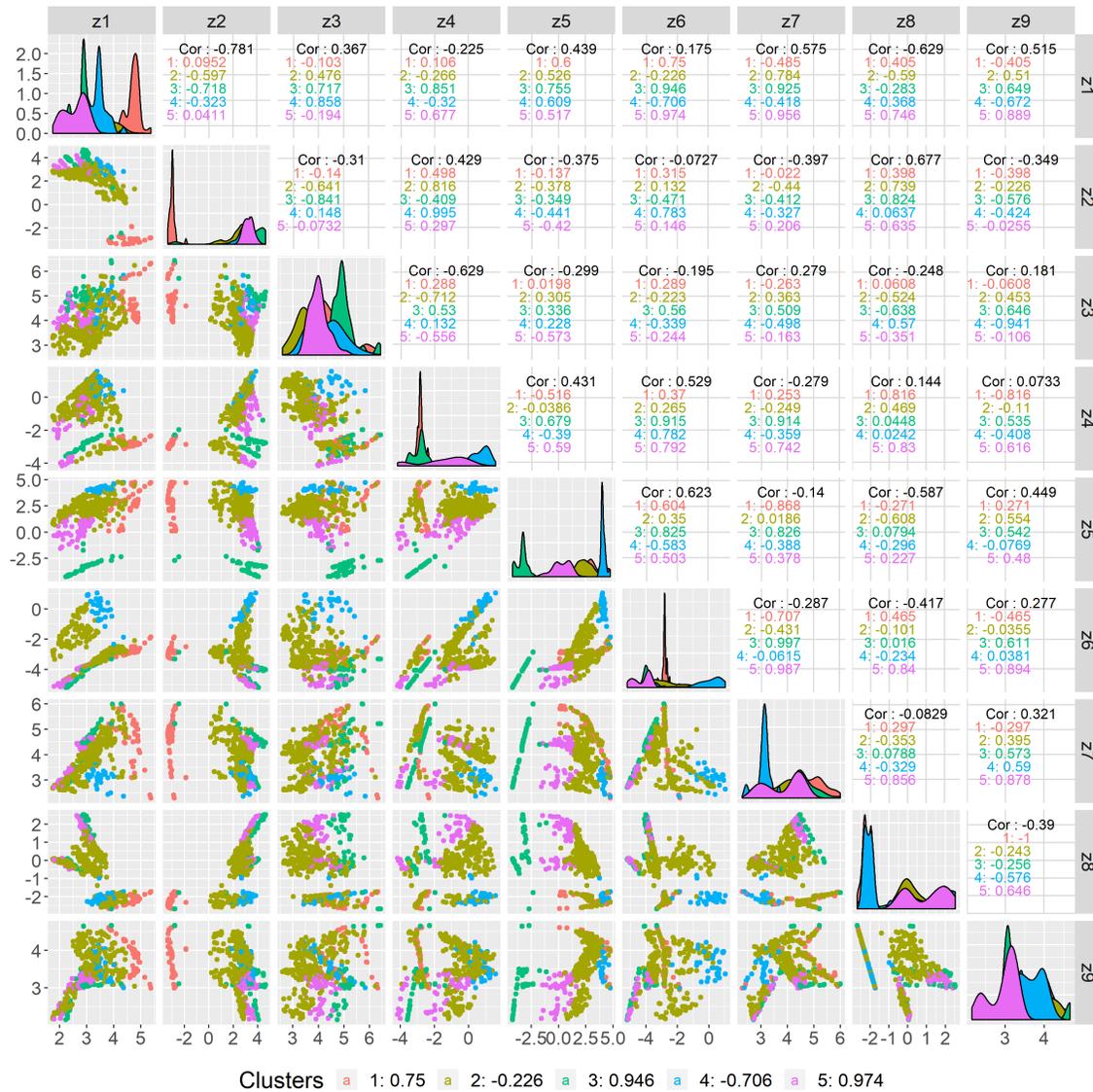


Figure 4.8. Scatter plot matrix of the pivot coordinates. The colours represent the cluster that the scenario belongs to based on the taxi status data (Table 4.6)

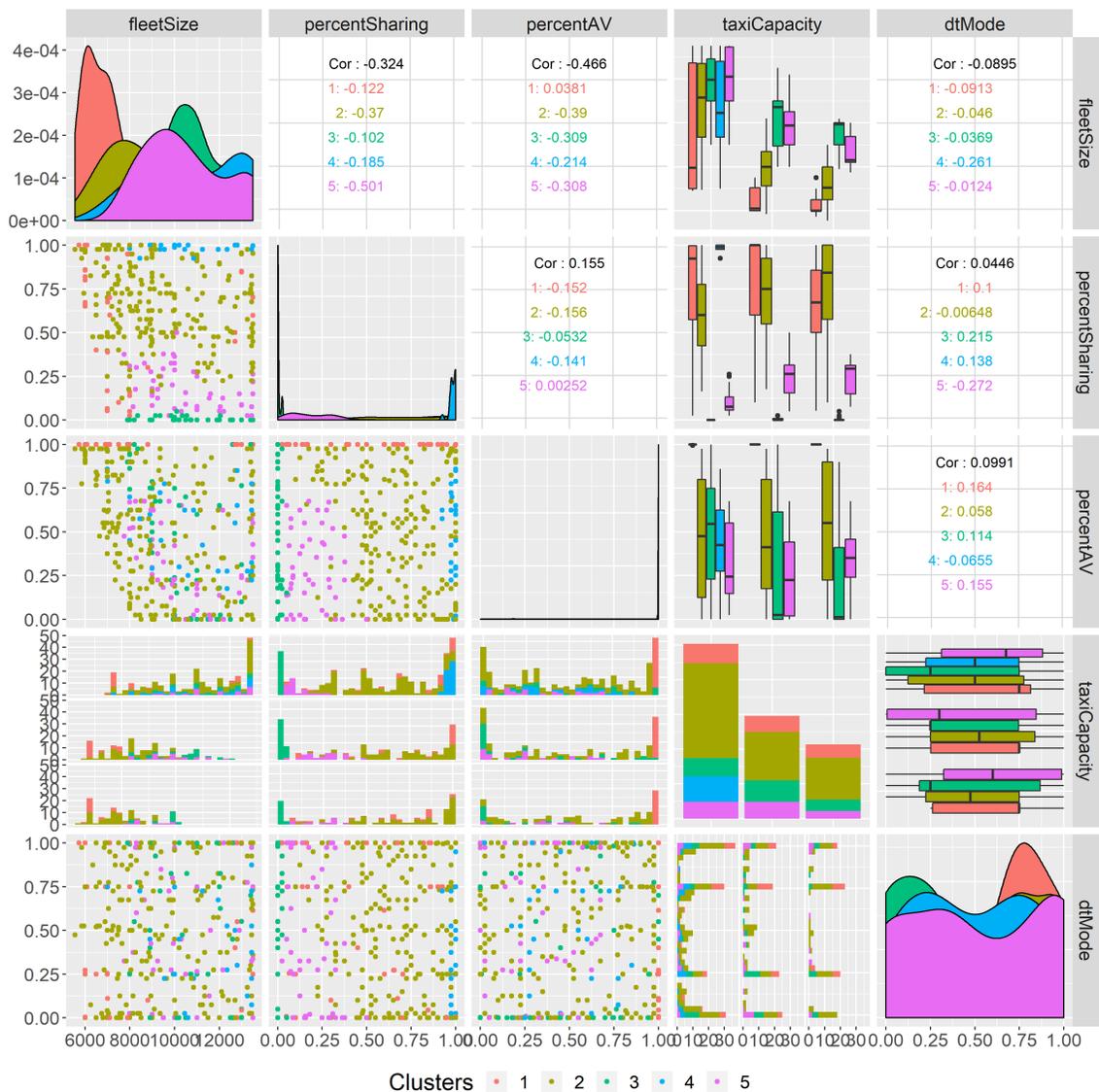


Figure 4.9. Scatter plot matrix of input data. The colours represent the cluster that the scenario belongs to based on the taxi status data (Table 4.6)

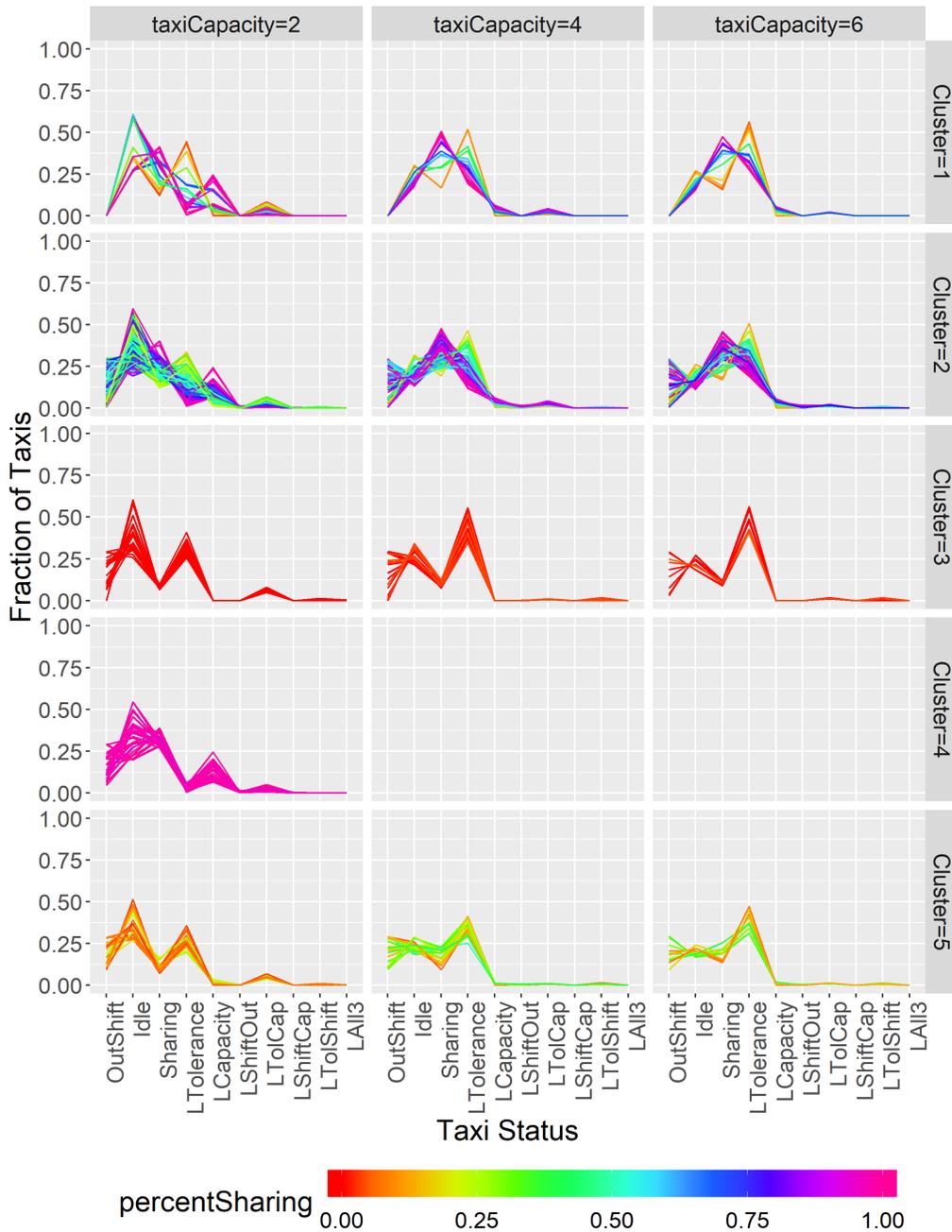


Figure 4.10. Composition plot for the hierarchical clustering on the pivot coordinates of the central taxi status composition (Figure 4.8), in original coordinates colored by *percentSharing*. The y axis is the fraction of the taxis in each status (x-axis) for each level of *taxiCapacity* (column) and each cluster (row). Each path line of the composition plot represents a single scenario

For each of these clusters, and for $taxiCapacity = \{2, 4, 6\}$ I plotted a composition plot (Figure 4.10) in order to visualize the status of the taxis within each cluster. In each composition plot, each scenario is represented using a path line connecting the fraction of taxis (y-axis) for each of the status codes mentioned in Table 4.6. The plots can be used to identify trends within each cluster to make inferences and recommendations for system improvements.

From each of the composition plots a few general trends can be observed. First, for all the scenarios where $taxiCapacity = 2$ the fraction of taxis in Idle state are higher than similar scenarios in $taxiCapacity = 3, 4$. Additionally, even in cases where the fraction of Idle taxis are not high (for example a few scenarios in cluster 2 and 4) most of the taxis are available to share (Sharing) but find no riders to share with due to their limited capacity (for example, if a taxi of capacity 2 has one occupied seat, it can only accept ride shares where the group size is 1). We also observe that the fraction of taxis who cannot share due to capacity limitations (LCapacity) is much higher for $taxiCapacity = 2$ as compared to those scenarios where $taxiCapacity = 4, 6$. Additionally, in cases with high sharing adoption (for example clusters 2 and 4) on an average the taxis do not reach their sharing tolerance limit (LTolerance). This tells us that taxi fleets with $taxiCapacity = 2$ are likely to be inefficient at serving passengers. Second, in all of the clusters, the trends of the scenarios with capacity 4 and capacity 6 are very similar to each other. In each of these scenarios the vehicles rarely reach their capacity limit (LCapacity). In the cases where the sharing participation is high (Cluster 2) the fraction of taxis which do not share due to the riders tolerance is slightly higher for scenarios with capacity 6. This indicates that capacity 6 taxis do have a marginal benefit at high levels of sharing adoption. Since, taxi fleets with capacity 6 emit more CO₂ emissions, this

marginal benefit may not be enough to convince policy makers to use capacity 6 taxis.

The cluster 1 scenarios all have 100% AV adoption. From Figure 4.10 we see that in cases where the sharing participation is low, most taxis reach their sharing tolerance limit (high LTolerance), however in scenarios with high sharing participation, most of the taxis are still open to accepting new shares. This means that when the sharing participation is high, there may be a surplus of vehicles available for sharing, but there may not be enough riders on common routes to share rides. Potentially, the system could serve even more riders with the same number of vehicles if they were travelling on the same route. This suggests that if the system operator induces demand along the routes of the taxis by lowering prices, the same fleet of taxis would still be able to service the additional demand. However, in these scenarios the fraction of idle taxis is low in these scenarios, the taxi fleet may not be able to serve additional non-shared rides. Hence, care will have to be taken to ensure that the system operators design policies in these cases to induce ride sharing demand but not additional demand for non-shared rides. Cluster 2 contains a large number of scenarios with varying levels of AV adoption and high ride sharing adoption. Since the ride sharing adoption is mostly high in these scenarios, we see that on average a large number of taxis are able to accept new shares. Hence, these scenarios will also be able to accept new induced demand that is willing to share rides. On the other hand, when the sharing participation is low (cluster 5) not many vehicles reach their maximum capacity, very few vehicles are available for sharing and most of the vehicles are limited from sharing from the tolerances of riders. This can tell us that there must be a minimum level of sharing participation required to have a successful ride sharing participation. If we compare the scenarios in cluster 2 with those in

cluster 5, we see that most of the scenarios where sharing seems to be successful have sharing participation above 30%.

4.4.3.1 Where can ride sharing be induced?

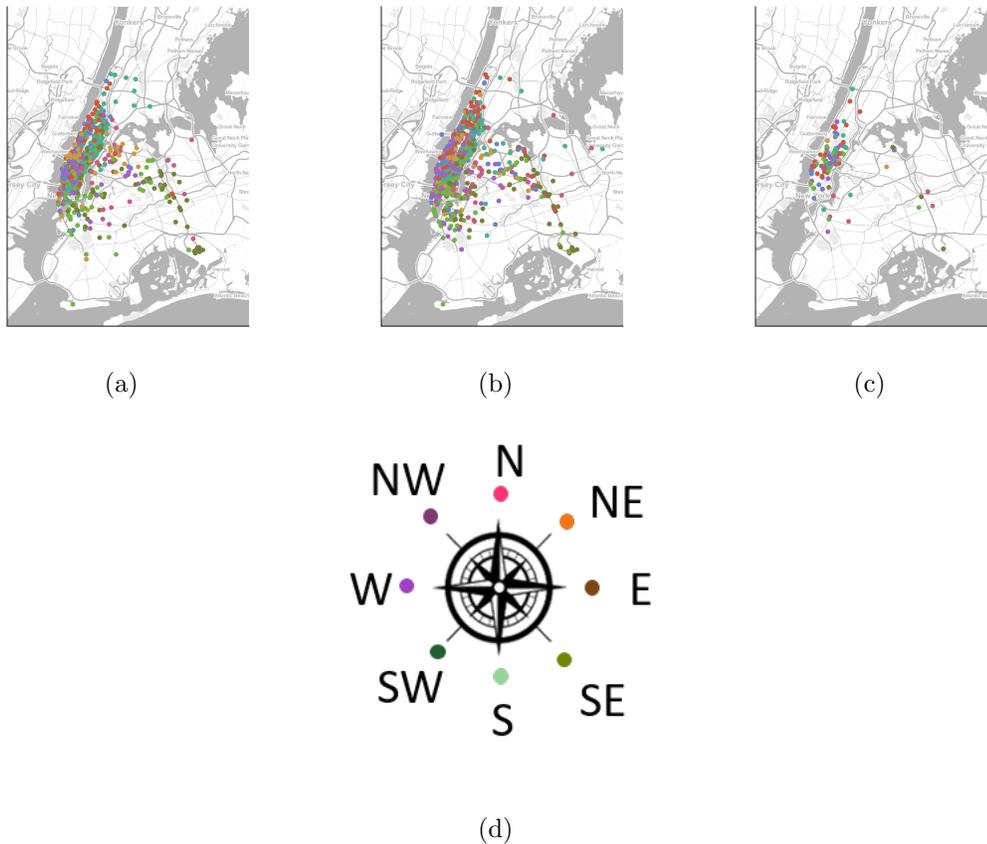


Figure 4.11. Location and direction of taxis that can accept shared rides

From Figure 4.10, it can be observed that scenarios in Cluster 1 and 2 which have high RS adoption can accept additional ride shares, while cluster 5 may not

be able to accept additional ride shares. In order to know where can new rides be introduced such that they may be served in using ride shares, I plotted the locations of the taxis that are in Sharing status and the direction in which these taxis are moving at 6:00pm (the time of the day at which there are maximum number of pickups) in Figure 4.11 for one scenario each from Cluster 1 (55), cluster 2 (36) and cluster 5 (154) (see Table C.1). We see in each of the scenario, the taxis that are in the Sharing status are moving from regions of low demand density (Brooklyn and Queens) to regions of high demand density (Manhattan). However, we see that the number of taxis in sharing mode in the case of the scenario from cluster 5 is much less. This suggests that if ride sharing demand were induced with pick up location at a low demand region and moving towards high demand regions the scenarios may be able to serve this additional demand with the same input parameters for scenarios in Cluster 1 and 2. However, scenarios in cluster 5 may not be able to accommodate large amounts of this additional induced demand.

4.5 Conclusion

In this chapter, I used meta model based simulation optimization to find comparable agent-based modeling scenarios which serve the same number of riders, while varying 4 continuous parameters (*percentAV*, *percentSharing*, *fleetSize*, *dtMode*) and one discrete parameter (*taxiCapacity*) of the PP-SAEV. Using these equivalent scenarios, we were able to build models to predict the appropriate fleet size for scenarios with a certain AV and ride sharing adoption for taxi fleets of capacity 4 and 6. The prediction models were not able to make accurate predictions for *taxiCapacity* = 2 since for a single setting of *percentAV*, *percentSharing*, *dtMode* there may be multiple fleet sizes that can serve the same number of riders. By using

clustering, we identified that scenarios with high sharing participation and low AV adoption with taxi fleets of capacity 4 had the lowest CO₂ emissions.

Then, by studying the proportion of taxi status during the peak demand period, and using hierarchical clustering, it was found that taxis of capacity 2 are not sufficient for a good taxi sharing program. On the other hand, taxis with capacity 6 are only marginally better than taxis with capacity 4 for high ride sharing participation. However, since vehicles with capacity 6 have higher CO₂ emissions, taxis with capacity 6 are not recommended for use in ride sharing. Additionally, a minimum sharing participation of 30% is required for a successful ride sharing program. Finally, by studying the spatial patterns of the taxis, I inferred that scenarios with high sharing participation may be able to accept additional ride sharing demand in the direction of low demand density to high demand density.

It is notable that this chapter did not consider different adoptions of electric vehicles. In order to extend this analysis to electric vehicles, at each step charging stations would have to be sited thus making such an extension not trivial. An experimental design study that incorporates the analysis of electric vehicles and performs similar analysis could provide an understanding of the combined effects of EV, AV and RS adoption changes.

5. UNDERSTANDING THE IMPACT OF HETEROGENEOUS RIDER PREFERENCES ON A SHARED AUTONOMOUS VEHICLE SYSTEM

5.1 Introduction

Chapters 3 and 4 show that ride sharing is a potentially lucrative method to curb carbon emissions from the transportation sector, reduce the number of vehicles on road, and increase service levels for cities (Ma et al., 2015; Alonso-Mora et al., 2017). Additionally, with the introduction of autonomous vehicles, the number of vehicles on road needed to serve the same number of passengers are expected to reduce even further. However the scenarios evaluated in Chapters 3 and 4 consider restrictive ride sharing rules. This chapter aims to loosen this restriction by allowing riders to have heterogeneous sharing preferences. The presence of such heterogeneous sharing preferences is well studied in the literature as documented in Section 1.1.2.

Not all riders may have a binary sharing preference (either to accept a share or not). Some cost sensitive riders may stay in areas with limited public transportation and would prefer to use ride sharing but would accept even a non-shared one to reach their destination if they cannot get a shared ride. Similarly, those riders who would not like to share rides due to concerns with safety, may be forced to use shared rides if a non-shared ride is unavailable (for example, when all the available vehicles at the moment are shared ones and it will take much longer to wait for a non-shared one). Other riders may be completely flexible and may choose to use either shared or

non-shared rides since they have no such restrictions and only want to get from their pick up location to their destination. Such heterogeneous sharing preferences could impact the performance of ride sharing systems. Additionally, the compositions of the rider preferences in the system may impact the service quality of a group of riders with specific sharing preferences, based on the proportion of the other rider types in the system. Understanding the impacts of these sharing preferences on system service quality as well as individual service quality may help system operators design policies that meet both system service targets and cater to a variety of rider types. By not considering these heterogeneous preferences, models could be limited in the capability to fully capture the system dynamics and make system level performance inferences.

The goal of this chapter is to address the question: how does the performance of a ride sharing system change with different mixes of heterogeneous rider preferences? To the best of my knowledge, no existing literature has evaluated a ride sharing system in which the riders in the system are modeled with varied sharing preferences. Additionally, no study has been able to quantify the result of varying the proportion of these rider types in the system. The results from this study could aid system planners to design incentives to promote riders to adopt a certain sharing preference to benefit the system (for example, increasing system service level or reducing average waiting time for customers). By varying the *riderTypes* parameter of the PP-SAEV, these heterogeneous preferences can be studied (Table 5.1). Then the output from the running the PP-SAEV with the parameter settings in Table 5.1 can be used to study the system service level as well as the service level for each rider type. To do this, I use the PP-SAEV model that was developed in Chapter 3 to study the impact of these heterogeneous sharing preferences of riders. Algorithm 2 is used to classify the searching preferences of the riders into one of the 5 rider-group types 1)

Table 5.1.
Simulation scenarios for sharing preference

Parameter	Setting
<i>riderTypes</i>	{1, 2, 3, 4, 5}, proportions of rider types varied using mixture experiments
<i>fleetSize</i>	{4000, 6000}.
<i>AVPercent</i>	{100%}
<i>taxiCapacity</i>	{4}
<i>dtMode</i>	{0, 1}

unwilling to share a ride, 2) prefer not to share, 3) indifferent to sharing, 4) prefer to share, 5) will only accept a shared ride. Then, I use mixture experiment design to set parameters for the PP-SAEV so that we can study the impact of each of these rider types on the SAV system. The parameters varied in this study are as per Table 5.1.

The rest of this chapter is divided into 3 sections, Section 5.2 explains the methodology that we used to vary the parameters as per Table 5.1, Section 5.3 illustrates the use of the method on a case study of New York City Taxis and Section 5.4 summarizes my findings from the case study.

5.2 Parameter Variation using Mixture Experiments

In order to estimate how changing the proportions of each *riderType* impacts the system, I ran the PP-SAEV model with different rider mixes. In my model, the proportion of rider types is represented by a vector $X = \{x_i : i \in 1 \dots 5\}$ x_i

for each i th *riderType*. Additionally, I also wanted to study the impact that the *dtMode* parameter has on the system for different proportions of *riderTypes* and for different *fleetSize*. Since, in each case $\sum_{i=1}^5 x_i = 1$, I used mixture designs to test different proportions of rider types in the system.

Mixture models are traditionally used in chemical industries where a product is commonly a combination of two or more components in a set proportion and the manufacturer wishes to study the effect of these compositions (Cornell, 1990). By using mixture experiments, the proportions of each of the rider types in the system can be varied. These proportions are varied in such a way that the effect of each of these rider types is understood, while running only those experiments that are needed to develop the understanding of the surface (Cornell, 1990). Since there are 5 rider types, and I would like to understand the impact of having different proportions of each rider type, a Simplex Lattice Design (Cornell, 1990) of degree 4 for the 5 rider types ($SLD(5, 4)$) with additional axial points to augment the design was used. The resulting design is able to distinguish interaction effects for up to 4 degrees. A degree 5 design is not used, because a degree 4 design is sufficient to represent the surface at interior design points, and the analysis showed that the models required polynomials of degree 2 only. Also, a degree 5 design would require additional 200 runs to complete, which is 2/3rd of the number of runs than a degree 4 design. Since it is possible that different fleet sizes (*fleetSize*), different deviation tolerance distributions (*dtMode* represents the mode of the triangular deviation tolerance distribution), and the different mix of riders could affect the system differently, the $SLD(5, 4)$ design is crossed with a 2^2 factorial design 2^2 factorial design with the process variables *fleetSize* and *dtMode*. The resulting design has 304 runs (Table D.1).

In order to make inferences regarding a particular system performance parameter (for example, service level and waiting time), linear regression models are built for that performance parameter with the composition variables and the process variables (Table 5.2). In order to construct the model, the quadratic Scheffe model crossed by a linear model with linear interactions for the process variables is used (Equation 5.1) as recommended by Cornell (1990).

$$\begin{aligned}
y = & \sum_{i=1}^5 \beta_i^{(0)} x_i + \sum_{i=1}^4 \sum_{j=i+1}^5 \beta_{ij}^{(0)} x_i x_j + \sum_{i=1}^5 \sum_{l=1}^2 \beta_{il}^{(1)} x_i z_l + \\
& \sum_{i=1}^4 \sum_{j=i+1}^5 \sum_{l=1}^2 \beta_{ijl}^{(1)} x_i x_j z_l + \sum_{i=1}^5 \beta_{ilm}^{(2)} x_i z_1 z_2 + \\
& \sum_{i=1}^5 \sum_{j=i+1}^5 \beta_{ijlm}^{(2)} x_i x_j z_1 z_2 + \epsilon
\end{aligned} \tag{5.1}$$

Where x_i is the i th mixture component; $z_1 = fleetSize$; $z_2 = dtMode$ are the process variables and y is the response variable. Then, the terms in the model that were not significant at 5% level are eliminated using step-wise regression. In order to evaluate the fit of the regression model, the root mean squared error (RMSE) and the corrected $R - sq^1$ is used.

Since the mixture has 5 factors, it can be difficult to visualize the results and deliver meaningful insights. In mixture experiments, when two or more mixture components produce the similar main effects and interaction effects (the coefficients of the regression terms involving the mixture components does not significantly differ), the components can be combined and analyzed them as a single component

¹For models fitted using Scheffe polynomials, since there is no constant term, the regular goodness of fit statistic $R^2 = 1 - \frac{SSE}{SST}$ where SSE is the error sum of squares and SST is the total sum of squares gives an inflated value. Cornell (2011) suggests to use the corrected $R^2 = 1 - \frac{SSE/(N-p)}{SST/(N-1)}$, where N is the number of observations, p is the number of terms in the model. All the R^2 values reported in this paper are the corrected R^2

(Cornell, 1990) to simplify the analysis. In this chapter, this model simplification technique is used to aid my discussion of the effect of the different rider types on the model. The notation $x_{ij} = x_i + x_j$ is used to indicate this simplification.

5.3 Case Study Results

In order to demonstrate the inferences that the PP-SAEV model can provide, the case study described in Section 2.5 is used. The factor levels that were run are given in Table 5.2 (a detailed list of runs is listed in Table D.1). The fleet size variable was set as per the findings of Chapter 3, where it was shown that a fleet of 5500 AVs with ride sharing could serve the same number of riders as the present day case (13500 non-autonomous vehicles with no ride sharing). By setting the $fleetSize = \{4000, 6000\}$ I tested the system performance in an under supplied (4000) and an over supplied (6000) cases. The $dtMode$ parameter was simply set to the highest and lowest values the variable could take. These settings help us evaluate the system in extreme cases of these parameter settings. Then the system level impacts of heterogeneous rider types are analyzed, which can be useful for system planners and policy makers, in Section 5.3.1. The impact on each rider type in Section 5.3.2 are also analyzed. Such analysis could help policy makers make decisions on parity among different rider types.

5.3.1 System level implications

In order to understand the system level implications of having different heterogeneous mixes of rider sharing preferences in the system, we examined the service

Table 5.2.

Simulation scenarios to study heterogeneous sharing preferences in ride sharing

Parameter	Value Settings	Design
x_1, \dots, x_5	{0, 0.1, 0.2, 0.25, 0.5, 0.6, 0.75, 1}	$SLD(5, 4)$
<i>fleetSize</i>	{4000, 6000}	2^2 full factorial
<i>dtMode</i>	{0, 1}	2^2 full factorial

level, waiting time, the percent of rides that were shared, the occupied capacity of the vehicles, and the environmental impact at a system level.

5.3.1.1 Service Level

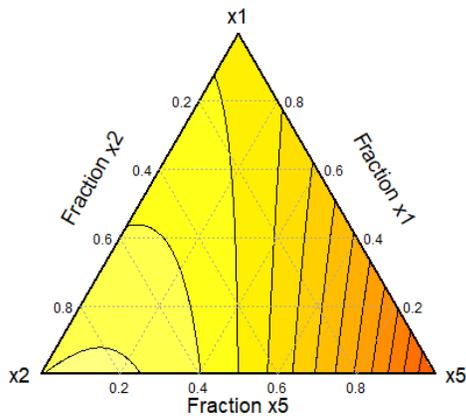
For a taxi system, the fraction of riders that are served by the system during the peak demand period is an important metric to understand its efficiency. Hence, the changes in the fraction of riders that were served by the system between 6:00pm and midnight (this is the time of the day that NYC has the highest taxi ridership (NYC DOT, 2014)) were modeled using regression based on Sheffe polynomials (Table 5.3). It was initially found that rider type 3 and 4 do not differ significantly in their main effects and their interactions and hence, I combined the responses for these two rider types. Figure 5.1 visualizes the predicted service level from this model. In each surface plot, such as Figure 5.1, the corners of the triangles represent the case where the component at the corner is set to one and the other components of the mixture are set to 0. A point on the edge represents a mixture of two components while the others are set to 0. A point in the interior of the triangle represents a mixture of 3

components while the other components that are not in the plot are set to 0. The color of the triangle is the predicted response variable value (which is the service level in the case of Figure 5.1)

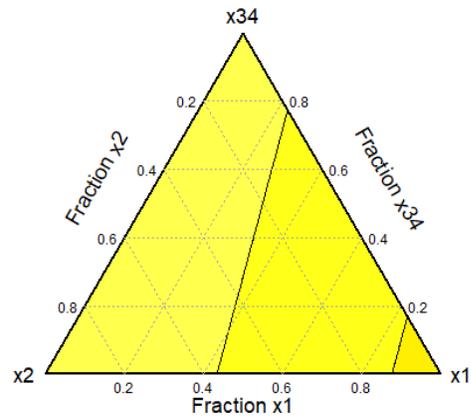
Table 5.3.
Regression model for served PC. $RMSE = 0.057$ and $R^2 = 0.86$

term	β	Std. error	t value	p value
x_1	1.24e-01	2.15e-02	5.74	2.34e-08
x_2	2.36e-01	2.15e-02	10.979	< 2e-16
x_{34}	2.06e-01	1.89e-02	10.908	< 2e-16
x_5	-3.72e-01	2.69e-02	-13.832	< 2e-16
$dtMode$	2.38e-02	6.58e-03	3.614	0.000354
$fleetSize$	9.82e-05	3.29e-06	29.822	< 2e-16
$x_1 : x_5$	4.29e-01	8.07e-02	5.315	2.11e-07
$x_2 : x_5$	8.19e-01	8.07e-02	10.142	< 2e-16
$x_{34} : x_5$	1.27e+00	6.15e-02	20.643	< 2e-16

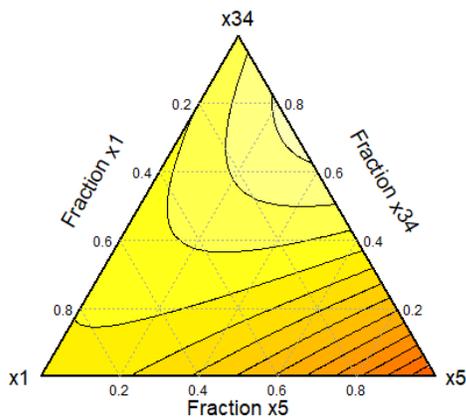
Figure 5.1 shows that the highest service level reached when the $fleetSize = 6000$ and the $dtMode = 1$ is when either rider type 3 or 4 is approximately 70% with 30% of rider type 5 in the model. This result shows that higher service levels are reached when all of the riders in the system are open to sharing, with most willing to accept non-shared rides and a few (30%) unwilling to accept a non-shared ride. While it is intuitive that a higher percent of sharing preferred/ indifferent riders would increase the service level of the system, it is interesting that, in order to achieve a higher system service level, a small number of riders who accept only shared rides is



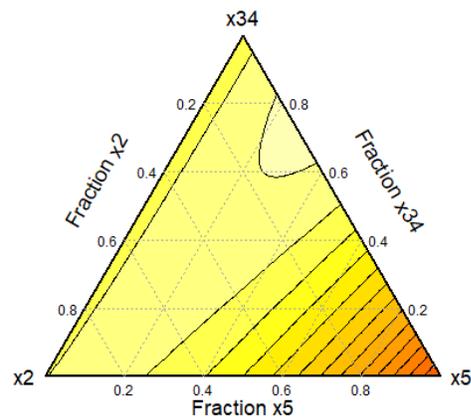
(a)



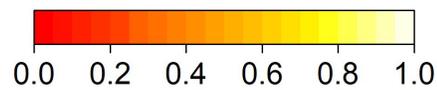
(b)



(c)



(d)



(e)

Figure 5.1. Surface plots of the predicted service level using the model represented by 5.3 at different compositions of rider types at $fleetSize = 6000$ and $dtMode = 1$

desired. Because, rider type 5 stays in the system longer as compared to the other rider types, and in a system with a limited number of taxis, riders are more likely to find shared rides as most of the vehicles will be occupied by other riders that can accept ride shares.

The service level depends linearly upon the process variables *fleetSize* and *dtMode* as seen in Table 5.3. By removing 2000 taxis from the system, the number of riders served by the system can decrease by 20%. Also, if the riders in the system are less tolerant to ride sharing (*dtMode* = 0), the service level could decrease by 2%. These results indicate that it is more important for a ride sharing system to have riders that are at-least open to sharing (as indicated by Figure 5.1) than having riders with a higher tolerance for deviating from their original route.

We can also see from Figure 5.1 that the least system service level is reached when no riders in the system will accept a non-shared ride (100% rider type 5), since there will be no riders to initiate a shared ride if no one is willing to accept a non-shared ride. Thus, policies and incentives should not be set in a way to encourage all riders in a system to only accept shared rides. On the other hand, if there are no riders in the system that are willing to participate in ride sharing, the service level will be reduced by approximately 20%. Hence, if the goal of the ride sharing service provider is to maximize service level, incentives should be set in such a way to bring the proportion of riders in type 3 or 4 to 70% and type 5 to 30%.

5.3.1.2 Sharing efficiency

In a ride sharing system, the efficiency of ride sharing can be measured by studying the percent of shared rides and the average vehicle occupancy during the peak demand periods. In order to do this, I built regression models for the percent of

shared rides (Table E.1) and the average taxi capacity (Table E.2) between 6:00pm and midnight, using the methodology outlined in Section 5.2 and examined the predictions from the model.

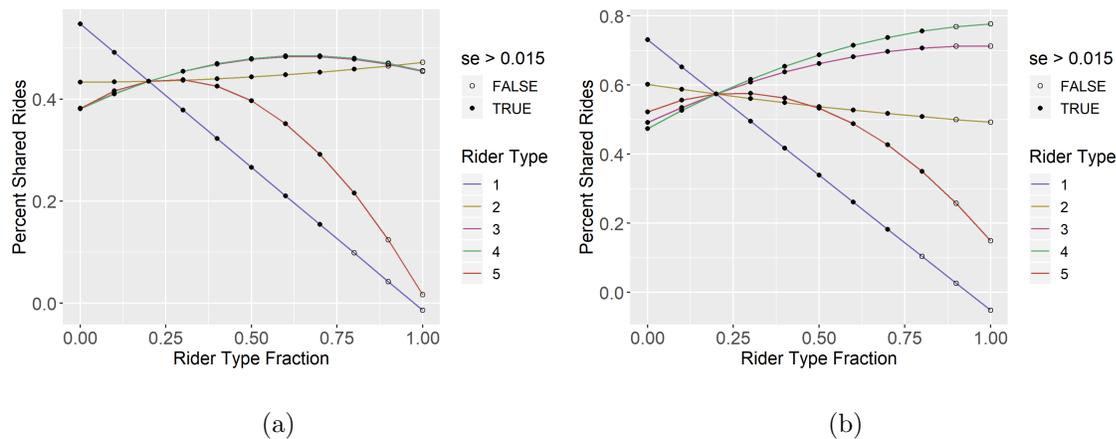


Figure 5.2. Effect plots of the predicted percent shared rides using the model represented by Table E.1 at different compositions of rider types at (a) $fleetSize = 4000$ and $dtMode = 0$ (b) $fleetSize = 6000$ and $dtMode = 1$. I have plotted the points with large standard error of prediction ($se > 0.015$) using hollow circles. Each line on the effect plot belonging to rider type i indicates the predicted response if the amount of that rider type increased/decreased by a fraction of Δ and the rest of the mixture components decreased/increased in the same proportion as they were in the reference mixture $\bar{X} = \{0.2, 0.2, 0.2, 0.2, 0.2\}$.

In order to visualize the effect of the mixture components and the process variables ($fleetSize$ and $dtMode$), I plotted effect plots for the percent shared rides (Figure 5.2) and taxi capacity (Figure 5.3). A reference mixture is first selected (in this case $\bar{X} = \{0.2, 0.2, 0.2, 0.2, 0.2\}$, the centroid of the space formed by the 5 dimensional tetrahedron with corners x_1, \dots, x_5) and hold the process variables ($fleetSize$ and $dtMode$) at a constant value. Then, for each mixture component i (colors), we

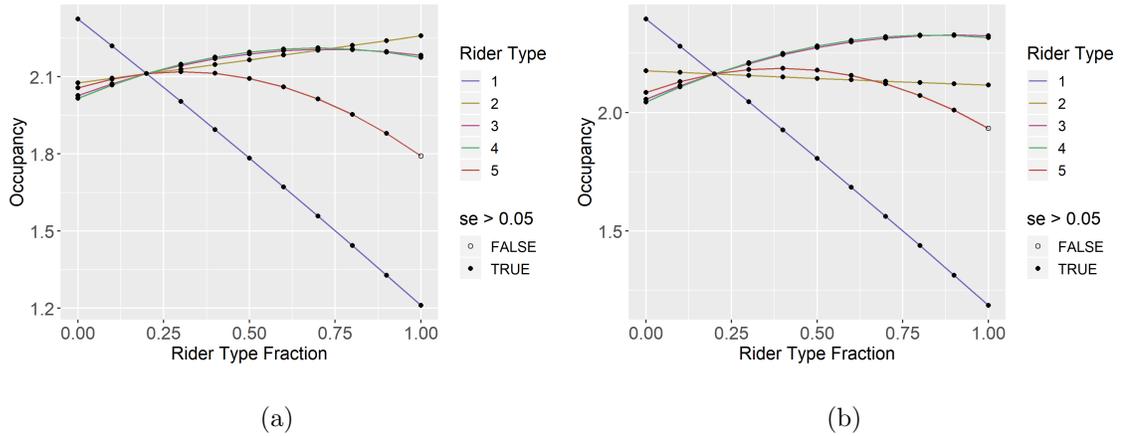


Figure 5.3. Effect plots of the predicted occupancy using the model represented by Table E.2 at different compositions of rider types at (a) $fleetSize = 4000$ and $dtMode = 0$ (b) $fleetSize = 6000$ and $dtMode = 1$. I have plotted the points with large standard error of prediction ($se > 0.005$) using hollow circles. Each line on the effect plot belonging to rider type i indicates the predicted response if the amount of that rider type increased/decreased by a fraction of Δ and the rest of the mixture components decreased/increased in the same proportion as they were in the reference mixture $\bar{X} = \{0.2, 0.2, 0.2, 0.2, 0.2\}$.

increase / decrease its composition in the mixture by an amount $\Delta \in [-\bar{x}_i, 1 - \bar{x}_i]$ ($x_i = \bar{x}_i + \Delta$ is on the x-axis). The other mixture components are calculated using the formula $x_j = \bar{x}_j - \frac{\Delta \bar{x}_j}{1 - \bar{x}_i} \quad \forall \quad j \in \{1, 2, 3, 4, 5\}; j \neq i$. This is done for several values of Δ , and for each mixture proportion setting X , I plot the predicted response (y-axis) using the regression model (Table E.1 for Figure 5.2) for the set value of the process variables and the mixture components. Each line on the effect plot belonging to rider type i indicates the predicted response if the amount of that rider type increased/decreased by a fraction of Δ and the rest of the mixture components decreased/increased in the same proportion as they were in the reference mixture \bar{X} .

From Figures 5.2 and 5.3, we see that increasing the proportion of rider type 1 (non-sharing) results in a linearly decreasing proportion of shared rides and average occupancy during the peak period (which is expected). Additionally, for rider type 5, as I increase its proportion in the system, both the fraction of shared rides and the occupancy increase till a certain point. Beyond this point, increasing the proportion of rider type 5 would decrease the system efficiency. This suggests that if rider type 5 is present in the system, its proportion should be adjusted, using incentives, in such a way that its proportion remains at 0.3 when $fleetSize = 4000$ and $dtMode = 0$ and 0.4 when $fleetSize = 6000$ and $dtMode = 1$. The effect that rider type 2 has on the system depends on the number of vehicles in the system. We can see in Figures 5.2(a) and 5.3(a) that since there are less taxis in the system, rider type 2 is forced to share rides, thus increasing the system ride sharing efficiency. However, when there are sufficient vehicles in the system, increasing the proportion of rider type 2 would decrease the system ride sharing efficiency, since they would take non-shared rides. Lastly, when there are low number of taxis, increasing the proportion of rider type 3 and rider type 4 have similar effects to increase the percent of shared rides. However, when the number of taxis in the system is higher, in order to increase the percent of shared rides, it is helpful to incentivize other rider types to become rider type 4.

5.3.1.3 CO₂ Emissions

The emission reduction as a result of ride sharing as estimated in the literature (Caulfield, 2009; Fagnant and Kockelman, 2014; Levofsky and Greenberg, 2001) and in Chapter 3 could be improved by considering the heterogeneous ride sharing preferences. We estimated the total distance travelled by all the taxis over the day in miles for each scenario in Table D.1, then, we multiplied this by a factor of 404 g/mile

which is the USEPA estimate for the CO₂ equivalent of one mile of gasoline vehicle travelled (EPA, 2017) to find the GHG emissions for one day of operation for each scenario.

From Figure 5.1, we see that, since different mixes of riders can have large changes in service level, in order to compare the scenarios on a fair ground, we considered only those scenarios that had a service level between 0.85-0.86 (since as per Table 3.2, a service level of 0.86 is equivalent to the present day taxi operations in NYC in the base case). In the set of scenarios that had a service level between 0.85-0.86, all the scenarios had *fleetSize* = 6000 vehicles and *dtMode* = 1. Figure 5.4 shows the composition of rider types in these 19 scenarios, which are ranked in increasing order of their GHG emissions (in tonnes of CO₂-equivalent). The difference between the GHG emission levels of scenario 1 and scenario 19 is 34 tonnes CO₂-equivalent.

From Figure 5.4, we see that when there are low number of vehicles in the system, having all riders in the system sharing rides is essential. However, there does not seem to be a clear pattern in the composition of rider types (specially in relation to rider type 2,3, and 4) in relation to the environmental impacts, as long as all the riders in the system are willing to share (sharing does not have to be preferred). However, we do see in Figure 5.4 that, if we do wish to reduce GHG emission, it is beneficial to have a small fraction of type 5 riders (sharing only) in the system.

5.3.2 Individual rider types

The aggregated system level implications presented in Section 5.3.1 are useful for city planners to study the overall impact of implementing ride sharing for a given rider type mix. However, it is also important to understand how the rider composition affects the service quality for each rider type. Such analysis can help

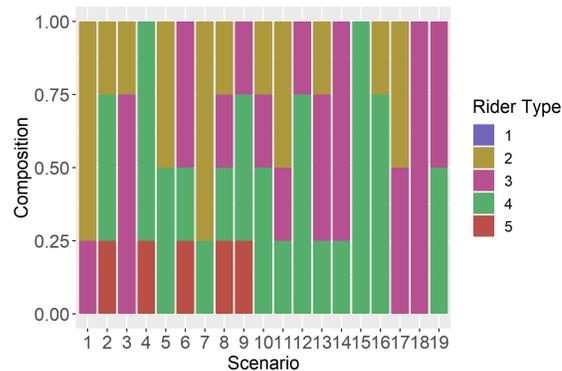


Figure 5.4. rider type compositions for scenarios with service level between 0.85 to 0.86 ranked in decreasing order of

transportation planners and service providers to implement policies that can either target a particular segment of the population, or ensure that all potential customers have equal service quality. Ride sharing companies could also provide such information to particular riders to incentivize them to adopt a particular ride sharing strategy.

The riders service quality depends on how likely they are to be served, and how long they need to wait for the service. I can use the simulation experiments to find the fraction of served riders, and their waiting time, for each rider type. Figure 5.5 shows that there are major difference in the service quality of the 5 rider types. In general, we can see that those rider types that are rigid in their sharing preferences (that is, sharing only or non-sharing only) have the least service level, the maximum variability in service level and lower waiting times, since sometimes the riders preferred type of vehicle may not be available. We also see that when rider type 1 does get served, they have the lowest waiting time of all in the system because if they are served, it needs to be within their low *exitTime* of 5 minutes (see Table

2.2). On the other hand, the most flexible rider type (rider type 3) has relatively high service level, low waiting time, and the least variability in both service level and waiting time. Interestingly, rider type 2 (non-sharing preferred) has the highest average waiting time as well as the most variability in waiting time.

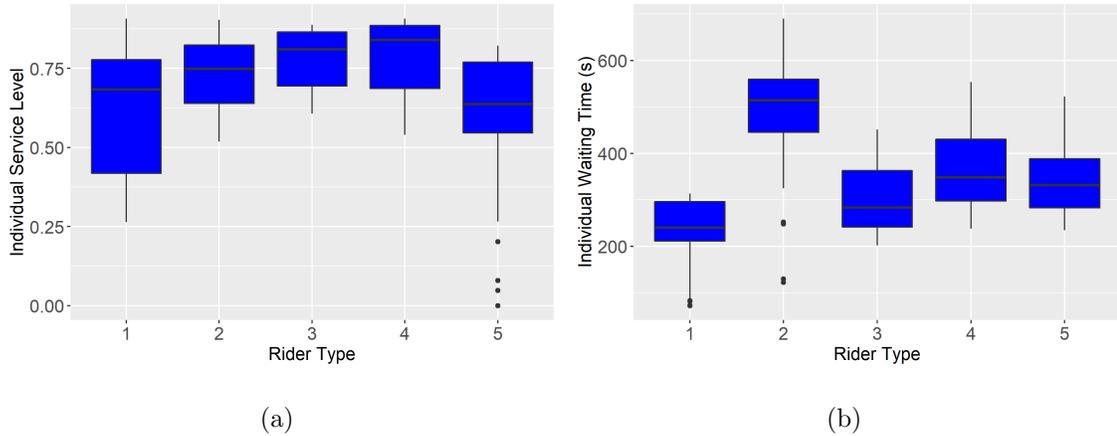
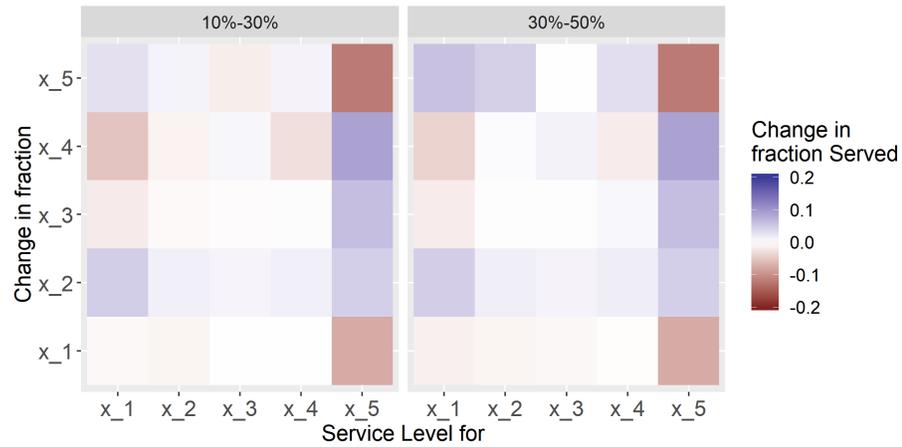


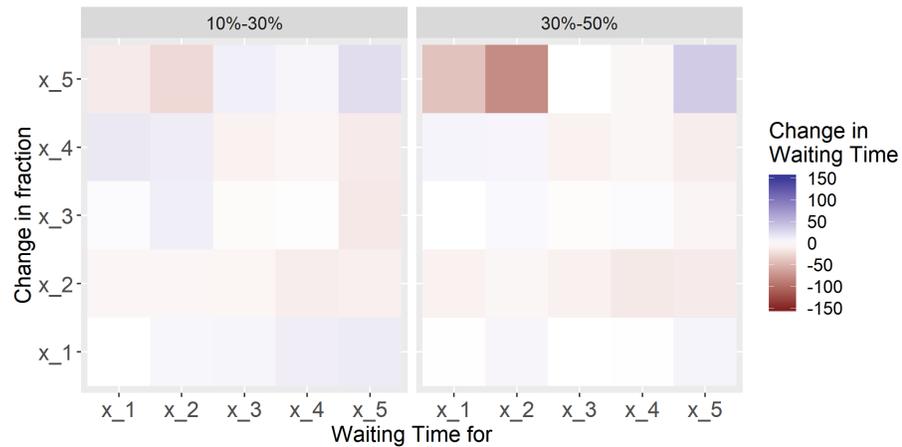
Figure 5.5. Box plots of (a) fraction of riders served and (b) Waiting time for the five rider types

In order to understand the effect of individual rider types on each other's service quality, I built regression models for the fraction of riders served for rider type 1 - 5 (Table E.3 - E.7) and the average waiting time for rider types 1 - 5 (Table E.8 - E.12). Then, in Figure 5.6 and 5.7 for each rider types service fraction and waiting time, I plotted the predicted change in these variables (x-axis) when the fraction of rider type x_j (y-axis) was increased from 10% to 30% and from 30% to 50%.

From Figure 5.6, we see that the fraction of riders served for rider type 1 and type 5 is most dependent on the fraction of other riders present in the system, since they are the most inflexible. For rider type 1, we see that its service level and waiting time are relatively independent of changes in its own fraction in the system, but are

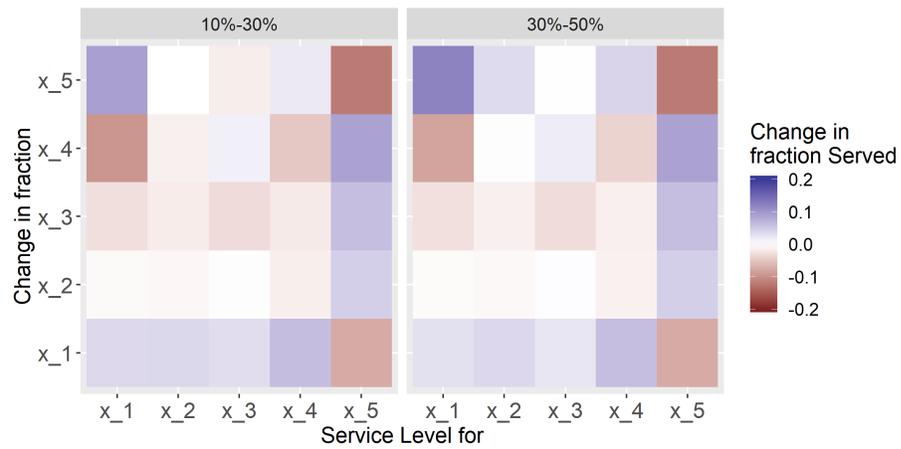


(a)

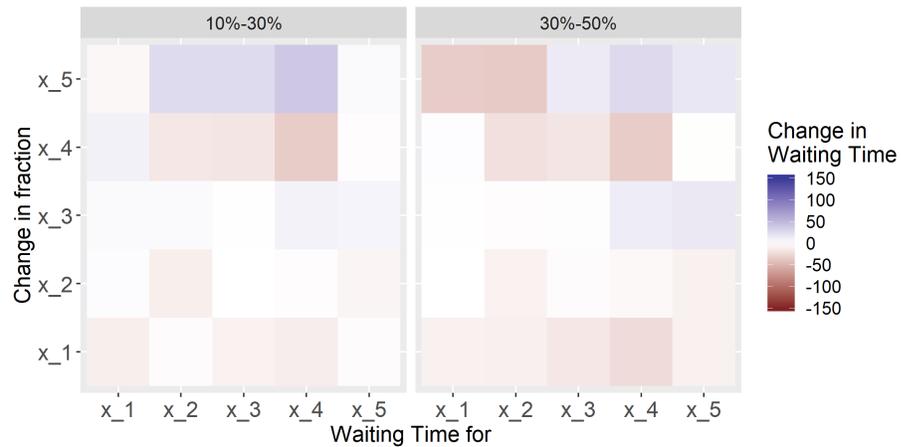


(b)

Figure 5.6. 2D bar plots for the change in the (a) served fraction (b) waiting time (color) for a rider type on the x axis with a change in fraction of the rider type on the y axis between the values indicated by the panel labels when $fleetSize = 6000$ and $dtMode = 1$



(a)



(b)

Figure 5.7. 2D bar plots for the change in the (a) served fraction (b) waiting time (color) for a rider type on the x axis with a change in fraction of the rider type on the y axis between the values indicated by the panel labels when $fleetSize = 4000$ and $dtMode = 0$

positively influenced by increasing fractions of rider type 2 (they would behave similar to rider type 1 but, if needed, would only share among themselves) and rider type 5 (since, rider type 5 would not get served themselves as there would be less riders to initiate shared ride. The effective outcome of increasing rider type 5 would be to reduce the total number of riders eligible for service in the system). Also, increasing rider type 4 results in the largest decrease in service level for rider type 1 and the largest increase in waiting time, due to lower number of non-sharing taxis being available (since rider type 4 prefers shared rides, and can also initiate them). Figure 5.7 shows similar trends when the number of vehicles and the deviation tolerance is lower. For rider type 5, its service level is actually most negatively affected by increasing its own fraction in the system, since rider type 5 needs flexible rider types to initiate ride sharing. In fact, at higher fractions of rider type 5, an increase in its fraction would increase its own waiting time by a large quantity. The service fraction of rider type 5 is positively influenced by adding more flexible riders in the system as they would help initiate ride shares. However, its service level is relatively unaffected by the number of vehicles in the system. The change in waiting time though is larger when there are more vehicles in the system.

The service quality for flexible rider types (2,3,4) are less affected by changes in other rider type fractions when the number of vehicles is higher (Figure 5.6). However, when the number of vehicles are low, the fraction of other rider types greatly affect the fraction of served riders for the flexible rider types as well (Figure 5.7). The service quality of rider type 2 is most positively influenced by increasing rider type 5 when the number of vehicles are high (since a higher number of vehicles allow the rider type 2 to find non-shared rides as per their preference). However, when the number of vehicles are low, increasing type 5 riders actually decreases the quality of service for rider type 2, because low number of vehicles will actually result

in more shared rides, which is not preferred for rider type 2 but is required for rider type 5. Rider type 3 is the least affected by changes in other riders proportions in the system, specially when the number of vehicles are high, since rider type 3 is mostly self sufficient by giving equal priority to shared and non-shared rides. When there are less vehicles in the system though, we see that the service fraction of type 3 riders are increased with increasing fraction of type 4 and decreased by increasing its own fraction in the system. These insights regarding service quality can help planners to design incentives that target specific groups of riders, thus encouraging riders to adopt a certain sharing preference.

5.4 Conclusions and future scope

The PP-SAEV model was applied on a case study of New York City taxis, and it was found that changes in rider sharing behaviors can have a large impact on the performance of a ride sharing system. Specifically, different compositions of different sharing preferences can change the service level by as much as 10%. The highest service level can be achieved with a composition of 70% of Type 3/4 riders (riders which are either indifferent to sharing or prefer to share) and 30% of riders of type 5 (which will only accept shared rides). However, rider sharing preferences do not have a major impact on GHG emissions.

In order to inform incentive formulation, the impact of changing rider compositions on the service level and waiting time of individual rider types was studied. The service quality of rider type 1 and rider type 5 are most affected by changes in compositions of other rider types, while the service level of the flexible rider types (2, 3, and 4) are most impacted by changes in composition only when the number of vehicles in the system are low.

It is notable that this study did not consider how variations in rider compositions in space and time could affect the performance of the system. In reality, research has shown that sharing preferences of riders could be correlated to the the pick up and drop off time as well as the demographics of people (which could differ in different areas of the city). A model that considers these variations could improve the understanding of how different sharing preferences could impact SAEV systems in a city.

While studied the impact of different rider preferences was studied, the reasons and incentives that cause a rider to adopt a certain sharing preference are beyond the scope of this study. Future studies that do identify these correlations between incentives and rider preferences would help us gain a deeper understanding of the ride sharing system.

6. SITING CHARGING STATIONS FOR INCREMENTAL ELECTRIC VEHICLE ADOPTION IN TRADITIONAL FLEETS AND FUTURE SHARED AUTONOMOUS VEHICLE FLEETS

6.1 Introduction

Electric vehicles (EV), especially battery electric vehicles, have the potential to reduce the transportation sector's greenhouse gas emissions (Choi et al., 2018). Traditionally, EV adoption has been slow due to the higher initial costs, limited battery range, long charging time, and low availability of public charging (Clean Technica, 2018). However, in recent years, with the fast dropping battery prices and the introduction of electric cars with longer and longer ranges, EV adoption has been increasing rapidly (Bloomberg NEF, 2018). Many cities and countries have built an initial charging infrastructure to support this increasing EV adoption. However, as the share of EVs grows, the charging infrastructure will also need to be expanded to support the additional charging demands.

The additional charging infrastructure can be built in two ways: (1) installing new charging stations (one charging station may have one or more charging ports) or (2) selecting existing ones to install more charging ports (station expansion), if space allows. Each option has different advantages and disadvantages. New station installation can help expand the spatial coverage of the charging network. If sited closer to the charging demands, new charging stations can help reduce the distances

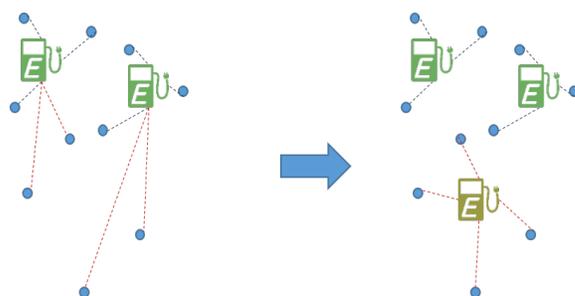


Figure 6.1. Adding a new charging station can decrease the total travel distance for charging. The blue dots indicate the charging demand (e.g., vehicles with the need for charging), the green chargers indicate the locations of the existing charging stations, the yellow charger indicates the additional charging station, and the dotted lines indicate the route to the closest charging station for each charging demand.

vehicles/drivers have to travel to visit a charging station (Figure 6.1). However, building new charging stations is expensive. Expanding the capacity of the existing charging stations, on the other hand, can reduce the installation costs by leveraging the existing infrastructure (e.g., existing electrical service) at that location (Smith and Castellano, 2015). Given a fixed budget, this saving could potentially allow installing more charging ports at the existing charging stations, having a total of more charging points than what can be offered by installing new charging stations. The expanded capacity at the existing stations can also reduce the waiting time at that station, if multiple EVs are queued for charging. However, the location of these additional ports are constrained to the locations of the existing stations, which may cause the drivers/vehicles longer travel distances to visit the stations. Therefore, when siting charging stations for incremental EV adoptions with a given budget, it is important to consider the trade-offs between reducing waiting time (expanding

existing stations to build more charging ports) and reducing overall travel distance for charging (building new charging stations to be close to the charging demands).

This chapter proposes a method to site charging stations for a growing fleet of electric vehicles, considering preexisting charging infrastructure. The model sites charging stations (including installation of new ones and expansion of existing ones) under a fixed budget constraint, optimizing the number of charging stations, the location of the charging stations, and the number of ports at each charging station. The charging demand is generated using an agent-based model that simulates the interactions among vehicles, rider groups, and charging stations. The optimization model minimizes the total time wasted due to charging, evaluating the trade-offs between travel time to charging stations and waiting time spent in the charging station queue. This method is applied to the case study of New York City taxis (Section 2.5) with different adoption pathways, evaluating the co-adoption of EV and other emerging transportation technologies (e.g., ride sharing and autonomous driving). The rest of the chapter is organized as follows. Section 6.2 discusses the existing literature of EV charging station siting, identifies the research gaps, and summarizes the contributions of our proposed method. Section 6.3 provides the details of our proposed method to site EV charging stations for a growing EV fleet. Section 6.4 introduces the case study and discusses the results. Lastly, the conclusions and directions for future research are reported in Section 6.5.

6.2 Literature Review

Charging station siting for EVs have received increasing attentions in recent years. For an extensive review on alternative fuel vehicle (which includes EV) fuel-

ing/charging station siting, I refer the readers to Ko et al. (2017). In this Section, I would like to focus the discussion on the major limitations of the existing literature.

First, the existing literature sites charging stations for static (or fixed) EV adoption (non-increasing) EV adoption which represent extreme cases. For example, He et al. (2015) and Chen et al. (2016b) considered a case in which all trip requests are fulfilled using EVs; Arslan and Karaşan (2016) assumed that there is a fixed demand for EV charging; Shahraki et al. (2015); Yang et al. (2017b); Tu et al. (2016) considered the travel patterns of a fixed number of electric taxis to site charging stations; and Liu (2012) evaluated a scenario in which the future charging demand for electric vehicles could be approximated by the current demand for gasoline (this represents 100% EV adoption). In reality, EV adoption will not happen over night but incrementally increase year by year (Bloomberg NEF, 2018). To keep pace with the increasing EV adoption, charging stations will need to be built at different stages of EV adoption. When a city is planning for 100% EV adoption, very likely there are a number of charging stations already existing in the system, which were built to satisfy charging demand at an earlier stage of EV adoption (e.g., 50%). However, existing literature (He et al., 2015; Chen et al., 2016b; Arslan and Karaşan, 2016; Shahraki et al., 2015; Yang et al., 2017b; Tu et al., 2016; Liu, 2012) does not take into account pre-existing charging stations in the network while siting new ones. If we solve separate optimization problems for each EV adoption level, this will imply that we need to relocate many, if not all, of our existing charging stations, which would not be economically feasible. Additionally, siting EV charging stations for 100% EV adoption in the present day (when the actual EV adoption is just 3%) would also not be economically feasible, considering that the useful life of an EV charging station

is limited¹. Some recent studies, Li et al. (2016a); Davidov and Pantoš (2017) have considered siting charging stations for different periods in time. Davidov and Pantoš (2017) considered locating stations for an EV car sharing program in a simulated road network at different points in time based on the demand generated by its users. Li et al. (2016b) considered the case where new cities would become candidates for EV adoption. However, Davidov and Pantoš (2017) and Li et al. (2016b) did not consider that EVs could queue at charging stations (queuing at charging stations is discussed in subsequent paragraphs)

Second, there are very few papers (Sadeghi-Barzani et al., 2014; Zhu et al., 2016; Jung et al., 2014; Xi et al., 2013) that optimize the number of charging stations, the location of the charging stations, and the number of charging ports at each station integratedly at the same time. For example, Davidov and Pantoš (2017); Brandstätter et al. (2017); He et al. (2015) used a binary variable to select whether a site would be developed at a certain time period or not, and Chen et al. (2016b) used a simulation model to determine the number of sites, but they did not set the number of ports at each charging station. They instead either assumed that charging stations would have sufficient capacity to service all demands (Davidov and Pantoš, 2017) or sited sufficient single port charging stations to serve all riders (Chen et al., 2016b; Brandstätter et al., 2017; He et al., 2015). Others, for example Shahraki et al. (2015), considered that cities have a fixed limit on the number of sites for charging stations. When considering adding new charging stations to a network, system planners will typically have a budget for expanding an existing network. The costs of installing and maintaining charging stations can vary from site to site and some of these costs

¹There has not been enough data to estimate the actual useful life of an EV charging station. NYCTLC (2013) considers a 5 year useful life, while Department of Energy (2018) considers a 10 year useful life.

can be shared by adding additional ports to the same charging stations (expanding). Thus, it is important to consider the trade-offs in deciding whether to add additional charging stations, which could cost more but reduce the travel time to a charging station, at new locations or to expand existing ones by adding more ports.

Additionally, very few papers considered that vehicles may need to queue at charging stations. For example, Sadeghi-Barzani et al. (2014); Brandstätter et al. (2017); He et al. (2015) considered the problem of sizing the charging stations but assumed that vehicles would not wait at charging stations for others to finish charging. They assumed that, at a given time, all charging demands would be met without the need of queuing. This would result in assigning more capacity to charging stations than what would be necessary. As shown by Yang et al. (2017b), by increasing the space for waiting at a charging station, the number of chargers can be reduced by 13-26%. There could be situations in which an EV currently in the process of charging only requires very little time (for example, 1 minute) to finish charging to its desired state of charge, while a second vehicle enters the station and requires charging. In this case, the optimization models not considering queuing would choose to add an additional charging station/port, instead of allowing the second vehicle to wait for a short period (e.g., 1 minute, which is a reasonable waiting time) for the charger to become available. Others, (Shahraki et al., 2015; Chen et al., 2016b; Liu, 2012) consider that all sited charging stations had sufficient ports to serve any demand. This could be infeasible in reality, because there is usually a limited budget to build charging stations and each additional port on a charging station has a certain cost associated with it. The number of cars that can charge at a charging station is also limited by capacity constraints of the space itself. A few papers do include the waiting time at charging stations as a part of their optimization procedure. For example, Yang et al. (2017b) used a $M/M/x/s$ (arrival rates and service rates defined

by a Poisson process) queuing model to consider the wait time at charging stations. However, in most cases, we cannot expect that either the arrival rate or the service rate for a charging station remains constant or follows an exponential distribution for the entire length of the day. Additionally, the service rate cannot be realistically modeled by an exponential distribution. For example, Morrissey et al. (2016) analyzed the charging patterns from a charging network in Ireland and found that the number of vehicles beginning to charge at a certain time is not constant through the day and peaks at about 3:00pm. Rao et al. (2018) used data from charging stations in Shenzhen, China and found that the daily distribution of EV taxis' charging start time can have multiple peaks and is better approximated by Gaussian Mixture Models as opposed to an exponential distribution, which has a constant mean through the day. Some papers, Jung et al. (2014); Tu et al. (2016) do consider waiting time for the vehicles at charging stations empirically from the data. however both Jung et al. (2014); Tu et al. (2016) site charging stations for only a single period of time (as discussed earlier in this Section).

Furthermore, all of the literature on charging station siting considers fixed charging rules for the EVs. For example, Shahraki et al. (2015) considered that the plug-in hybrid EVs in their model would maximize the distance that the vehicle would travel on electricity and charge at all times that the car was parked within the service range of a charging station. Chen et al. (2016b) considered that EVs would charge if the remaining range was less than 2 miles. Others (He et al., 2015) using a fixed constraint in an optimization program which thresholds the the vehicle's SOC to a certain minimum. Arslan and Karaşan (2016) set a rule that EVs must start and end a trip with at least 50% SOC. In reality, the reason why an EV (and specifically electric taxis) would seek charging is complex and cannot be defined by simple rules (Franke and Krems, 2013; Neubauer and Wood, 2014). For example, electric taxis would need

to maintain a balance between serving riders and seeking charging opportunities, to make sure that they have sufficient SOC to meet demands during the peak demand period. Rao et al. (2018) noted that most electric taxis seek charging opportunities as much as 4 times a day.

Lastly, most of the literature do not provide insights on how EVs and optimum charging station siting would affect the performance of a system of EVs as compared to a system of internal combustion engine vehicles. For example, He et al. (2018); Arslan and Karaşan (2016); He et al. (2015); Davidov and Pantoš (2017); Alhazmi et al. (2017); Lee and Han (2017); Wu and Sioshansi (2017); Brandstätter et al. (2017); Shahraki et al. (2015); Zhu et al. (2016) focused on finding better solutions and algorithms to find optimal charging station sites and did not provide any insights on system level impacts of EV adoptions. Some studies do report system level impacts but are mostly concerned with operational characteristics of EVs. Xi et al. (2013) used their model to show that, with an increasing budget to site charging stations, the expected daily energy used by EVs increases till a point and then remains steady, and the number of EVs that ran out of charge while serving customers also decreases. Li et al. (2016b) studied the impact of variability of future demand in their multi-period model and reported on parameter setting in their model that would make the model robust to future changes. Yang et al. (2017b) analyzed the trade-offs between installing more chargers and increasing the waiting space and showed that this depended on the ratio of cost of parking spots to the cost of the chargers. Some studies do focus on making system level inferences. For example, Cai et al. (2014b) compared the impact of relocating the existing charging stations in Beijing on the electric power grid and found that the charging peak in Beijing coincided with the peak demand for electricity. However, Xi et al. (2013); Li et al. (2016b); Yang et al. (2017b); Cai et al. (2014b) do not report system level statistics such as

the number of riders being served. These system level statistics are important to report as they help planners evaluate the performance of their system with respect to multiple parameters of interest. There are a few studies that do make system level inferences. For example, Chen et al. (2016b) used their simulation to show that driving to charging stations would not increase the total daily distance travelled by a large amount. They also showed that each SAEV could replace 5-9 privately owned vehicles. Bauer et al. (2018) estimated that, for Manhattan, NY, the costs would be the lowest for EVs with battery ranges between 50 to 90 miles and 66 chargers per square mile of 11kW each, and a SAEV fleet could reduce GHG emissions by 73%. However, none of the previous studies compared the differences in siting charging stations for a non-autonomous vehicle - ride sharing (non AV-RS) system with that for a autonomous vehicle - ride sharing (AV-RS) system for increasing levels of EV adoption. Since AVs, RS, and EVs are each emerging technologies that are growing at variable rates, it is important to understand the synergies between these systems to inform policy makers and city planners on adoption strategies and development plans for these emerging systems. I have summarized the existing literature in Table 6.1

This chapter proposes a method to site charging stations for a city given a certain budget and uses a case study to make inferences on non AV-RS and AV-RS scenarios. The unique contributions of this work are to build a model to site charging stations with the following considerations:

1. Increasing EV adoptions at several points in time, considering the continuous use of charging stations that have already been installed at previous points.

2. Optimizing both the number of charging stations and the number of ports simultaneously with a fixed budget, where the cost of a new charging station may be different from a new port at the same charging station.
3. Vehicles can wait (queue) at charging stations for other vehicles to finish charging. Our model does not assume fixed distributions for arrival and service rates at charging stations, but instead uses a simulation to model the arrival and service of vehicles that require charging at a charging station.

The demand for the model to site charging stations is estimated using using an agent-based model to simulate shared autonomous electric vehicles (the PP-SAEV model) in which the adoption of ride sharing, autonomous vehicles, and electric vehicles are variable parameters. The PP-SAEV model allows the EVs to decide when to charge based on a utility function that considers the distance from a charging station and the current SOC of the vehicle. Finally, I use a case study to show how increasing EV adoption would affect the number of riders served, the environmental impacts, and the effect that the charging EVs would have on the electric grid through the day using the charging station configurations proposed by the charging station siting model.

Table 6.1.: Gaps in recent literature in EV charging station siting. Station Counts refers to whether the number of charging stations were decided by the model, Charging Ports refers to whether the number of ports at each EV charging charging stations were decided by the model, Queuing at Station refers to whether the model considered that EVs would wait idle at a charging station if currently occupied, Multiperiod refers to whether the literature sited charging stations for multiple periods in time.

Reference	Network Type	Station Counts	Charging Ports	Queuing at Station	Multiperiod	Type of inferences
He et al. (2018)	Simulated					
Davidov and Pantoš (2017)	Simulated	✓			✓	
He et al. (2015)	Simulated	✓				
Chen et al. (2016b)	Simulated	✓				System level impact for AV-RS only
Arslan and Karaşan (2016)	Highway					
Alhazmi et al. (2017)	Highway	✓				
Lee and Han (2017)	Highway					
Wu and Sioshansi (2017)	Highway					
Xi et al. (2013)	Highway	✓	✓			
Li et al. (2016b)	Highway				✓	

continued on next page

Table 6.1.: *continued*

Reference	Network Type	Station Counts	Charging Ports	Queuing at Station	Multiperiod	Type of inferences
Brandstätter et al. (2017)	City	✓				
Cai et al. (2014b)	City					Impact on Grid
Liu (2012)	City					
Sadeghi-Barzani et al. (2014)	City	✓	✓			
Shahraki et al. (2015)	City					
Zhu et al. (2016)	City (Simplified)	✓	✓			
Yang et al. (2017b)	City	✓		✓		Waiting time vs driving time tradeoff
Bauer et al. (2018)	City					System level impact for SAEV only
Jung et al. (2014)	City (Simplified)	✓	✓	✓		
Tu et al. (2016)	City			✓		

continued on next page

Table 6.1.: *continued*

Reference	Network Type	Station Counts	Charging Ports	Queuing at Station	Multiperiod	Type of inferences
This Study	City	✓	✓	✓	✓	System level impact for AV-RS and non AV-RS

6.3 Method

The charging station siting framework includes three components: 1) demand generation using an agent-based simulation to estimate the charging demand (PP-SAEV) (**Generate demand**), 2) evaluation of objective function (**Evaluate objective function**), and 3) optimization of charging station using a modified genetic algorithm (**Genetic algorithm**). These three components are executed as per Algorithm 8 and are explained in details in the subsections of this Section. Figure 6.2 provides a visual representation of this method. The results from this charging station siting method are used as input to the PP-SAEV used in the demand generation step, so that I can analyze the system performance of the proposed siting infrastructure development.

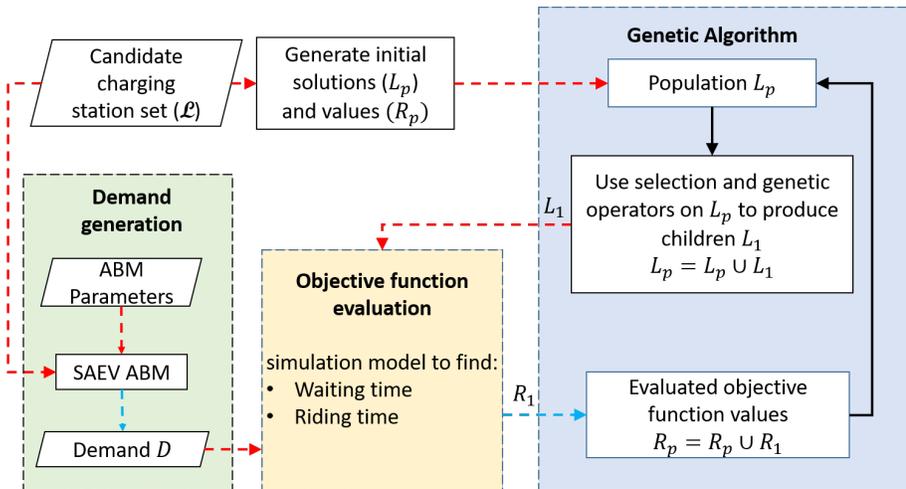


Figure 6.2. Flowchart of the method. The red dashed arrows represent an input and the blue dashed arrows represent an output. The colors in the flowchart represent the different parts of the method. green: demand generation (Section 6.3.2), orange: genetic algorithm (Section 6.3.3.1); yellow: objective function evaluation (Section 6.3.4)

I introduce the variables that are used in Algorithm 8 below:

- D : The set of charging demand locations, time, and energy requirements.
- i : An iteration counter

- I : The maximum number of iterations of the genetic algorithm.
- \mathcal{L}_{full} : A set of potential charging station locations, containing information on the location of the candidate site and the maximum possible capacity of that location (in terms of the total number of chargers that can be built on this site).
- \mathcal{L} : A subset of \mathcal{L}_{full} generated by the filtering procedure which is used to eliminate unlikely candidate solutions so as to improve the efficiency of the genetic algorithm (Section F.1.2).
- L_p : A set of all the siting plans (solutions) as found by the genetic algorithm. Also referred to as the chromosomes.
- L_1 : A set new of chromosomes formed by using the genetic operators on L_p in every iteration.
- R_p : The set of objective functions corresponding to each chromosome in set L_p
- R_1 : The set of objective functions corresponding to each chromosome in set L_1
- S : The selected charging station configuration (the number of charging ports to be built on each candidate sites). I select the charging station configuration $S \in L_p$ with the least objective function value (from set R_p).

6.3.1 Key Assumptions

Our model is built on the following key assumptions.

Algorithm 8 Main algorithm

- 1: **procedure** MAIN ALGORITHM(*riderGroup*)
 - 2: **Generate demand** D using the PP-SAEV as an estimate of the demand
 - 3: Filter the set of potential charging station locations \mathcal{L}_{full} to \mathcal{L} to improve the efficiency of the genetic algorithm.
 - 4: Generate an initial population of charging station locations L_p from the set of filtered locations \mathcal{L} using Algorithm F.1.2
 - 5: **Evaluate objective function** for L_p and add the evaluated objective values to set R_p ▷ **Genetic algorithm**
 - 6: **for** $i = 1 \dots I$ **do**
 - 7: use genetic operators (Section F.1.3) on set L_p to generate L_1 new solutions and $L_p \leftarrow L_p \cup L_1$
 - 8: **Evaluate objective function** for L_1 and $R_p \leftarrow R_p \cup R_1$
 - 9: Report the solution L_p that has the lowest objective function value (from R_p) as the proposed charging station set S
-

- A known total budget is available to develop the charging infrastructure at different steps in time with increasing levels of EV adoption. For example, if the EV adoption were to grow from 0% to 100%, the charging infrastructure may be built at discrete time points to support a total of 10%, 40%, 70% and 100% EV adoption, with different budget assigned at each time point. In this study, I refer to a single discrete time point as an epoch.
- I start the first epoch assuming that there are no charging stations currently in the system. Our model will optimize the charging infrastructure to support of the first stage's EV adoption (e.g., 10%) from the initial 0% EV adoption.

- For all subsequent epochs after the first one, it is assumed that all charging stations and charging ports in the previous epoch will continue their service as part of the expanded infrastructure.
- Because this model does not consider the number of years it takes to reach the next EV adoption epoch, it assumes the life time of charging stations to be infinite ¹.
- All the charging stations that we site are assumed to have the same charging rate.

6.3.2 Demand generation using PP-SAEV

The charging station demand is generated using the PP-SAEV from Chapter 2 with a set of parameters (*fleetSize*, *PercentAV*, *PercentEV*, *PercentSharing*, t_s , t_f) and with the candidate locations for building charging stations/chargers as inputs. The model then outputs a list of time and locations that the EVs demanded charging, along with the increase in SOC that was obtained from charging (Table 6.2) for that parameter setting. Because our model considers flexible charging rules (the decision to charge is made using a utility function described in Section A.3.5.3 of the Supplementary Information), the charging infrastructure can be better sited using the unconstrained charging demands. From the PP-SAEV, the unconstrained charging demand is recorded at the locations that the taxis needed / wanted to charge, assuming that they had access to all possible charging stations. To do this, the charging station locations are set to be equal to all possible locations where charging stations can be installed (\mathcal{L}_{full}). Additionally, these charging stations are assigned unlimited number of charging ports. We also needed to know the *SOC*

increase that the EV required (ΔSOC), since I allowed EVs to search for and accept rides while charging at a charging station. The ΔSOC is recorded as the difference between the SOC when the vehicle began charging and the SOC recorded when the vehicle leaves the charging station. Each row of Table 6.2 represents the time, location, and the (ΔSOC) that is desired, when an EV required charging. The demand generated is then used as input to the optimization model presented in the next subsection.

Table 6.2.
Sample demand data generated from the agent-based model

demand Time	demand Lat	demand Long	SOC Change
06/12/2014 0:00:00	-73.937	40.797	0.34
06/12/2014 0:10:00	-74.001	40.717	0.23
06/12/2014 0:12:20	-73.984	40.765	0.8
⋮	⋮	⋮	⋮

6.3.3 Optimization of charging station locations

Our paper sites EV charging stations for different levels of EV adoption. At each level of EV adoption, the charging stations that are present in the system from the previous period are considered to be fixed. Idling EVs are allowed to wait at a charging station while they are seeking rider requests. In order to do site the

charging stations, the following optimization model represented by Equations 6.1 to 6.5 is solved at each level of EV adoption.

I first define notation that I use to introduce the optimization model:

- B : The total allowable budget to site charging stations in the current epoch
- C_n : The cost to install charging stations at a new location
- C_u : The cost required to add one additional port to an existing charging station (station expansion)
- D : The demand that is generated using the PP-SAEV
- i : Index for the demand from set D
- \mathcal{I}^+ : The set of positive integers
- j : Index for the potential charging station location from set \mathcal{L}
- K_j : Maximum capacity of charging station j
- \mathcal{L} : A subset of \mathcal{L}_{full} generated by the filtering procedure (Section F.1.2).
- N : A vector $\{N_1, N_2, \dots, N_l\}$ where l is the length of \mathcal{L} . Each N_i is the of the number of charging ports at candidate charging station location i in \mathcal{L} . If the number of charging port is zero, it means that no charging station is built at this candidate site. The set L_p is made up of many different N as determined by the genetic algorithm.
- N_j : The number of charging ports at candidate location j in the current epoch
- \hat{N}_j : The number of charging ports at charging station j in the previous epoch

- T : The total wasted time (the objective function that we seek to minimize).
- T_w : The total waiting time for all the demands
- T_t : The total travel time for all the demands
- T_{w_i} : The waiting time of demand i
- T_{t_i} : The time spent in traveling to charging stations for demand i

$$\text{minimize } T = \sum_{i \in D} T_{w_i} + 2 \sum_{i \in D} T_{t_i} \quad (6.1)$$

Subject To

$$C_n \sum_{j \in \mathcal{L}} (\mathbb{I}_{N_j > 0} - \mathbb{I}_{\hat{N}_j > 0}) + C_u \sum_{j \in \mathcal{L}} (N_j - \mathbb{I}_{N_j > 0}) \leq B \quad (6.2)$$

$$N_j \geq \hat{N}_j \quad \forall j \in \mathcal{L} \quad (6.3)$$

$$N_j \leq K_j \quad \forall j \in \mathcal{L} \quad (6.4)$$

$$N_j \in \mathcal{I}^+ \quad (6.5)$$

Equation 6.1 represents the objective function minimizing the total wasted time T (in seconds), which I use as a measure of the inconvenience caused to an EV to satisfy its charging demand ². The objective function is the sum of the time that each of the vehicle represented by the demands in D waits at a charging station T_{w_i} and the round trip time T_{t_i} that vehicles have to spend on traveling to the nearest charging station from their original locations. The trip time is computed using the

²Total Wasted Time = 2(Total Time spent in driving to the charging station) + Total Time spent waiting at the charging station. I double the time spent driving to the charging station (e.g., considering a round trip) to account for the potential change in service patterns that could result from having the taxis moving away from its current location to charge.

distance traveled by road from the vehicle’s original location to the nearest charging station. The demands D (see Table 6.2) contain information on their location (the vehicle’s original location when charging demand arises) and the SOC increase that they obtained from charging at that location (Δ SOC). The waiting time T_{w_i} for taxi i is recorded from the simulation as the time that the taxi spent waiting in the charging station queue (the time elapsed between the taxis arrival at the charging station and the time that the taxi began charging). This waiting time is influenced by N_j . If N_j is larger, there will be more charging ports at location j , and thus more taxis can be served at the same time, thus potentially reducing T_{w_i} . The trip time (T_{t_i}) is dependent on which of the L_p potential charging stations are selected to install charging stations or ports by the model. Whether a charging station is selected by the model is found by applying the indicator function $\mathbb{I}_{N_j>0}$ ³. The goal of the optimization model is to find the vector of number of ports N which minimizes T . In order to make use of real-world road networks while computing the travel time T_{t_i} , and to implement queuing at charging stations (which is used to calculate T_{w_i}), I compute the objective function using a discrete event simulation model that is further explained in Section 6.3.4.

Equation 6.2 places budget restrictions on the infrastructure development in this EV adoption epoch. The left hand side (LHS) of Equation 6.2 represents the budget that is being used by the proposed charging infrastructure expansion configurations. The first term of the LHS represents the cost incurred to build the first charging port at each new site, where C_n is the cost of each new location added. The number of charging stations can be found from $\mathbb{I}_{N_j>0} - \mathbb{I}_{\hat{N}_j>0}$ where \hat{N}_j is the number of ports for location j that are carryovered from the previous epoch. The second term calculates

³The indicator function \mathbb{I}_{cond} takes the value 1 if the condition *cond* is true, and 0 if false.

the cost of installing additional ports to an existing or newly selected location, where C_u is the cost of adding a new port beyond the first one at a charging station. The number of new charging ports added to a location can be computed from $\sum_{j \in \mathcal{L}} (N_j - \mathbb{I}_{N_j > 0})$. The right hand side (RHS) of the equation represents the budget B that is available for siting the new charging stations or expanding existing ones in the current epoch.

Equation 6.3 disallows old charging stations and ports to be demolished. Equation 6.4 places a capacity restriction K_j on each potential location \mathcal{L} , which represents physical space restrictions (for example, a parking lot of a certain size can only hold a certain number of vehicles).

Finding an exact solution of the integer optimization program represented by Equations 6.1 to 6.5 is computationally challenging, particularly due to a potentially large number of candidate solutions and the difficulty in evaluating the objective function (since I needed to use simulation to allow the taxis to queue at charging stations and thus evaluate the waiting time). Hence I use a modified genetic algorithm as a meta-heuristic to solve this problem.

6.3.3.1 Genetic Algorithm

Genetic Algorithms (GA) are a class of evolutionary algorithms which search for the optimal solution of an optimization model by starting from an initial point and then progressing towards the optimal solution by applying genetic operators (such as mutation and crossover). Genetic algorithms are named after the natural process of genetic evolution where chromosomes of the parent undergo mutation and crossover to generate a new individual (Sivanandam, 2007). Genetic algorithms have been used in previous research on charging station location optimization (He et al., 2015;

Lee and Han, 2017; Sadeghi-Barzani et al., 2014; Zhu et al., 2016; Jung et al., 2014; Tu et al., 2016; Li et al., 2016b).

Here I present an overview of our modified genetic algorithm. The steps are explained in detail in Section F.1 of the Supplementary Information. In order to reduce the size of the candidate set, I first reduce the number of potential charging stations sites by applying a filtering procedure (Section F.1.2) to generate $\mathcal{L} = \{l_1, l_2, \dots, l_l\}$. The filtering procedure removes candidate locations that would increase the value of the objective function and thus would be unlikely to be a part of an optimal solution, or those that are located far away from the charging demand. By removing unlikely sites from the candidate set, I increase the likelihood that the genetic algorithm would find a better solution faster. A chromosome in the genetic algorithm is represented by a vector of numbers $N = \{N_1, N_2, \dots, N_l\}$, where N_i (also referred to as a gene) is the number of ports at location i . Then, I generate our initial set of chromosomes $L_p = \{n_1, n_2, \dots, n_l\}$, where each n_i is a chromosome which represents a potential valid set of charging stations that satisfies the constraints Equations 6.2 - 6.5. The objective function values R_p for the chromosomes in set L_p is then computed using the method described in Section 6.3.4. I then generate a set of new chromosomes L_1 with the goal of finding the optimal chromosome using the genetic operators (selection, crossover, mutate, perturbation, fission, and fusion) on L_p . The genetic operators are described in detail in Section F.1.3. I then evaluate the objective function of L_1 as R_1 , append L_1 to L_p and append the objective function values R_1 to R_p , and repeatedly apply the genetic operators till a termination criteria is reached (for example I iterations). In our genetic algorithm, I allow all the chromosomes that have been created to be selected as parents by the selection step in order to promote more diversity and hence decrease the chance that the population

gets stuck in a local minimum. The chromosome $S \in L_p$ with the lowest objective function value in R_p is chosen as the charging station setting for that epoch.

6.3.4 Objective function calculation

Below I define the variables that I use in this Section and in Algorithm 9:

- C : The installed charging stations in the objective function simulation model. It can be found by filtering the locations from \mathcal{L} where $N > 0$
- G : The number of best solutions that the genetic algorithm selects. It is a model parameter.
- I : The maximum number of iterations
- \mathcal{I}^+ : The set of positive integers
- K_j : Maximum capacity of charging station j
- \mathcal{L}_{full} : A set of potential charging station locations, containing information on the location of the charging station and the capacity of that location.
- L_p : A set of all the siting plans as found by the genetic algorithm. Also referred to as chromosomes.
- L_1 : The modified chromosomes after genetic operations
- L_{q_s} : Queue length of charging station s
- N : A vector $\{N_1, N_2, \dots, N_L\}$ of the number of charging ports at each charging station location in \mathcal{L} . The set L_p is made up of several N as determined by the genetic algorithm.

- N_j : The number of charging ports at potential charging station site j in the current epoch
- \hat{N}_j : The number of charging ports at potential charging station site j in the previous epoch
- Q_d : the current position of the demand d in the charger queue at station s
- R_p : The set of objective functions corresponding to each chromosome in set L_p
- $State_d$: A variable keeping track of the state of the demand d in the objective function simulation model. Possible values are : (0) Not Entered, (1) Moving to charger, (2) Queued, and (3) Charging.
- T : The total wasted time (the objective function that I seek to minimize).
- T_w : The total waiting time for all the demands
- T_t : The total travel time for all the demands
- T_{w_i} : The waiting time of demand i
- T_{t_i} : The time spent in traveling to a charging station for demand i
- T_{e_i} : The time that demand i enters the objective function simulation model.
- T_{s_i} : The time that demand i needs to charge so that its SOC increases by ΔSOC
- t : Time in objective function simulation model
- t_s : A variable to indicate the time of start of traveling or waiting

- U_s : The number of ports of the charging station s currently in use by the charging taxis.

The objective function (Equation 6.1), which is the total wasted time for all the demand points that need to charge, is computed using a simulation model (Algorithm 9). The demands (D) generated by the PP-SAEV (Section 6.3.2) enter into the objective function simulation at their scheduled time T_{e_d} and search for the nearest location s among the charging stations set $C \in L$ that were generated by the genetic algorithm (Section 6.3.3.1). The demands then start moving towards s and, when they reach s , they enter the charging station queue. All the charging stations serve the demands according to the first-come-first-served policy. The charging time required for each demand is given by $T_{C_d} = (\Delta SOC) \times T_{FC}$, where T_{FC} is the time required for the charging station to charge the vehicle from 0 to 1 SOC . Once the demand obtains the required SOC , it exits the simulation, recording the time it spent traveling T_{t_d} and the time it spent in queue T_{w_i} . I have illustrated this process in Figure 6.3.

6.4 Case Study

In order to demonstrate this method and understand the effects of increasing EV adoption in an SAEV system, I have applied the proposed method to three case studies. In Chapter 3, we found that 5500 autonomous shared taxis (A scenario) had the potential to serve as many riders as 13500 conventional taxi cabs without sharing (O scenario) for New York City taxi cabs. I also refer to the A scenario as the Autonomous Vehicle Ride Sharing (AV-RS) scenario, and the O scenario as the non AV-RS scenario. Using the method described in Section 6.3, I site charging stations for a progressively increasing EV fleet for the A and O scenarios, so that I

Algorithm 9 Compute objective function

```

1: procedure COMPUTE OBJECTIVE FUNCTION(riderGroup)
2:   Set  $t = 0, T_w = 0, T_t = 0$ 
3:   Set  $State_d = 0 \quad \forall \quad d \in Demand$  to 0
4:   Set charging station locations  $C$  for those  $N > 0$  and their number of ports
   to  $N$  for those  $N > 0$ 
5:   Set  $L_{qc} = 0; U_c = 0 \quad \forall \quad c \in C$ ;
6:   for All  $d$  in Demands do
7:     if  $State_d = 0$  then ▷ not entered
8:       if  $t = T_{e_d}$  then
9:         Set  $State_d = 1$ ;
10:        Find nearest charging station  $s$  and move towards it along the GIS
        route
11:        set  $t_s = t$ 
12:        if  $State_d = 1$  and  $d$  has reached  $s$  then ▷ moving to  $s$ 
13:          set  $State_d = 2$  ▷ Queued
14:           $T_{t_i} = t - t_s, t_s = t, Q_d = L_{q_s}, L_q = L_{q_s} + 1$ 
15:          if  $State_d = 2$  and  $Q_d = 0$  and  $U_s \neq N_s$  then
16:            set  $State_d = 3$ 
17:            Charging
18:             $T_{w_i} = t - t_s, t_s = t, U_s = U_s + 1, L_q = L_{q_s} - 1$ 
19:            if  $State_d = 3$  and  $T_{c_d} = t_s - t$  then
20:               $Q_d = Q_d - 1 \forall d$  which are in the queue of  $s$ 
21:               $U_s = U_s - 1, T_w = T_w + T_{w_d}, T_t = T_t + T_{t_d}$ 
22:              Exit the simulation
23:           $t = t + 1$ 

```

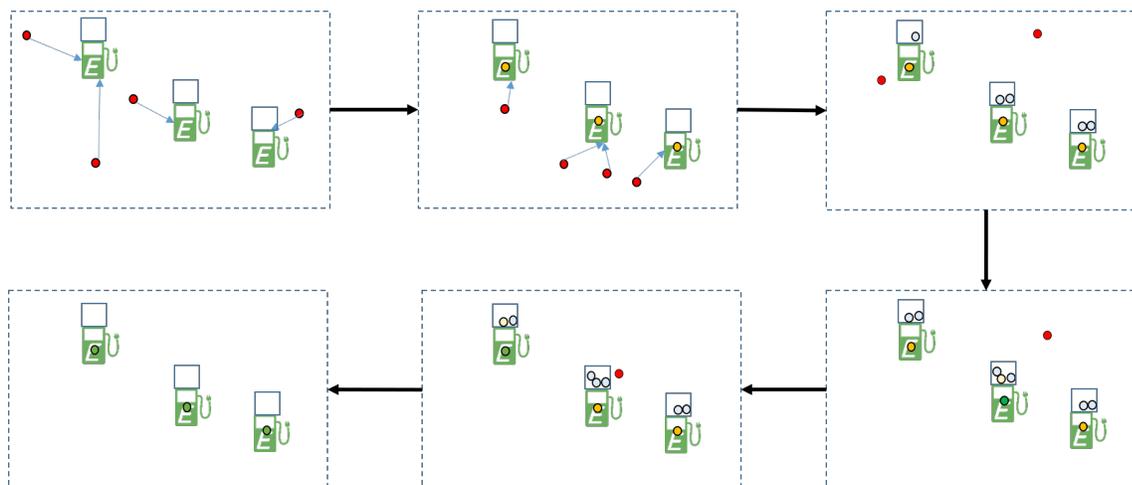


Figure 6.3. Illustration of steps used in the objective function evaluation. The 3 charging stations each have 1 charging port. The dots represent the current locations of the demands. The color of the dot represents the demands *State* - Red = 1 (Entered the simulation and moving towards charging station); Light Yellow = 2 (queued at charging station); Yellow = 3 (charging); Green = Exiting. The direction of the black arrows progressing time steps. The boxes above the charging stations are the queuing areas for the charging stations.

can understand the impacts of increasing EV adoption on the system performance indicators (e.g., service level, rider waiting time, total distance driven etc.) In order to examine the impact of switching from a non AV-RS case to a AV-RS case (the simultaneous adoption of autonomous driving, ride sharing, and electric vehicles), I consider a third case where we switch from the O scenario to the A scenario, while increasing EV adoption. In all three cases, I increase the EV adoption in four steps/epochs (10% \rightarrow 40% \rightarrow 70% \rightarrow 100%). For each scenario in the case studies, I assign the budget B , which is the total available resource units to site charging stations. In order to make scenarios at the same EV adoption level comparable, I

assign the same budget B for the same level of EV adoption across the three cases⁴. The three case studies are listed below. I use the notation RXX, where R refers to the scenario (A or O), while XX refers to the percent of EV adoption. For example, A10 refers to scenario A with 10% EV adoption (10% of the vehicles are EV). The parameter setting for each scenario is presented in Table 6.3.

- Case F (future) : A10 ($B = 150$) \rightarrow A40 ($B = 350$) \rightarrow A70 ($B = 500$) \rightarrow A100 ($B = 1000$)
- Case P (present): O10 ($B = 150$) \rightarrow O40 ($B = 350$) \rightarrow O70 ($B = 500$) \rightarrow O100 ($B = 1000$)
- Case M (mixed): O10 ($B = 150$) \rightarrow O40 ($B = 350$) \rightarrow A70 ($B = 500$) \rightarrow A100 ($B = 1000$)

I consider that $C_n = 2$ and $C_u = 1$, which means that siting a new charging station costs as much as adding an additional charging port to an existing one. While the numeric values of the budget were set arbitrarily in this study, I have conducted sensitivity analysis on the budgets for Case F and Case P in the Supplementary Information Section H.

6.4.1 Siting charging stations for the case study

I first generate the demand using the method outlined in Section 6.3.2 for each scenario using the parameter settings in Table 6.3, which correspond to the different epochs in our three case studies. I assume that all EVs can travel $range = 200$ miles on a single full charge (similar to the base version of the Tesla Model 3) and the

⁴The budget B was set to keep the average waiting time for the vehicles approximately the same in the A scenario (Figure 6.10(b))

Table 6.3.
Scenarios for which EV charging stations are sited

Scenario	Fleet Size	AV Adoption	EV Adoption	Ride Sharing Adoption
A10	5500	1	0.1	1
A40	5500	1	0.4	1
A70	5500	1	0.7	1
A100	5500	1	1	1
O10	13500	0	0.1	0
O40	13500	0	0.4	0
O70	13500	0	0.7	0
O100	13500	0	1	0

charging ports are fast charging ones that can fully charge an EV in $t_f = 30$ minutes (Tesla, 2019). Non-AVs can also charge off-shift at a rate of $t_s = 180$ minutes (which is the time taken for a level 2 charger to fully charge a Tesla Model 3 Tesla (2019)). The *SOC* of the EVs at the beginning of the day was drawn from a triangular distribution with mode 0.6⁵.

Even though the total charging demand in both cases is approximately the same (in all cases the O scenarios have approximately 10% less total demand than their corresponding A scenario), we see from Figures 6.4 and 6.5 that the spatial distribution of the demands is very different for the two sets of scenarios (A and O). Spatially, the charging demands in the A scenarios are more spread out in the suburban regions

⁵I tested the mode of the triangular distribution by varying it between the values of 0 to 1 and found that, if I set the mode to 0.6, the *SOC* distribution at the beginning of the day is similar to distribution at the end of the day.

of NYC than the O scenarios (Figure 6.4). Temporally, the charging demands in the O scenarios have a single large peak at the beginning of the day, where the demands in the A scenarios have three smaller peaks (Figure 6.5). Because, in the O scenarios, the taxis that are off-shift are charged using level 2 chargers, it is highly likely that they come in-shift with full charge, and wouldn't need additional charging during their shift. On the other hand, the autonomous taxis in the A scenarios do not go off-shift and can only access the fast chargers that are placed in the model. These spatial and temporal differences in the charging demand distributions are likely to have a significant impact in the placement of charging stations.

For each of these scenarios, the set of public parking stations in New York City (NYC Open Data, 2016) is considered as the potential charging station location set \mathcal{L} . Parking lots are considered as ideal sites to locate charging stations since they are more likely to be located near points of interests for taxi customers and can serve as good locations to wait while searching for new passengers. Additionally, taxis cannot charge in the middle of a trip, because it would be unreasonable to expect the customers to wait while the taxi is charging.

I run 10 instances of the genetic algorithm to account for the randomness of heuristic algorithms 8 and select the best solution in each scenario. The Genetic algorithm is run for $I = 80$ iterations (since after this point the objective function value does not decrease at a rapid rate as per Section G of the Supplementary Information) with $G = 20$, and then the best charging station sites are reported, which are visualized in Figures 6.6 to 6.8.

Figures 6.6 to 6.8 show that the optimal charging station configurations in the A and O scenarios are very different. The corresponding objective function values are presented in Figure 6.10(a). From these results, it can be seen that, first, the genetic algorithm selects less charging stations with more charging ports in the O

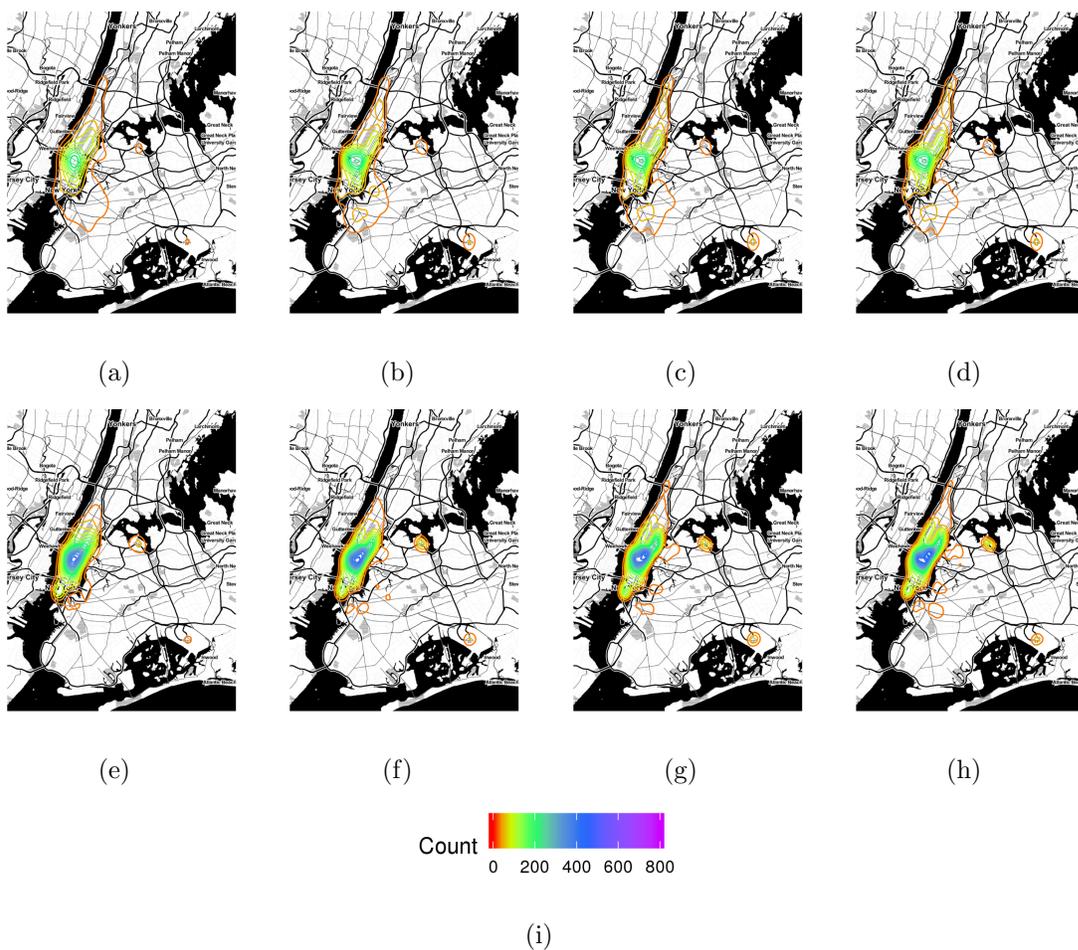


Figure 6.4. Spatial visualization of the charging demands for each scenario. The contours denote the demand distribution for that scenario and are superimposed on a map of New York City bounded by the coordinates $(-74.1, 40.55)$ and $(-73.7, 40.95)$. The demand concentrated in the Manhattan area.

scenarios, so that it can reduce the waiting time (T_w) during the large morning peak as T_w is the dominating contributor to the total wasted time (Figure 6.10(c) and 6.10(d)). On the other hand, the A scenarios have many more charging stations

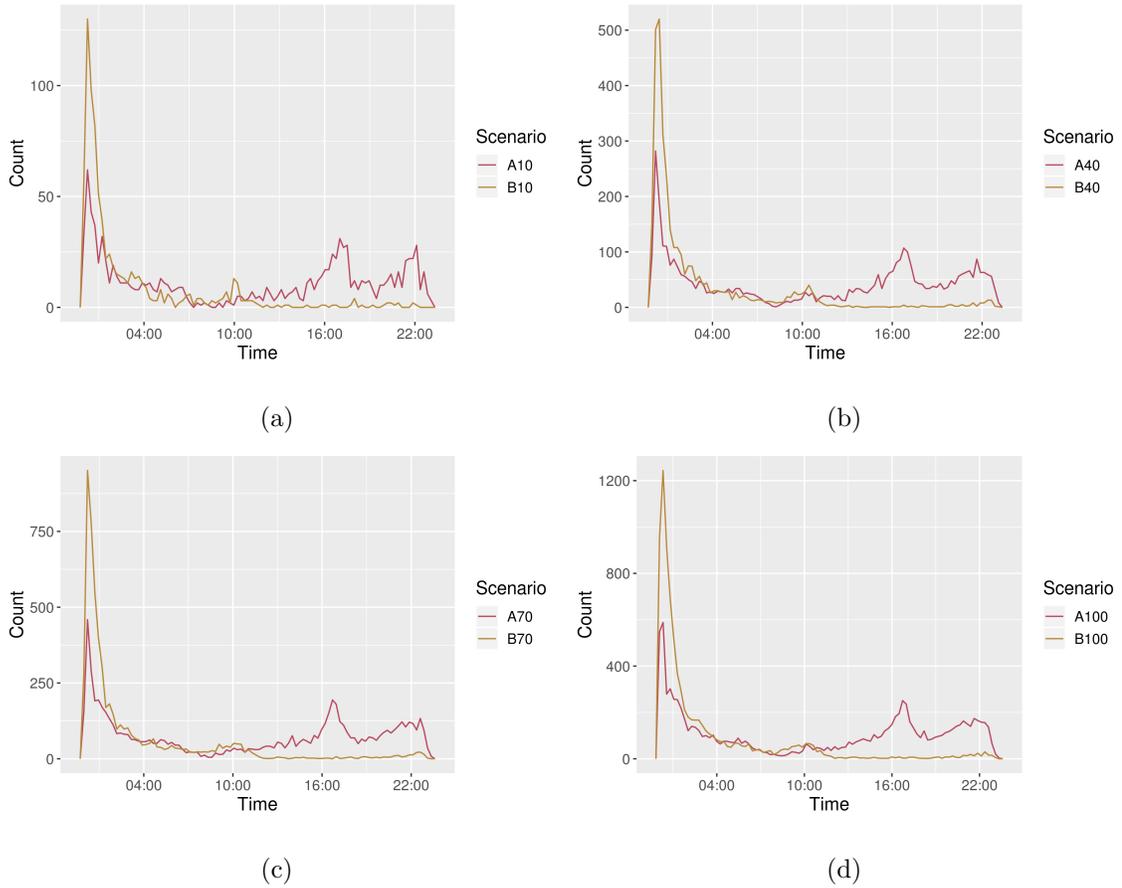
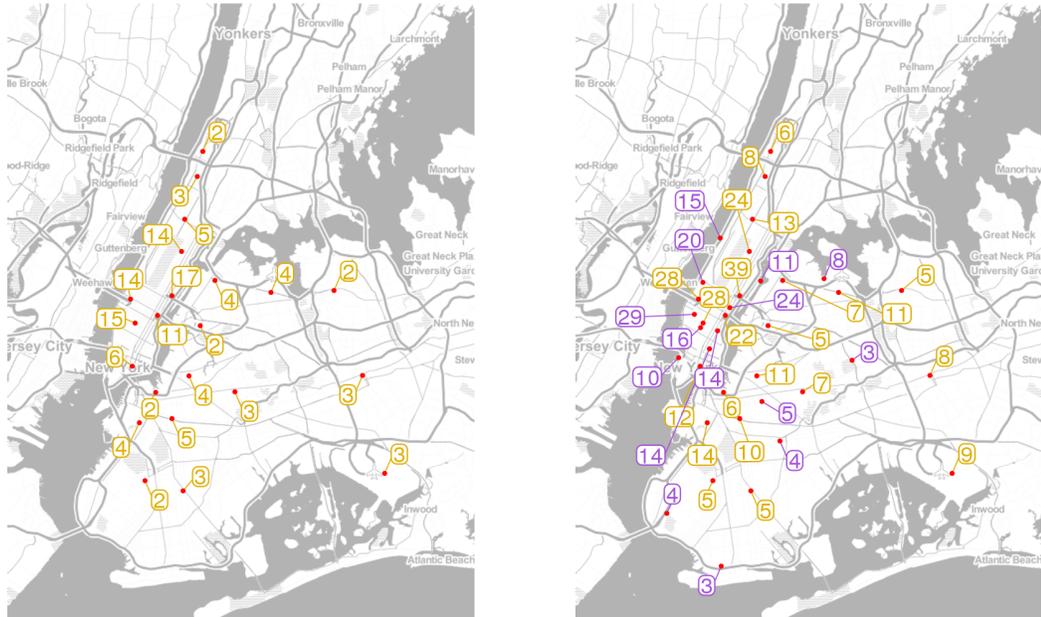


Figure 6.5. Count of unrestricted charging demand for every 15 minutes with different EV adoption scenarios: (a) 10%, (b) 40%, (c) 70%, and (d) 100%

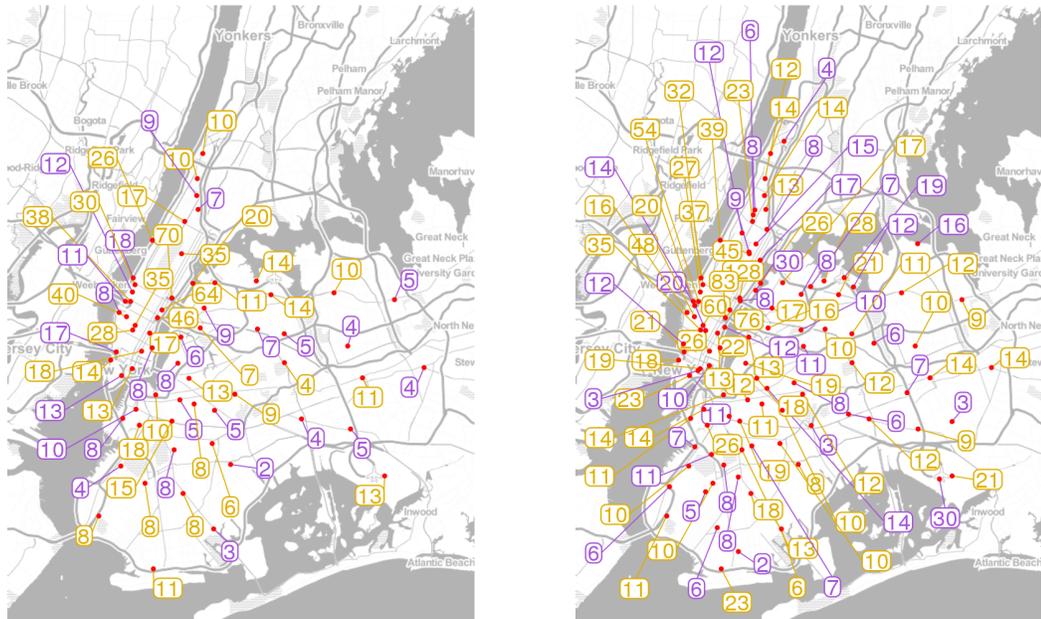
with lower number of charging ports, which would help minimize the time taken to travel to the charging stations (T_t), because, for the A scenarios, the driving time is the dominating contributor to the total wasted time (Figure 6.10(c) and 6.10(d)). Second, the objective function value reached is much larger (worse) for the O scenarios as compared to the A scenarios. From Figure 6.10(b), it can be seen that,

for the same level of EV adoption, the average wasted time due to charging for each taxi is much larger in the O scenarios. This indicates that, if we introduce electric taxis without autonomous driving and ride sharing, we would need more budget to site enough charging stations to maintain the same objective function value (charging infrastructure service quality) for the same level of EV adoption. Third, we see that the objective function values of A70 and A100 scenarios in case 3 are slightly higher than those in the A70 and A100 scenarios of case 1, because, in case 3, the charging stations sites for scenario A70, consider pre-existing charging stations of G40. Since the demand patterns of G40 and A70 are very different (see Figure 6.4 and 6.5), the sudden shift from O to A results in the system having to deal with more demands in the suburban regions of the city (where previously there was low demand).



(a)

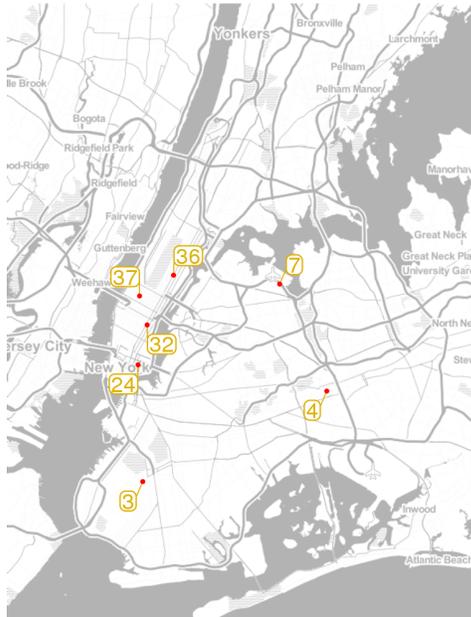
(b)



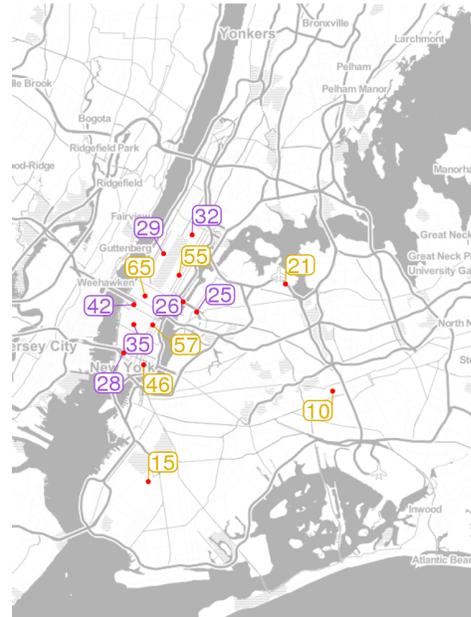
(c)

(d)

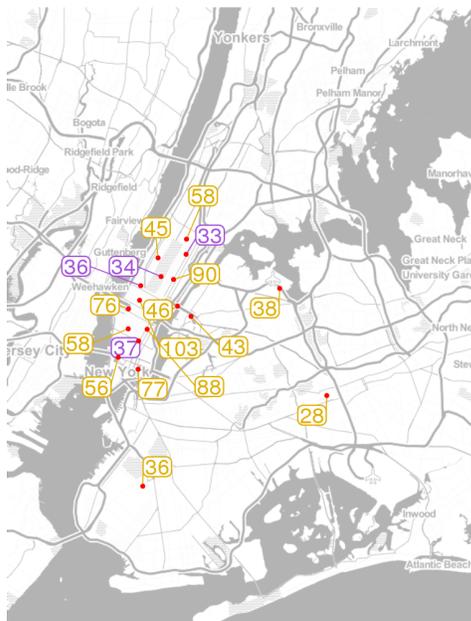
Figure 6.6. Future case : Best charging station siting plan as determined by the genetic algorithm. The red dots indicate the site and the numbers indicate the total number of charging ports at that station. The purple charging station labels indicate new stations built for that EV adoption level, while the yellow labels indicate station which exist from the previous epoch



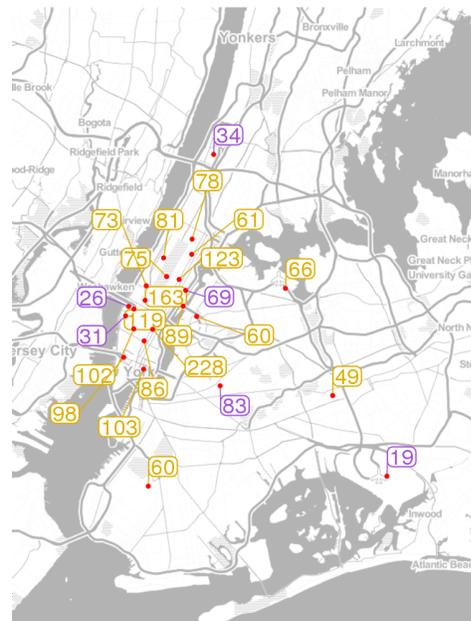
(a)



(b)

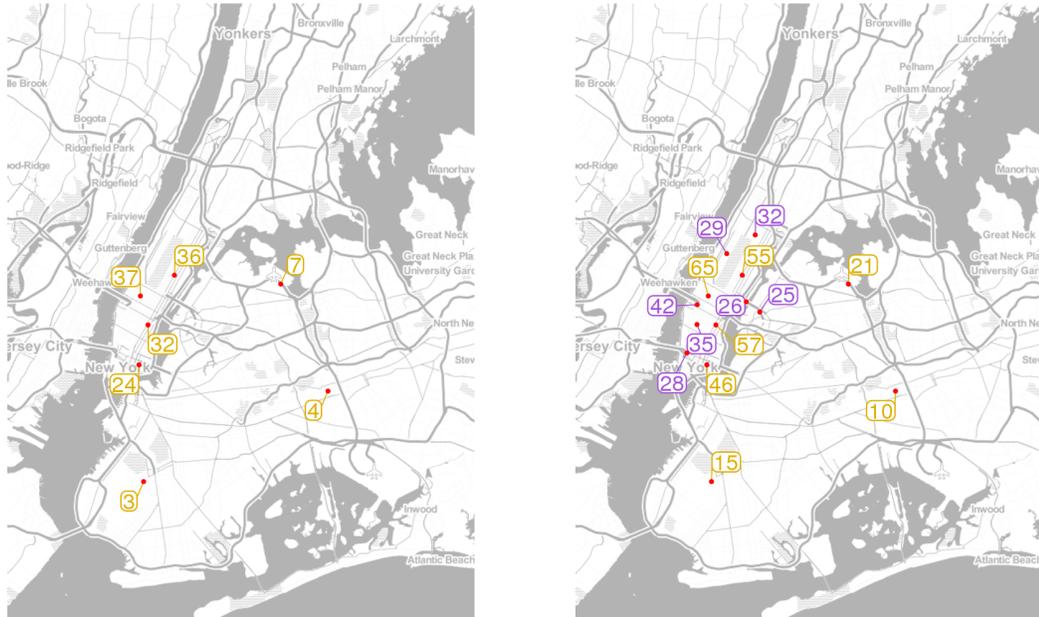


(c)



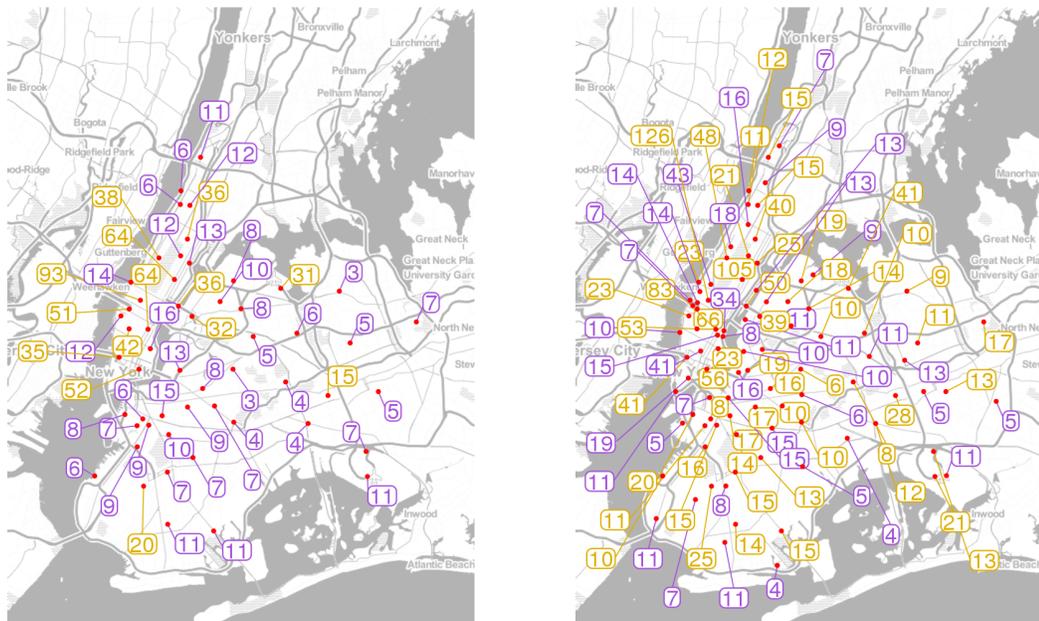
(d)

Figure 6.7. Present case: Best charging station siting plan as determined by the genetic algorithm. The red dots indicate the site and the numbers indicate the total number of charging ports at that station. The purple charging station labels indicate new stations built for that EV adoption level, while the yellow labels indicate station which exist from the previous epoch



(a)

(b)



(c)

(d)

Figure 6.8. Mixed case: Best charging station siting plan as determined by the genetic algorithm. The red dots indicate the site and the numbers indicate the total number of charging ports at that station. The purple charging station labels indicate new stations built for that EV adoption level, while the yellow labels indicate station which exist from the previous epoch

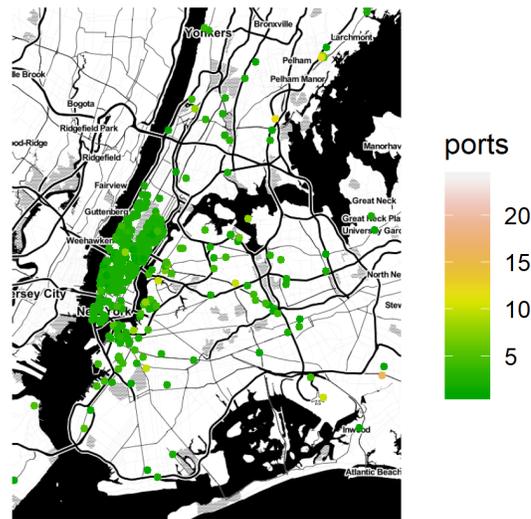


Figure 6.9. Present day charging station sites in NYC

In comparison to the present day charging station sites in NYC (Figure 6.9 obtained from US DOE (2019)) where the EV adoption levels are overall less than 2% of all the vehicles on road, there appear to be some similarities between the present day case (Figure 6.7). In both, the actual sites for public EV charging are more concentrated near Manhattan (the region where ridership demand is high). However, in the actual sites many of the charging stations located in suburban regions have high number of chargers. This difference can be observed because our case study focusses on NYC taxi operations, and the ridership demand for NYC taxis is very low in suburban regions. Alternatively, there may be a significant number of personal EVs in those suburban regions to justify a large number of charging stations.

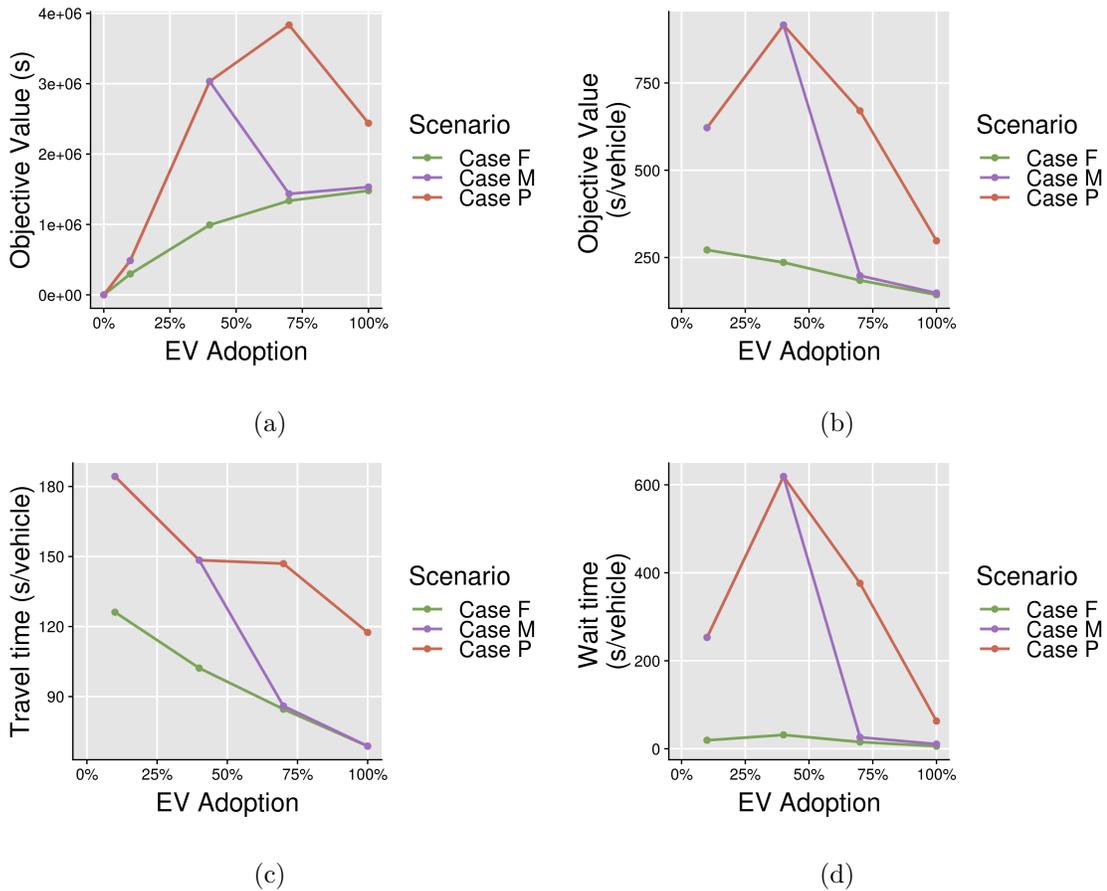


Figure 6.10. (a) Objective function values reached in different adoption scenarios, (b) Scaled objective function value based on the total number of EVs, (c) Scaled travel time, and (d) Scaled waiting time. The scaled objective function value is the ratio of the objective function value to the number of EV taxis in that scenario. The scaled objective function can be interpreted as the average time additional time that an EV taxi wastes in moving to and waiting at a charging station

6.4.2 Case Study Analysis

After siting the charging stations for each of these cases, I re-ran the PP-SAEV with the optimal charging station locations (as shown in Figures 6.6 to 6.8) as inputs

to analyze the impacts of increasing EV adoption on the performance of the taxi system.

6.4.2.1 Temporal charging trends

Since I used the unconstrained demand to site charging stations (as discussed in Section 6.4.1), I need to evaluate that, with the sited charging stations, how closely our optimal charging stations satisfy the EV charging demand. As shown in Figures 6.11 and 6.12, for the scenarios where the objective function value is relatively low (see Figure 6.10(b), 10% and 100% EV adoption in case P and all scenarios in case F), indicating that budget B is sufficient, the resulting charging load (Figure 6.5) closely matches that of the estimated demands (Figure 6.11), and the queues are relatively short (Figure 6.12). However, when the budget is not sufficient, the charging load in the morning peak is more spread out because more taxis need to queue at charging stations.

From the number of taxis charging throughout the day, I can also calculate the total power consumed for EV charging by multiplying the number of chargers that are in use at any given time with the power rating of the charger. Considering fast chargers with 120kW power rating (Tesla, 2019), I plotted the total power consumption of the city (Figure 6.13a) and the power consumption to support the charging of 100% EV adoption (Figure 6.13b) in each of the three cases. We can see that, the power consumption in the city will not be significantly affected due to the different scale, because the total number of taxis in NYC is relatively small. Additionally, the peak power consumption by the EV charging occurs during the time when the power consumption from other sources is low, hence we would not need to expand the capabilities of the electric power grid to account for EV taxis in NYC. How-

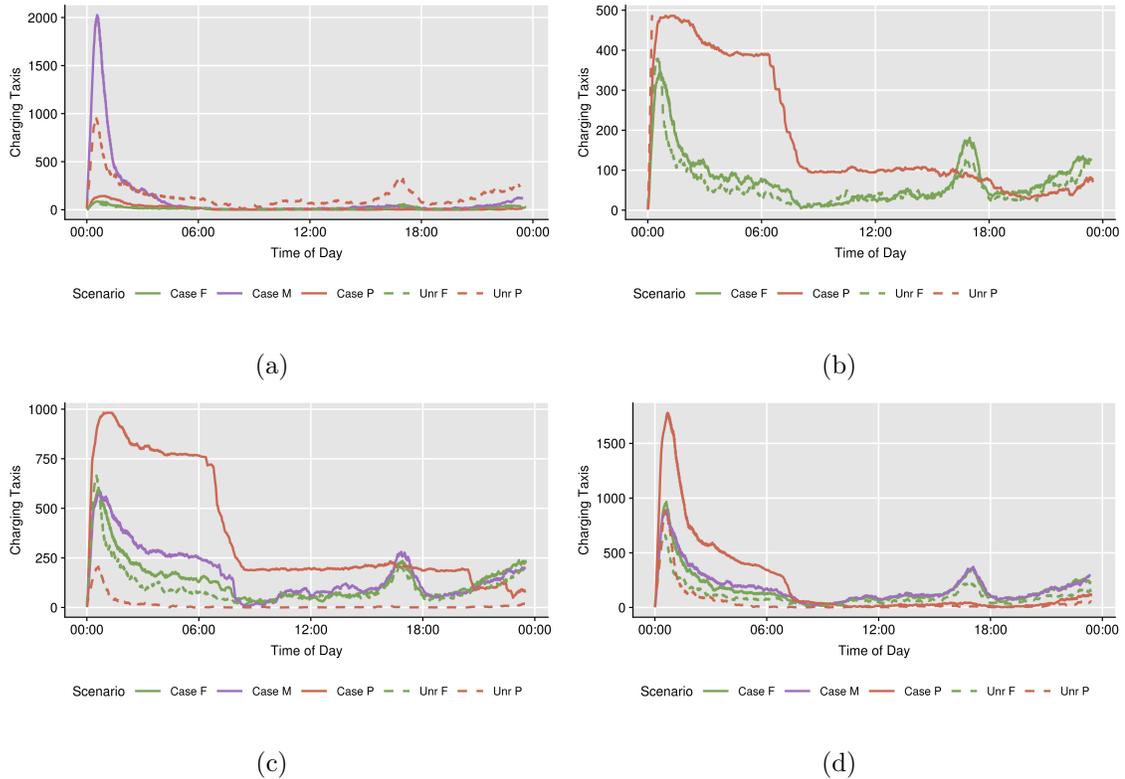


Figure 6.11. Comparison between the demands simulated based on the optimal charging station siting and the estimated unrestricted demands in terms of the number of vehicles charging at the charging station over time. The solid lines represent Case F, Case P, and Case M, while the dotted lines represent the unrestricted scenarios. (a)10%, (b)40%, (c) 70%, and (d) 100% EV adoption.

ever, it must be noted that this study only considers electric taxis and it is possible that personal EV adoption will also increase along with the adoption of EV taxis (Bloomberg NEF, 2018). Hence, a similar study would be needed to understand the overall impact on the grid due to widespread adoption of personal EVs.

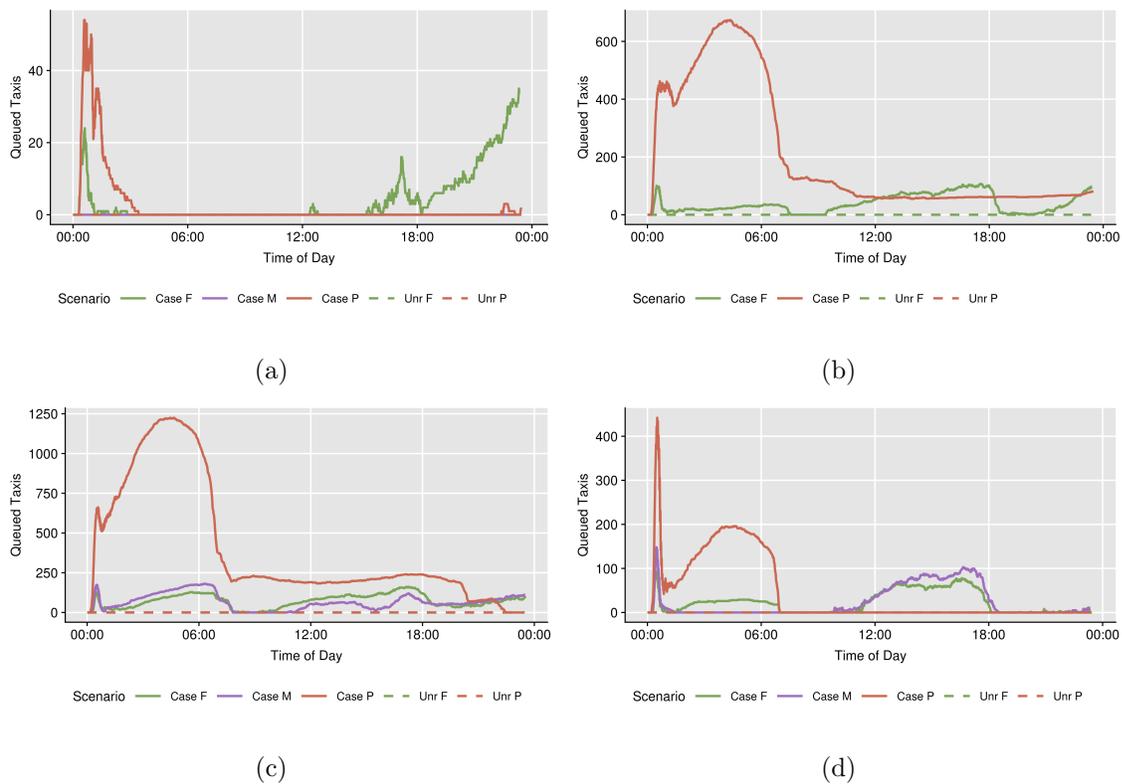
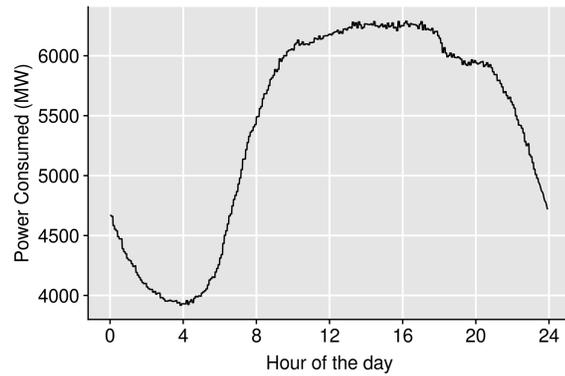


Figure 6.12. Count of vehicles queuing at the charging station. The solid lines represent Case F, Case P, Case M and the dotted lines represent the unrestricted scenario (a)10% (b)40% (c) 70% (d) 100%

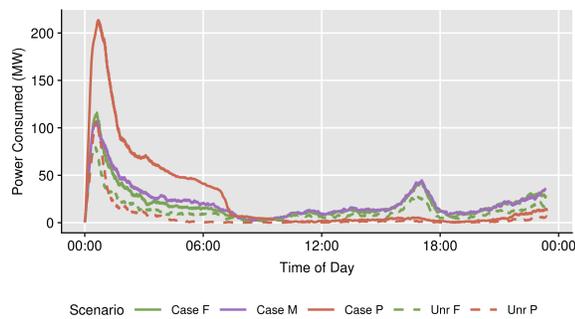
6.4.2.2 Service Level

For each scenario, I studied the service level (the fraction of riders served by the system) achieved during the peak demand period⁶ (Figure 6.14(a)) so that the effect of EVs on the system during rush hours could be understood. The service level

⁶The peak demand period in NYC is from 6:00pm to midnight (NYC DOT, 2014).



(a)



(b)

Figure 6.13. (a) The current power consumption profile of NYC (New York Independent System Operator, 2019), and (b) The power consumption by charging EVs for 100% EV adoption in the evaluated cases (color). The solid lines represent Case F, Case P, Case M and the dotted lines represent the unrestricted scenario (a)10% (b)40% (c) 70% (d) 100%

achieved during the peak charging demand period⁷ was also evaluated, because this would be the time that less vehicles would be available to serve customers (Figure

⁷From Figure 6.5, we see that the peak charging demand occurs at the beginning of the day (1:00am to 7:00am), when the demand for taxis is low.

6.14(b)). In order to understand the impact that EVs would have on the system, assuming they had access to unlimited charging infrastructure, I also plotted the service levels for scenarios that had access to all potential charging locations from set \mathcal{L} . We see that EVs have minimal effect on the service level in all cases for the peak charging time (1:00 am to 7:00 am). This indicates that, at the peak charging time, sufficient vehicles are available to serve riders. For the unrestricted charging case, more EVs are able to charge as compared to the restricted charging case, and hence the number of vehicles available to serve riders is less. However, in the peak demand period (6:00pm to 12:00am), we see that, for the O scenarios (Case P, and the G10 & G40 scenarios of Case M), there is little change in the service level. The reason is that, during the peak demand period, there is very little charging demand. Therefore, the number of vehicles that are available do not decrease by a large number by EV charging. For the A scenarios there is a larger decrease in the service level for increasing EV percentage. However, we see an increase in service when the percent EV increased from 70% to 100%. We can expect this increase in service level because the objective function value of the optimal sites in the A100 scenario is actually lower than that of the A70 scenario (Figure 6.3). This may be caused by the fact that I assigned the A100 scenario a budget that was relatively larger (for the same EV adoption increase, an additional budget of 500 was allocated) than the A70 scenario. We also see that the service level of the A70 and A100 scenarios of Case M is higher than the corresponding A70 and A100 scenarios of Case F. This happens because there are more vehicles charging during the peak demand period in Case F than in Case M (see Figure 6.11), reducing the service level.

Since there are limited number of charging stations and the EVs would need to relocate to charge at charging stations, the service levels in different parts of the city would change with increasing EV adoption. For the three cases, I plotted the

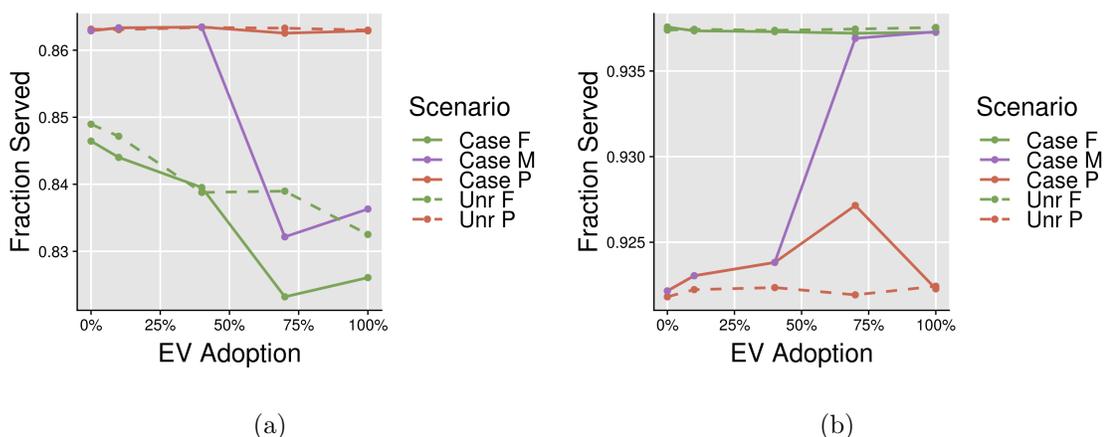


Figure 6.14. Service level for (a) rider demand peak (6:00pm to 12:00am) (b) charging demand peak (1:00 am to 7:00 am); The solid lines represent Case F, Case P, Case M and the dotted lines represent the unrestricted scenario (a)10% (b)40% (c) 70% (d) 100%

change in the service levels between 100% EV adoption and 0% EV adoption. In each case though, we observe no major spatial trend⁸ in the change in service levels throughout the city for the entire day (Figure 6.15). This result indicates that EV adoption is unlikely to significantly impact the taxi service in different regions of the city, as long as the charging stations are appropriately placed.

6.4.3 Environmental Impacts

In order to estimate the environmental impact of EV adoption in the NYC taxi fleet, I studied the total distance travelled by all vehicles (Figure 6.16(a)) distance

⁸In some of the cells we see a 100% change in service level. This occurs since the demand for taxi rides in that region is very low, and even a small change in the absolute number of served riders in that cell would result in a large change in service level

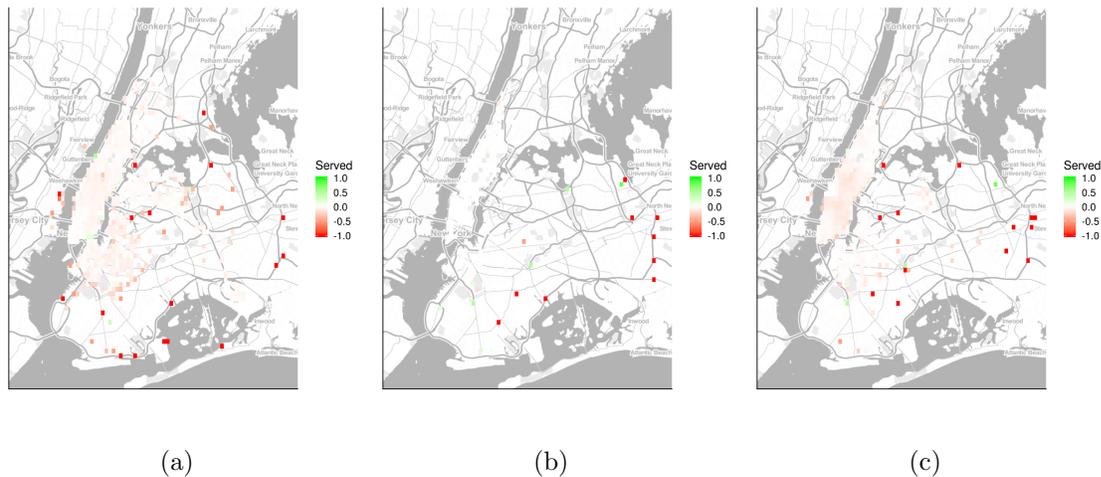


Figure 6.15. Change in service levels over the city between 100% EV adoption and 0% EV adoption for the entire day. The color in each cell corresponds to the change in service level, which is measured as the ratio of the change in the number of riders served by the two scenarios in that cell to the total number of riders in that cell

travelled by the EVs (Figure 6.16(b)). While the total distance decreases by about $1 \times 10^5 km$ throughout the day in both the F and P cases, we expectedly see an increase in the distance travelled by the EVs in each case. Additionally, we see that the total distance travelled between the P case and the F case is approximately different by $2 \times 10^5 km$ per day (similar to Section 3.3.3). This reduced distance is mainly on account of having 100% ride sharing adoption in the F case (and hence less total distance travelled for the same number of rides) as compared to the P case (0% ride sharing). In each case though, we see from the difference in travel distance between Case M and case P (for 70% and 100% EV adoption) that placing the charging stations optimally has a very small effect on the total travel distance. This is in line with the observations of Bauer et al. (2018) that the additional distance

travelled due to charging will likely not have a major impact on the total travel distance.

For ICEVs, every mile driven results in tailpipe emissions and is estimated to contribute 404 g of CO₂ (US EPA, 2019b). However, for electric vehicles, while the tailpipe emissions are 0, CO₂ emissions are produced at the source of electricity generation. For New York City, this is estimated to be 110 g (US EPA, 2019a) of CO₂ for a mid range Tesla Model 3 (The Tesla model 3 has a range closest to 200 miles per charge as considered in our model). For AVs, we consider an additional 9% reduction in CO₂, which is predicted by Gawron et al. (2018) as a result of potential changes in driving efficiencies. Using these conversion factors, I estimated the daily CO₂ emissions for our three cases (Figure 6.17). While increasing the EV adoption from 0 to 100%, the CO₂ emissions reduced by about 861 Tonnes (74%) (this is consistent with Bauer et al. (2018)) in the A scenario and about 1100 Tonnes (78%) in the O scenario. Interestingly, due to a larger increase in the distance travelled by EVs for the P scenario, increasing EV adoption will result in a larger reduction in CO₂ emissions than the F scenarios.

6.5 Conclusion and Future Scope

In this chapter, I presented a method to site charging stations considering a fixed budget, while optimizing the number of charging stations and the number of charging ports at each station. The proposed method considers queuing at charging stations and the presence of existing charging infrastructure. The charging demand was estimated using an agent-based model to simulate shared autonomous electric vehicles which have flexible charging rules. A genetic algorithm was used to find the optimal charging infrastructure expansion plan.

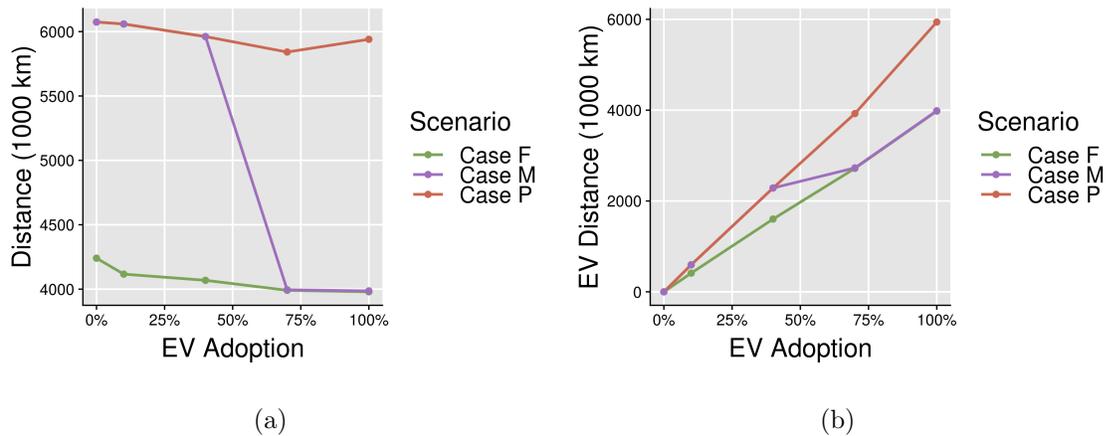


Figure 6.16. The distance travelled by (a) all the taxis (b) Only the electric taxis with increasing EV adoption

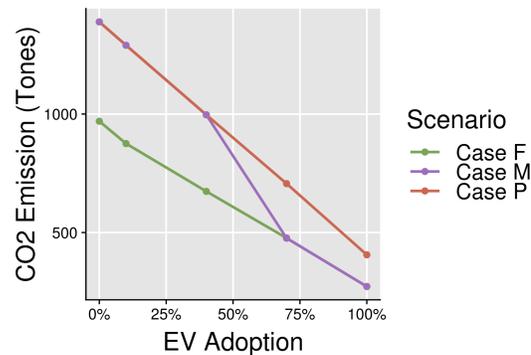


Figure 6.17. CO₂ emissions in Tonnes for increasing EV adoption

I used the proposed method on three case studies where the EV adoption was increased from 10% to 100% in four steps. By studying the outputs from running the case studies with the optimal charging station identified using the genetic algorithm, I found that for non AV-RS (P) case, there is no overall change in service levels due

to the presence of EVs in the system. For the AV-RS case (F), I found that the overall service level decreased by approximately 2%. Additionally, I estimated that EV taxis would be able to reduce CO₂ emissions by about 790 Tonnes per day in the F scenario and 1100 Tonnes in the P scenario.

While this study improves the existing charging station siting method from multiple aspects, it has a few limitations that I would like to note to shed lights for future research. First, I sited charging stations for one epoch at a time considering the charging stations that were sited in the previous epoch. Future work could look at siting charging stations for multiple periods at the same time considering the uncertainty in future AV, ride sharing, and EV adoption. Second, more research could be done in finding exact solutions to our problem. Genetic algorithms do not guarantee global optimal. It is possible that better solutions can be found. Third, the model can incorporate more complex charging behaviors. Our simulation model considers that taxis would only go to the nearest charging station and the decision of whether to seek charging opportunities only depended on the *SOC* and the distance to the charging station. In reality, the taxis would consider a number of factors in evaluating whether to charge or not, or select an alternative charging location (for example, the queue length and the expected waiting time at the charging station, the potential of finding new riders at the current location of the taxi and at the charging station, and the cost of charging etc.) Fourth, since I was siting charging stations for a fleet of taxis, I did not consider a mix of charging stations with different power ratings. While assuming that all charging stations have the same power rating may be reasonable in this case, it may not be reasonable to make the same assumption for other cases (for example personal EVs). Additionally, I considered the cost of building/expanding charging stations at different sites to be identical. However, the cost of building the same type of charging stations at different locations could vary

considerably based on prevailing land prices and the types of surface on which the charging stations are installed (NYCTL, 2013). Studies that include more accurate charging stations building costs can generate more realistic solutions. However, this will depend on information such as land price, surface type, and existing electric service to be available at the large-scale for all candidate sites.

7. CONCLUSION

This dissertation uses agent-based modeling to better understand emerging transportation systems that are used for personal mobility. These emerging transportation systems could be comprised of a mixed fleet of autonomous vehicles, electric vehicles and may have a varied adoption of ride sharing. Additionally, the preferences of the riders towards ride sharing in the system may also be different. The understanding of such systems is crucial to formulating policies that guide the adoption of these new technologies (ride sharing, autonomous vehicles and electric vehicles).

The major contribution of this dissertation is in building the Parameterized Preference-based Shared Autonomous Electric Vehicle (PP-SAEV) model, which is an agent-based model that simulates systems of SAEVs. The model is parameterized so that it is able to evaluate varying adoption levels of AVs, EVs and RS. The PP-SAEV model also includes settings to define the heterogeneity of riders towards choosing a shared ride or a non-shared ride, and can also accept heterogeneous charging rules for EVs. The output from the model can help estimate the aggregate performance of a system with a particular parameter setting in terms of service level, waiting time, riding time, percent shared rides, extra distance travelled by riders, occupancy of vehicles and the distance travelled by the vehicles. Additionally, more specific inferences can also be made from the model output for example, the spatial and temporal service patterns of riders, the status changes of vehicles through the day, locations of charging demand, and potential routes where additional demand can be induced.

By comparing the outputs of several scenarios against each other, the impacts of changing a particular parameter setting can be studied. The dissertation also introduces a method to comparatively analyze multiple scenarios by identifying those scenarios that are comparable to each other based on a selected performance metric (for example, service level). It is important to find scenarios that are comparable to each other and compare their secondary performance metrics (for example waiting time, or environmental impacts) so that the systems are being compared on a fair ground. Using meta-model based simulation optimization, comparable scenarios can be identified even when there are many continuous and discrete parameter settings. By identifying comparable scenarios researchers are able to identify multiple scenarios that can be implemented and then evaluate these options on the basis of other important parameter settings.

The dissertation introduces a new method to site charging stations for electric vehicles when the budget to site charging stations is fixed. The method selects charging stations from potential charging station locations and allocates the number of ports to each charging station, considers queuing of EVs at charging stations, and considers charging stations already in the system while siting new ones.

The key results that were obtained by applying the methods developed in this dissertation to a case study of NYC taxis are discussed in Section 7.1. The insights that these results can provide for system planners are discussed in Section 7.2. While this dissertation does improve our understanding of these sustainable transportation technologies, it also has some limitations which are outlined in Section 7.3.

7.1 Key Results from Case Studies

This section outlines the key results that were obtained from running the PP-SAEV model with several parameter variations. In each case study, the parameters of the PP-SAEV were varied so that the questions that were asked in Section 1.5 were addressed.

The PP-SAEV was used to identify scenarios that had comparable service level with respect to the present day taxi operation using a case study of New York City taxis for systems with AVs and non-AVs by varying the parameters of the PP-SAEV in discrete steps. It was found that 8000 traditional vehicles with 100% ride sharing adoption or 5500 autonomous vehicles with 100% ride sharing adoption could service the same demand as a system of 13500 traditional taxi cabs with no ride sharing. Comparable scenarios were also developed for systems which had a mixture of AVs and non-AVs in varying proportions, for different taxi capacities. Using these comparable scenarios, a fleet size prediction model was developed and it was found that to reduce the fleet size by 1000 vehicles, the system operator could increase AV adoption by approximately 25% or increase the RS adoption by 35% for taxis of capacity 4. For taxis of capacity 6, about 50% AV adoption or 37% RS adoption could bring about the same reduction in taxi fleets. However, the reduction in the number of vehicles in the system could result in 1) increased wait times and 2) a reduction of service in suburban areas.

The service level determines the usability of the transportation system, and hence it is important to understand how different system parameters (for example rider heterogeneity and EV adoption) impact the service level. While EV adoption was not found to have a significant impact on service level, the rider heterogeneity did impact the service level significantly. Mixture experiments were used to vary the

proportions of the riders in the PP-SAEV to understand the impact of different mixes of riders in a SAV system. The analysis shows that system with about 70% riders that are either indifferent to sharing or those that prefer sharing with about 30% of riders accepting only shared rides have the best service level for 6000 SAVs. Increasing the proportion of non-sharing riders could potentially reduce the service level of a RS system by 17%. Individual service levels for each rider type were also analyzed, and sets of rider types that positively and negatively affected the service quality of the other were identified. The riders with the most flexible ride sharing preference were least affected in terms of service level by changing the proportions of other rider types in the SAV system. These insights can help formulate policies to design ride sharing systems that are able to target riders with a particular sharing preference.

While analyzing the CO₂ emissions of these comparable scenarios, it was found that AVs with RS could reduce approximately 866 metric tonnes of CO₂ per day in capacity 4 taxis mainly by reducing the total distance travelled by all the vehicles. Additionally, capacity 4 taxis have the most potential to reduce CO₂ emissions when compared to taxis with capacity 2 or 6. While studying the impact of rider heterogeneity on the RS system, no significant impact of varying rider compositions was found on the CO₂ emissions. By introducing EVs in SAV systems we are able to reduce approximately 810 metric tonnes of CO₂ emissions and by introducing EVs in traditional taxi systems a reduction of 1100 metric tonnes of CO₂ can be achieved.

Ride sharing systems are not able to reach maximum capacity thus limiting their efficiency in serving more riders. In order to further understand the limitations of ride sharing, the status of the taxis through the day was studied and used composition data analysis techniques to analyze the reasons for sharing limitations from the taxi status data. Using hierarchical clustering groups of scenarios were identified which

had similar patterns of taxi statuses and it was found that for taxis with a fleet size of 2, ride sharing would not be successful due to most of the taxis either reaching their maximum capacity, or being unable to find new riders to share with since they may have more than two persons in their group. Taxis with fleet size 4 and 6 are potentially capable to add additional riders to their vehicle. However, since appropriate ride shares may not be available, taxis of size 4 and 6 are unable to reach their maximum capacity. This indicates that if appropriate incentives are provided to encourage ride sharing on similar routes taxis with capacities 4 and 6 may be able to serve additional demand. Because, taxis with capacity 4 and 6 have approximately similar taxi status patterns, it would be better to use taxis with capacity 4 for ride sharing since taxis with higher capacities would emit more tailpipe emissions.

The adoption of ride sharing has a significant impact on the average waiting time of riders in an SAEV system. The waiting time of the riders increased by approximately 30% when switching from 0% ride sharing participation and 100% ride sharing participation. By building a regression model of waiting time as a function of fleet size, AV adoption, RS adoption, capacity, and deviation tolerance it was found that by increasing the fleet size by 1000 taxis or increasing the AV adoption by 25% the average waiting time for the riders could be reduced by 30s for capacity 4 taxis. The waiting time of riders is also affected by the sharing strategy of the riders and the overall composition of rider types (sharing preferences) in the system, and in general, the trends in the dependence of waiting time on the rider type composition were negatively correlated with the trends of service level.

7.2 Insights for SAEV System Development

The PP-SAEV model and the insights as discussed in Section 7.1 can guide policy makers to design future transportation systems which may have varying adoption levels of ride sharing, electric vehicles and autonomous vehicles.

- Policy makers can test various “what if” scenarios using the PP-SAEV model developed. The model outputs fine grained status data that can be used to find aggregated level system performance metrics for example the service level, average waiting time, percent of rides shared. Additionally, the fine grained status output from the PP-SAEV can be used to make inferences at spatial and temporal levels for system key performance indicators.
- By analyzing scenarios that are comparable to each other (for example, in Chapter 3 and 4), relationships between other system performance indicators such as average waiting time, environmental impact and system ride sharing efficiency can be identified. Based on these system performance indicators, a policy maker can choose to drive efforts to achieve a certain adoption level of AVs and RS.
- By inducing additional shared rides (as discussed in Section 4.4.3) from the suburban regions (where the demand is low) to the city center (where the demand is high), potentially extra rides could be served by the system. Thus policy makers could incentivize riders traveling in the same direction but using other modes of transportation (for example personal vehicle) to instead adopt the ride sharing system as a preferred transit mode. This could be done by lowering rates for such rides.

- The heterogeneous preferences of riders (as discussed in Chapter 5) do impact the system performance (in terms of service level, waiting time and ride sharing efficiency) as well as the individual riders service quality (service level and waiting time). If the rider types of a rider can be identified, the system performance can be studied more accurately using the PP-SAEV model, since it can account for these rider types. Additionally, in order to achieve a certain system performance, the system operators can use monetary or other incentives to modify the sharing preference of riders to bring the overall composition of rider types towards a desired level.
- Charging stations can be sited for a fleet of electric vehicles given a fixed budget using the method developed in Chapter 6, so as to cause least inconvenience to the taxi fleet. Also, by studying the switchover case (switch from non AV-RS to AV-RS system at 70% EV adoption) I showed that such a sudden adoption of future technologies could increase the inconvenience caused to EV taxis that seek charging when compared to the AV-RS case. In order to minimize this inconvenience, if future systems are likely to have high AV and RS adoption levels this switch-over should be made as early as possible by promoting fast adoption of AV-RS systems as compared to EVs to avoid having greater constraints in siting charging stations and thus having more inconvenience to EVs looking to charge.

7.3 Limitations and Future Research Directions

The limitations of this dissertation and potential future research directions are listed below. Some of these limitations can be addressed using the PP-SAEV model

itself or by minor extensions, while others may need additional research work beyond the scope of the current PP-SAEV.

- The causes and impacts of induced rider demand need to be analyzed. It is possible that due to incentive policies (lowering taxi fares for shared rides), many additional riders may be added to the system and it is important to analyze the impact of this induced demand on the system performance. In Chapter 4 I inferred that if the induced demand is along existing shared routes, it may be possible to accommodate some additional sharing riders. The demand supplied to the PP-SAEV can be tested with varying fractions of induced demand to see how the system performance parameters change. Additionally, research is needed into how new policies can control this induced demand.
- In order to study the impacts of incentives, a cost structure for using the system should be incorporated in the PP-SAEV model. In reality, the adoption of ride sharing, electric vehicles and autonomous vehicles would depend upon the costs that are associated with using these systems. For example, in order to incentivize riders to share rides or adopt a certain rider type sharing strategy, system operators would have to subsidize shared rides appropriately. If electric vehicles are a part of the system, planners may want to implement a variable pricing scheme for charging at different charging stations in order to balance the utilization. Studies that consider such pricing schemes and model the demand according to these schemes would provide a better understanding of the system.
- Although the PP-SAEV model does match riders with taxis for shared and non-shared rides, it makes these matching decisions locally. There have been several optimization algorithms to find optimal matches in literature (Lin et al.,

2012; Santos and Xavier, 2015; Hosni et al., 2014; Mahmoudi and Zhou, 2016; Masoud and Jayakrishnan, 2017) which can potentially find better matches thus potentially allowing a further reduction in fleet size to serve the same rider demand. However, neither of these algorithms have been shown to have practical inferential applications on real-world road networks due their long running time. Research is needed to find globally optimal real-time matching algorithms for shared autonomous electric vehicles and how they would improve system performance.

- To demonstrate the use of the PP-SAEV model in making inferences for SAEV systems in cities, this dissertation uses a case study of NYC taxis. If input data is available or input data can be approximately generated, multiple cities can be compared to see if there are any peculiar features of cities that enhance the sustainability of these new technologies.
- All the taxis in a single scenario had taxis with the same capacity ($Capacity = 4$ in Chapter 3, 6 and 5 and $Capacity = 2, 4, 6$ in Chapter 4). However, it is possible to have taxi fleets where each taxi may have a different seating capacity. For example, in areas where the number of rides shared are higher, and additional ride sharing demand can be served, larger capacity vehicles can be introduced. Research is needed to find how these heterogeneous sized taxi fleets could help improve system performance
- Although in Chapter 4, varying adoption of RS and AV is considered, this dissertation does not consider varying the adoption of EVs, AVs and RS all at the same time. In Chapter 4, I varied the adoption of RS and AV but did not consider changing adoptions of EVs since in order to extend this analysis to

electric vehicles, at each step charging stations would have to be sited (using the method developed in Chapter 6) thus making such an extension not trivial. An experimental design study that incorporates the analysis of electric vehicles and performs similar analysis could provide an understanding of the combined effects of EV, AV and RS adoption changes.

- Chapter 5 considers heterogeneous preferences in ride sharing. However, it was noted that the sharing preferences defined in Chapter 5 actually depended on demographic and logistical factors of the riders. A study that understands the relationship between these factors of the riders and their sharing preference may provide a more direct intuition regarding how incentives should be designed to produce a desired effect (for example,) reduce waiting time or increase the service level for a particular rider type in the ride sharing system.
- Charging stations were sited by considering the demand generated by fixing the charging utility function (Equation A.1) for each rider in Chapter 6. However, each taxi may have their own charging preferences with regards to trading off waiting in queue at a nearer charging station vs driving further in search of a potentially empty charging station. Further, each taxi may have a different *SOC* threshold beyond which they would like to charge. These thresholds likely impact optimal charging station placement, and by studying the impact of these heterogeneous charging rules on a system with electric vehicles could lead to better insights on the system performance parameters.
- The optimization model in Chapter 6 considered siting charging station for increasing levels of EV adoption considering each EV adoption level sequentially. However, if charging stations are sited in such a way at each stage of

EV adoption without considering its impact on future stages of EV adoption, we may not get an optimal charging station configuration at a later stage. However, since we may not be able to clearly predict the future EV/AV/RS adoption states, it might be risky to site charging stations for different levels of these adoptions. Research is needed into studying how an optimal charging station configuration would change considering variability in adoption of new technologies like AV and RS.

In summary, future systems of shared autonomous electric vehicles offer great potential to reduce carbon emission by reducing VMT, and lowering tailpipe emissions. The proposed PP-SAEV model is a helpful tool to analyze systems with various mixes of AV, EV and RS adoption and can guide decision making. When evaluating system performance of several scenarios, comparable scenarios can help discover important trends relating the performance indicators to the input parameters. The proposed meta-model based simulation optimization method can help identify such comparable scenarios while using less computational resources. Large capacity vehicles may not be environmentally efficient because they emit more GHG emissions per mile and ride sharing efficiency can be limited by low sharing tolerance and spatially variable demand density. Heterogeneous sharing preferences of riders should be considered while modeling SAEV systems since they do have an impact on the service quality of riders. Last, EVs, though environmentally friendly, require an extensive charging station network and while evaluating an SAEV system, it is important to site these charging stations for incremental adoption of EVs. EV charging station siting models should consider queuing at stations, the locations of existing charging stations and budget limitations while determining the appropriate charging station sites. The genetic algorithm based approach used in this dissertation incorporates

these considerations and can be used to site charging stations for EVs. Future work needs to be done in exploring the impact of a gradual adoption of EV, AV and RS in SAEV systems.

8. REFERENCES

Bibliography

- Agatz, N., Erera, A. L., Savelsbergh, M. W., and Wang, X. (2011). Dynamic ride-sharing: A simulation study in metro atlanta. *Procedia-Social and Behavioral Sciences*, 17:532–550.
- Alemi, F., Circella, G., Handy, S., and Mokhtarian, P. (2018). What influences travelers to use uber? exploring the factors affecting the adoption of on-demand ride services in california. *Travel Behaviour and Society*, 13:88 – 104.
- Alhazmi, Y. A., Mostafa, H. A., and Salama, M. M. (2017). Optimal allocation for electric vehicle charging stations using trip success ratio. *International Journal of Electrical Power & Energy Systems*, 91:101–116.
- Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., and Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, page 201611675.
- Amaran, S., Sahinidis, N. V., Sharda, B., and Bury, S. J. (2016). Simulation optimization: a review of algorithms and applications. *Annals of Operations Research*, 240(1):351–380.
- Amirkiaee, S. Y. and Evangelopoulos, N. (2018). Why do people rideshare? an experimental study. *Transportation research part F: traffic psychology and behaviour*, 55:9–24.

- Ankenman, B., Nelson, B. L., and Staum, J. (2010). Stochastic kriging for simulation metamodeling. *Operations Research*, 58(2):371–382.
- Arslan, O. and Karaşan, O. E. (2016). A benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles. *Transportation Research Part B: Methodological*, 93:670–695.
- Barbati, M., Bruno, G., and Genovese, A. (2012). Applications of agent-based models for optimization problems: A literature review. *Expert Systems with Applications*, 39(5):6020–6028.
- Barth, M. and Todd, M. (1999). Simulation model performance analysis of a multiple station shared vehicle system. *Transportation Research Part C: Emerging Technologies*, 7(4):237–259.
- Barton, R. R. (2009). Simulation optimization using metamodels. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 230–238. IEEE.
- Bauer, G. S., Greenblatt, J. B., and Gerke, B. F. (2018). Cost, energy, and environmental impact of automated electric taxi fleets in manhattan. *Environmental Science & Technology*, 52(8):4920–4928. PMID: 29589439.
- Bazzan, A. L. and Klügl, F. (2014). A review on agent-based technology for traffic and transportation. *The Knowledge Engineering Review*, 29(3):375–403.
- Bloomberg NEF (2018). Electric Vehicle Outlook 2018 — Bloomberg NEF. <https://about.bnef.com/electric-vehicle-outlook/toc-download>. Accessed : 10/11/2018.

- Boesch, P. M., Ciari, F., and Axhausen, K. W. (2016). Autonomous vehicle fleet sizes required to serve different levels of demand. *Transportation Research Record*, 2542(1):111–119.
- Box, G. E. and Behnken, D. W. (1960). Some new three level designs for the study of quantitative variables. *Technometrics*, 2(4):455–475.
- Box, G. E. and Hunter, J. (1961a). The 2 k—p fractional factorial designs part ii. *Technometrics*, 3(4):449–458.
- Box, G. E. and Hunter, J. S. (1961b). The 2 k—p fractional factorial designs. *Technometrics*, 3(3):311–351.
- Box, G. E. and Wilson, K. B. (1992). On the experimental attainment of optimum conditions. In *Breakthroughs in statistics*, pages 270–310. Springer.
- Box, J. and Wilson, W. (1951). Central composites design. *JR Stat Soc*, 1:1–35.
- Brandon Greenwell (2016). Pdp: An R Package for Constructing Partial Dependence Plots. <https://journal.r-project.org/archive/2017/RJ-2017-016/index.html>. Accessed : 4/14/2017.
- Brandstätter, G., Kahr, M., and Leitner, M. (2017). Determining optimal locations for charging stations of electric car-sharing systems under stochastic demand. *Transportation Research Part B: Methodological*, 104:17–35.
- Brownell, C. and Kornhauser, A. (2014). A driverless alternative: fleet size and cost requirements for a statewide autonomous taxi network in new jersey. *Transportation Research Record: Journal of the Transportation Research Board*, (2416):73–81.

- Cai, H., Jia, X., Chiu, A. S., Hu, X., and Xu, M. (2014a). Siting public electric vehicle charging stations in beijing using big-data informed travel patterns of the taxi fleet. *Transportation Research Part D: Transport and Environment*, 33:39 – 46.
- Cai, H., Jia, X., Chiu, A. S., Hu, X., and Xu, M. (2014b). Siting public electric vehicle charging stations in beijing using big-data informed travel patterns of the taxi fleet. *Transportation Research Part D: Transport and Environment*, 33:39 – 46.
- Cai, H. and Xu, M. (2013). Greenhouse gas implications of fleet electrification based on big data-informed individual travel patterns. *Environmental science & technology*, 47(16):9035–9043.
- Cai, H., Zhan, X., Zhu, J., Jia, X., Chiu, A. S., and Xu, M. (2016). Understanding taxi travel patterns. *Physica A: Statistical Mechanics and its Applications*, 457:590–597.
- Caulfield, B. (2009). Estimating the environmental benefits of ride-sharing: A case study of dublin. *Transportation Research Part D: Transport and Environment*, 14(7):527–531.
- Chen, H. and Rakha, H. A. (2016). Multi-step prediction of experienced travel times using agent-based modeling. *Transportation Research Part C: Emerging Technologies*, 71:108 – 121.
- Chen, T. D., Hanna, J. P., and Kockelman, K. M. (2016a). Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. *Transportation Research Part A*, 94:243–254.

- Chen, T. D. and Kockelman, K. M. (2016a). Carsharing's life-cycle impacts on energy use and greenhouse gas emissions. *Transportation Research Part D: Transport and Environment*, 47:276–284.
- Chen, T. D. and Kockelman, K. M. (2016b). Management of a shared autonomous electric vehicle fleet: Implications of pricing schemes. *Transportation Research Record: Journal of the Transportation Research Board*, (2572):37–46.
- Chen, T. D., Kockelman, K. M., and Hanna, J. P. (2016b). Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. *Transportation Research Part A: Policy and Practice*, 94:243–254.
- Choi, H., Shin, J., and Woo, J. (2018). Effect of electricity generation mix on battery electric vehicle adoption and its environmental impact. *Energy policy*, 121:13–24.
- Chow, J. Y. and Regan, A. C. (2014). A surrogate-based multiobjective metaheuristic and network degradation simulation model for robust toll pricing. *Optimization and Engineering*, 15(1):137–165.
- Clean Technica (2018). 25 Factors That Will Affect EV Adoption (Part 1) — CleanTechnica. <https://cleantechnica.com/2018/05/05/25-factors-that-will-affect-ev-adoption-part-1/>. Accessed : 3/30/2019.
- Cornell, J. A. (1990). *Experiments with mixtures : designs, models, and the analysis of mixture data*. Wiley series in probability and mathematical statistics. Applied probability and statistics section. Wiley, New York, 2nd ed.. edition.
- Cornell, J. A. (2011). *Experiments with mixtures: designs, models, and the analysis of mixture data*, volume 403. John Wiley & Sons.

- Davidov, S. and Pantoš, M. (2017). Planning of electric vehicle infrastructure based on charging reliability and quality of service. *Energy*, 118:1156–1167.
- Delhomme, P. and Gheorghiu, A. (2016). Comparing french carpoolers and non-carpoolers: Which factors contribute the most to carpooling? *Transportation Research Part D: Transport and Environment*, 42:1–15.
- Department of Energy (2018). Electric Vehicle Basics — Department of Energy. <https://www.energy.gov/eere/electricvehicles/>. Accessed : 4/14/2018.
- Ding, H., Benyoucef, L., and Xie, X. (2005). A simulation optimization methodology for supplier selection problem. *International Journal of Computer Integrated Manufacturing*, 18(2-3):210–224.
- d’Orey, P. M. and Ferreira, M. (2014). Can ride-sharing become attractive? a case study of taxi-sharing employing a simulation modelling approach. *IET Intelligent Transport Systems*, 9(2):210–220.
- Egbue, O. and Long, S. (2012). Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy Policy*, 48:717 – 729. Special Section: Frontiers of Sustainability.
- Ekström, J., Kristoffersson, I., and Quttineh, N.-H. (2016). Surrogate-based optimization of cordon toll levels in congested traffic networks. *Journal of Advanced Transportation*, 50(6):1008–1033.
- EPA (2017). Greenhouse Gases Equivalencies Calculator - Calculations and References. <https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references>. Accessed : 4/16/2017.

- EPA (2018). Fast Facts on Transportation Greenhouse Gas Emissions. <https://fueleconomy.gov/feg/Find.do?action=sbs&id=39342&id=39955&id=39414&id=38551>. Accessed : 10/11/2018.
- Fagnant, D. J. and Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 77:167 – 181.
- Fagnant, D. J. and Kockelman, K. M. (2014). The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, 40:1–13.
- Fagnant, D. J. and Kockelman, K. M. (2016). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. *Transportation*, pages 1–16.
- Fagnant, D. J., Kockelman, K. M., and Bansal, P. (2015). Operations of shared autonomous vehicle fleet for austin, texas, market. *Transportation Research Record: Journal of the Transportation Research Board*, (2536):98–106.
- Farhan, J. and Chen, T. D. (2018). Impact of ridesharing on operational efficiency of shared autonomous electric vehicle fleet. *Transportation Research Part C: Emerging Technologies*, 93:310–321.
- Farrington, R. and Rugh, J. (2000). Impact of vehicle air-conditioning on fuel economy, tailpipe emissions, and electric vehicle range: Preprint. Technical report, National Renewable Energy Lab., Golden, CO (US).
- FHWA (2011). Summary of Travel Trends: 2009 National Household Travel Survey.

- FHWA (2016). Electric Vehicle Feasibility: Can EVs take US Households to where they need to go?
- Filzmoser, P. (2018). *Applied Compositional Data Analysis With Worked Examples in R*. Springer Series in Statistics. Springer International Publishing : Imprint: Springer.
- Fisher, R. A. (1935). The design of experiments.
- Fortune (2016). Uber Now Has 40 Million Monthly Riders Worldwide — Fortune.com. <http://fortune.com/2016/10/20/uber-app-riders/>. Accessed : 8/29/2017.
- Franke, T. and Krems, J. F. (2013). What drives range preferences in electric vehicle users? *Transport Policy*, 30:56–62.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):1–67.
- Galland, S., Knapen, L., Yasar, A.-U.-H., Gaud, N., Janssens, D., Lamotte, O., Koukam, A., and Wets, G. (2014). Multi-agent simulation of individual mobility behavior in carpooling. *Transportation Research Part C: Emerging Technologies*, 45:83 – 98.
- Gawron, J. H., Keoleian, G. A., De Kleine, R. D., Wallington, T. J., and Kim, H. C. (2018). Life cycle assessment of connected and automated vehicles: Sensing and computing subsystem and vehicle level effects. *Environmental Science & Technology*, 52(5):3249–3256. PMID: 29446302.

- Gheorghiu, A. and Delhomme, P. (2018). For which types of trips do french drivers carpool? motivations underlying carpooling for different types of trips. *Transportation Research Part A: Policy and Practice*, 113:460–475.
- Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *nature*, 453(7196):779.
- Greene, D. L. and Plotkin, S. (2011). Reducing greenhouse gas emission from us transportation. *Arlington: Pew Center on Global Climate Change*.
- Hawkins, T. R., Singh, B., Majeau-Bettez, G., and Stromman, A. H. (2013). Comparative environmental life cycle assessment of conventional and electric vehicles. *Journal of Industrial Ecology*, 17(1):53–64.
- He, F., Yin, Y., and Zhou, J. (2015). Deploying public charging stations for electric vehicles on urban road networks. *Transportation Research Part C: Emerging Technologies*, 60:227–240.
- He, J., Yang, H., Tang, T.-Q., and Huang, H.-J. (2018). An optimal charging station location model with the consideration of electric vehicle’s driving range. *Transportation Research Part C: Emerging Technologies*, 86:641–654.
- Heath, B., Hill, R., and Ciarallo, F. (2009). A survey of agent-based modeling practices (january 1998 to july 2008). *Journal of Artificial Societies and Social Simulation*, 12(4):9.
- Hong, Z., Chen, Y., Mahmassani, H. S., and Xu, S. (2017). Commuter ride-sharing using topology-based vehicle trajectory clustering: Methodology, application and impact evaluation. *Transportation Research Part C: Emerging Technologies*, 85:573 – 590.

- Hosni, H., Naoum-Sawaya, J., and Artail, H. (2014). The shared-taxi problem: Formulation and solution methods. *Transportation Research Part B: Methodological*, 70:303–318.
- Huang, H. and Zabinsky, Z. B. (2013). Adaptive probabilistic branch and bound with confidence intervals for level set approximation. In *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*, pages 980–991. IEEE Press.
- Jianqin, Z., Peiyuan, Q., Yingchao, D., Mingyi, D., and Feng, L. (2015). A space-time visualization analysis method for taxi operation in beijing. *Journal of Visual Languages & Computing*, 31:1–8.
- Joines, J., Gupta, D., Gokce, M., King, R., and Kay, M. (2002). Supply chain multi-objective simulation optimization. volume 2, pages 1306–1314, USA. IEEE.
- Jung, J., Chow, J. Y., Jayakrishnan, R., and Park, J. Y. (2014). Stochastic dynamic itinerary interception refueling location problem with queue delay for electric taxi charging stations. *Transportation Research Part C: Emerging Technologies*, 40:123–142.
- Kamel, J., Vosooghi, R., Puchinger, J., Ksontini, F., and Sirin, G. (2019). Exploring the impact of user preferences on shared autonomous vehicle modal split: A multi-agent simulation approach. *Transportation Research Procedia*, 37:115 – 122. 21st EURO Working Group on Transportation Meeting, EWGT 2018, 17th – 19th September 2018, Braunschweig, Germany.

- Kim, H., Yang, I., and Choi, K. (2011). An agent-based simulation model for analyzing the impact of asymmetric passenger demand on taxi service. *KSCE Journal of Civil Engineering*, 15(1):187–195.
- Kleiner, A., Nebel, B., and Ziparo, V. A. (2011). A mechanism for dynamic ride sharing based on parallel auctions. In *IJCAI*, volume 11, pages 266–272.
- Ko, J., Gim, T.-H. T., and Guensler, R. (2017). Locating refuelling stations for alternative fuel vehicles: a review on models and applications. *Transport Reviews*, 37(5):551–570.
- Krueger, R., Rashidi, T. H., and Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transportation research part C: emerging technologies*, 69:343–355.
- Lee, C. and Han, J. (2017). Benders-and-price approach for electric vehicle charging station location problem under probabilistic travel range. *Transportation Research Part B: Methodological*, 106:130–152.
- Levofsky, A. and Greenberg, A. (2001). Organized dynamic ride sharing: The potential environmental benefits and the opportunity for advancing the concept. In *Transportation Research Board 2001 Annual Meeting*, pages 7–11.
- Li, B., Krushinsky, D., Woensel, T. V., and Reijers, H. A. (2016a). The share-a-ride problem with stochastic travel times and stochastic delivery locations. *Transportation Research Part C: Emerging Technologies*, 67:95 – 108.
- Li, S., Huang, Y., and Mason, S. J. (2016b). A multi-period optimization model for the deployment of public electric vehicle charging stations on network. *Transportation Research Part C: Emerging Technologies*, 65:128–143.

- Lin, Y., Li, W., Qiu, F., and Xu, H. (2012). Research on optimization of vehicle routing problem for ride-sharing taxi. *Procedia-Social and Behavioral Sciences*, 43:494–502.
- Liu, J. (2012). Electric vehicle charging infrastructure assignment and power grid impacts assessment in beijing. *Energy Policy*, 51:544–557.
- Loeb, B., Kockelman, K. M., and Liu, J. (2018). Shared autonomous electric vehicle (saev) operations across the austin, texas network with charging infrastructure decisions. *Transportation Research Part C: Emerging Technologies*, 89:222–233.
- Luo, H., Kou, Z., Zhao, F., and Cai, H. (2019). Comparative life cycle assessment of station-based and dock-less bike sharing systems. *Resources, Conservation and Recycling*, 146:180 – 189.
- Ma, S., Zheng, Y., and Wolfson, O. (2015). Real-time city-scale taxi ridesharing. *IEEE Transactions on Knowledge and Data Engineering*, 27(7):1782–1795.
- Maciejewski, M. and Bischoff, J. (2015). Large-scale microscopic simulation of taxi services. *Procedia Computer Science*, 52:358–364.
- Mahmoudi, M. and Zhou, X. (2016). Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: A dynamic programming approach based on state-space-time network representations. *Transportation Research Part B: Methodological*, 89:19–42.
- Martinez, L. M., Correia, G. H., and Viegas, J. M. (2015). An agent-based simulation model to assess the impacts of introducing a shared-taxi system: an application to lisbon (portugal). *Journal of Advanced Transportation*, 49(3):475–495.

- Masoud, N. and Jayakrishnan, R. (2017). A decomposition algorithm to solve the multi-hop peer-to-peer ride-matching problem. *Transportation Research Part B: Methodological*, 99:1–29.
- McKinsey & Company (2018). Self-driving car technology: When will the robots hit the road? — McKinsey & Company. <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/self-driving-car-technology-when-will-the-robots-hit-the-road>. Accessed : 4/14/2018.
- Mishra, P. P. (2018). Genetic algorithm approach for inventory and supply chain management: A review. In *Handbook of Research on Promoting Business Process Improvement Through Inventory Control Techniques*, pages 468–478. IGI Global.
- Morrissey, P., Weldon, P., and O’Mahony, M. (2016). Future standard and fast charging infrastructure planning: An analysis of electric vehicle charging behaviour. *Energy Policy*, 89:257–270.
- Mourad, A., Puchinger, J., and Chu, C. (2019). A survey of models and algorithms for optimizing shared mobility. *Transportation Research Part B: Methodological*.
- Nazari, F., Noruzoliaee, M., and Mohammadian, A. K. (2018). Shared versus private mobility: Modeling public interest in autonomous vehicles accounting for latent attitudes. *Transportation Research Part C: Emerging Technologies*, 97:456 – 477.
- Neoh, J. G., Chipulu, M., Marshall, A., and Tewkesbury, A. (2018). How commuters’ motivations to drive relate to propensity to carpool: Evidence from the united kingdom and the united states. *Transportation Research Part A: Policy and Practice*, 110:128–148.

- Neubauer, J. and Wood, E. (2014). The impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. *Journal of power sources*, 257:12–20.
- New York Independent System Operator (2019). Load Data - NYISO. <https://www.nyiso.com/load-data>. Accessed : 2/20/2019.
- Nielsen, J. R., Hovmøller, H., Blyth, P.-L., and Sovacool, B. K. (2015). Of “white crows” and “cash savers:” a qualitative study of travel behavior and perceptions of ridesharing in denmark. *Transportation Research Part A: Policy and Practice*, 78:113–123.
- Nourinejad, M. and Roorda, M. J. (2016). Agent based model for dynamic ridesharing. *Transportation Research Part C: Emerging Technologies*, 64:117–132.
- NYC DOT (2014). 2014 Taxicab Factbook. Technical report.
- NYC DOT (2016a). 2016 Taxicab Factbook. Technical report.
- NYC DOT (2016b). NYC Taxi and Limousine Commission - Trip Record Data. http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml. Accessed : 4/14/2017.
- NYC Open Data (2016). Parking Lot — NYC Open Data. <https://data.cityofnewyork.us/City-Government/Parking-Lot/h7zy-iq3d>. Accessed : 1/9/2019.
- NYCTLIC (2013). Take Charge: A roadmap to Electric New York City Taxis. Technical report, New York City Taxi and Limousine Commission.

- Osorio, C. and Bierlaire, M. (2013). A simulation-based optimization framework for urban transportation problems. *Operations Research*, 61(6):1333–1345.
- Qian, X., Zhang, W., Ukkusuri, S. V., and Yang, C. (2017). Optimal assignment and incentive design in the taxi group ride problem. *Transportation Research Part B: Methodological*.
- Rao, R., Cai, H., and Xu, M. (2018). Modeling electric taxis' charging behavior using real-world data. *International Journal of Sustainable Transportation*, 12(6):452–460.
- Richardson, D. B. (2013). Electric vehicles and the electric grid: A review of modeling approaches, impacts, and renewable energy integration. *Renewable and Sustainable Energy Reviews*, 19:247–254.
- Romero, J. P., Ibeas, A., Moura, J. L., Benavente, J., and Alonso, B. (2012). A simulation-optimization approach to design efficient systems of bike-sharing. *Procedia-Social and Behavioral Sciences*, 54:646–655.
- Sadeghi-Barzani, P., Rajabi-Ghahnavieh, A., and Kazemi-Karegar, H. (2014). Optimal fast charging station placing and sizing. *Applied Energy*, 125:289–299.
- Salanova, J. M., Romeu, M. E., and Amat, C. (2014). Aggregated modeling of urban taxi services. *Procedia-Social and Behavioral Sciences*, 160:352–361.
- Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S. H., and Ratti, C. (2014). Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37):13290–13294.

- Santos, D. O. and Xavier, E. C. (2015). Taxi and ride sharing: A dynamic dial-a-ride problem with money as an incentive. *Expert Systems with Applications*, 42(19):6728–6737.
- Shaheen, S. A., Chan, N. D., and Gaynor, T. (2016). Casual carpooling in the san francisco bay area: Understanding user characteristics, behaviors, and motivations. *Transport Policy*, 51:165–173.
- Shahraki, N., Cai, H., Turkay, M., and Xu, M. (2015). Optimal locations of electric public charging stations using real world vehicle travel patterns. *Transportation Research Part D: Transport and Environment*, 41:165–176.
- Simonetto, A., Monteil, J., and Gambella, C. (2019). Real-time city-scale ridesharing via linear assignment problems. *Transportation Research Part C: Emerging Technologies*, 101:208–232.
- Sivanandam, S. N. (2007). *Introduction to genetic algorithms*. Springer, Berlin ; New York.
- Smith, M. and Castellano, J. (2015). Costs associated with non-residential electric vehicle supply equipment: Factors to consider in the implementation of electric vehicle charging stations. Technical report.
- Taiebat, M., Brown, A. L., Safford, H. R., Qu, S., and Xu, M. (2018). A review on energy, environmental, and sustainability implications of connected and automated vehicles. *Environmental science & technology*, 52(20):11449–11465.
- Tesla (2019). Supercharger — Tesla. <https://www.tesla.com/supercharger?redirect=no>. Accessed : 1/29/2019.

Tu, W., Li, Q., Fang, Z., Shaw, S.-l., Zhou, B., and Chang, X. (2016). Optimizing the locations of electric taxi charging stations: A spatial-temporal demand coverage approach. *Transportation Research Part C: Emerging Technologies*, 65:172–189.

US DOE (2019). Alternative Fuels Data Center: Data Downloads. afdc.energy.gov/data_download. Accessed : 5/22/2019.

U.S. DOE (2019). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990 - 2017. Technical report, U.S. Department of Energy.

US EPA (2019a). Beyond Tailpipe Emissions: Results. <https://www.fueleconomy.gov/feg/Find.do?zipCode=10001&year=2019&vehicleId=41188&action=bt3>. Accessed : 2/20/2019.

US EPA (2019b). Greenhouse Gas Emissions from a Typical Passenger Vehicle — Green Vehicle Guide — US EPA. <https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle>. Accessed : 2/20/2019.

Wadud, Z., MacKenzie, D., and Leiby, P. (2016). Help or hindrance? the travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practice*, 86:1 – 18.

Wang, H., Mostafizi, A., Cramer, L. A., Cox, D., and Park, H. (2016). An agent-based model of a multimodal near-field tsunami evacuation: Decision-making and life safety. *Transportation Research Part C: Emerging Technologies*, 64:86 – 100.

Wang, Y., Zheng, B., and Lim, E.-P. (2018). Understanding the effects of taxi ride-sharing—a case study of singapore. *Computers, Environment and Urban Systems*, 69:124–132.

- Waymo (2019). Journey - Waymo. <https://waymo.com/journey/>. Accessed : 5/22/2019.
- Wu, F. and Sioshansi, R. (2017). A stochastic flow-capturing model to optimize the location of fast-charging stations with uncertain electric vehicle flows. *Transportation Research Part D: Transport and Environment*, 53:354–376.
- Xi, X., Sioshansi, R., and Marano, V. (2013). Simulation–optimization model for location of a public electric vehicle charging infrastructure. *Transportation Research Part D: Transport and Environment*, 22:60–69.
- Xu, R. et al. (2015). *Machine learning for real-time demand forecasting*. PhD thesis, Massachusetts Institute of Technology.
- Yang, J., Dong, J., and Hu, L. (2017a). A data-driven optimization-based approach for siting and sizing of electric taxi charging stations. *Transportation Research Part C: Emerging Technologies*, 77(2):462–477.
- Yang, J., Dong, J., and Hu, L. (2017b). A data-driven optimization-based approach for siting and sizing of electric taxi charging stations. *Transportation Research Part C: Emerging Technologies*, 77:462–477.
- Yilmaz, M. and Krein, P. T. (2012). Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles. *IEEE transactions on Power Electronics*, 28(5):2151–2169.
- Yu, B., Ma, Y., Xue, M., Tang, B., Wang, B., Yan, J., and Wei, Y.-M. (2017). Environmental benefits from ridesharing: A case of beijing. *Applied energy*, 191:141–152.

- Zhang, Y. and Zhang, Y. (2018). Examining the relationship between household vehicle ownership and ridesharing behaviors in the united states. *Sustainability*, 10(8):2720.
- Zhu, S. and Kornhauser, A. (2017). The interplay between fleet size, level-of-service and empty vehicle repositioning strategies in large-scale, shared-ride autonomous taxi mobility-on-demand scenarios. In *Transportation Research Board 96th Annual Meeting*.
- Zhu, Z.-H., Gao, Z.-Y., Zheng, J.-F., and Du, H.-M. (2016). Charging station location problem of plug-in electric vehicles. *Journal of Transport Geography*, 52:11–22.
- Zou, M., Li, M., Lin, X., Xiong, C., Mao, C., Wan, C., Zhang, K., and Yu, J. (2016). An agent-based choice model for travel mode and departure time and its case study in beijing. *Transportation Research Part C: Emerging Technologies*, 64:133 – 147.

9. CURRICULUM VITA

9.1 Education

9.1.1 Academic Qualifications

- Doctor of Philosophy, Industrial Engineering, Purdue University, West Lafayette, IN (2015–2019)
- Bachelor of Technology, Mechanical Engineering College of Engineering, Pune, India (2009–2013)

9.1.2 Publications

- **Lokhandwala, M., & Nateghi, R.** (2018). Leveraging advanced predictive analytics to assess commercial cooling load in the US. *Sustainable Production and Consumption*, 14, 66-81.
- **Lokhandwala, M., & Cai, H.** (2018). Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of NYC. *Transportation Research Part C: Emerging Technologies*, 97, 45-60.
- **Lokhandwala, M. & Cai H.** (2019). Using Comparable Scenarios to Analyze Taxi Fleets with Variable Ride Sharing and Autonomous Vehicle Adoption. *Transportation Research A: Policy and Practice (Under Preparation)*

- **Lokhandwala, M.** & Cai H. (2019). Siting Electric Vehicle Charging Stations for a Growing Fleet. Transportation Research D: Transport and Environment (Submitted)
- **Lokhandwala, M.** & Cai H. (2019). Understanding the Sharing Limit in an SAV System. Transportation Research Record (Under Preparation)
- **Lokhandwala, M.** & Cai H. The Influence of Rider Behavior in Shared Mobility. Transportation Research Part C: Emerging Technologies (Under Preparation)

9.1.3 Presentations

- **Lokhandwala, M.** and Cai H. (2016). The environmental benefits of taxi sharing in New York City. Poster Presentation *INFORMS AGM Nashville, TN*
- **Lokhandwala, M.**, Nateghi R. and Shevade P (2016). Analysis of Energy Consumed due to Cooling. Poster Presentation *SRA Meeting San Diego, CA*
- **Lokhandwala, M.** and Cai H. (2017). Ride Sharing: An Agent Based Modeling Analysis. Poster Presentation *ISIE ISSST Annual Meeting, Chicago IL 2017*
- **Lokhandwala, M.** and Cai H. (2017). Ride Sharing: An Agent Based Modeling Analysis. Oral Presentation *INFORMS AGM Houston, TX 2017*
- **Lokhandwala, M.** and Cai H. (2017). Ride Sharing: An Agent Based Modeling Analysis. Poster Presentation *Dawn or Doom, West Lafayette IN 2017*

- **Lokhandwala, M.** and Cai H. (2018). The future of sustainable Transportation. Poster Presentation *GRC-Industrial Ecology, Les Diabrelets, Switzerland*
- **Lokhandwala, M.** and Cai H. (2018). Siting Charging Stations for Electric Vehicles using Simulation Optimization *INFORMS AGM Phoenix, AZ*
- **Lokhandwala, M.** and Cai H. (2019). Agent Based Models for Shared Autonomous Electric Vehicle Systems *Transportation Research Board Washington D.C*

9.1.4 Teaching Assistant Experience

- IE 530: Quality Control
- IE 431: Industrial Engineering Design
- IE 343: Engineering Economy
- IE 230: Probability and Statistics 1
- IE 335: Operations Research: Optimization

9.1.5 Mentorship

Mentored three undergraduate students and three masters students

- Anthony Neimic - Freshman Computer Science Purdue University
- Jackson Bennett - Senior Industrial Engineering Purdue University
- Nikhil Carnerio - Sophomore Industrial Engineering Purdue University

- Brijesh Poojary - Masters Industrial Engineering Purdue University
- Rahulkumar Baranwal - Masters Industrial Engineering Purdue University
- Varun Srivastava - Masters Industrial Engineering Purdue University
- Nicholas Mori - Masters Industrial Engineering Purdue University

9.1.6 Projects

- CFD Analysis of a rocket, Armament Research and Development Institute, Pune, India (January 2013–May 2013)
- BAJA SAE Design Series India (2012)
- BAJA SAE Design Series India (2013)
- SASOL BAJA South Africa (2012)

9.2 Work Experience

- iPhone Product Operations Intern, Apple Inc, Cupertino, CA (May 2018–Aug 2018)
- Experience Insights and Analytics Intern, Walt Disney World, Orlando, FL (May 2017–August 2017)
- Assistant Manager - Purchasing, Tata Motors, Pune, India (October 2013–July 2015)

9.3 Technical skills

- **Programming Languages:** R, Python, Java, SAS, VBA, SQL, Matlab
- **Software Skills:** Anylogic, ARENA, Minitab, Excel (advanced), Tableau, CPLEX

9.4 Achievements

9.4.1 Grants

- Bisland Dissertation Award 2018 - Full Tuition Waiver with Stipend
- Andrews Environmental Travel Grant 2017 - \$1500
- Ismael Interdisciplinary Travel Grant 2017 - \$700
- Purdue Climate Change Research Center Travel grant 2016 - \$1000
- College of Engineering Grant for the IEGSO - \$2200
- Travel Scholarship by ISIE ISSST to attend ISIE ISSST Conference in Chicago- \$1145
- Purdue Graduate Student Government Travel Grant 2018 - \$250
- Purdue Graduate Student Government Travel Grant 2017 - \$250
- Purdue Climate Change Research Center Travel grant 2018 - \$1000

9.4.2 Awards

- Best Paper Award from Chinese Society of Industrial Ecology - 2018
- 4th Place in ISIE ISSST Poster Competition - \$100
- 1st Overall, BAJA India 2013
- 1st Acceleration, BAJA India 2013
- 1st Build Quality, BAJA India 2013
- 1st CAE, BAJA India 2013
- 1st Endurance, BAJA India 2013
- 1st Fastest Lap, BAJA India 2013
- 1st Acceleration, BAJA India 2012

9.5 Leadership

- Academic Affairs Chair - Industrial Engineering Graduate Student Organization (2016 - 2017)
- President - Industrial Engineering Graduate Student Organization (2017 - 2018)
- Communications Chair - Industrial Engineering Graduate Student Organization (2018 - 2019)

9.6 Affiliations

- Institute for Operations Research and Management Sciences (INFORMS)
- International Society for Industrial Ecology (ISIE)
- Transportation Research Board (TRB)
- Purdue Industrial Engineering Graduate Student Organization (IEGSO)

APPENDICES

APPENDIX A

PSEUDO CODES FOR PP-SAEV MODEL

I gave a simplified description of the model in Chapter 2. Below I expand on the details of the parameters and the actions of the individual agents. The pseudo code here gives additional details to complement Algorithm 1-7 of the main text. Several steps of Algorithm 1-7 have been broken down into several sub steps to demonstrate the flow of the agent-based model.

A.1 Main Program and Input Data

The execution of the simulation begins with the first step of the main algorithm. The algorithm's main task is to call two sub algorithms for the *taxi* agents and the *riderGroup* agents, which it does by looping through the collection of agents defined in Section A.1.1. My algorithm gives control of shifts starts for taxis to the main section.

I restrict our area of focus to the city of New York which I further divide into two distinct regions Manhattan and Brooklyn. I have also defined a sub region of Times Square within Manhattan to account for the significantly low speeds of traffic in that region. The Regions are defined as per Table A.1. These regions are primarily used to limit the search area for shares and for rides and also to define average speeds of taxis.

Table A.1.
Boundary points defining the regions of NYC

Manhattan			Brooklyn			Times Square		
	Latitude	Longitude		Latitude	Longitude		Latitude	Longitude
1	40.71196	-73.9795	1	40.66567	-73.7416	1	40.77035	-73.987513
2	40.73795	-73.9774	2	40.68595	-73.7267	2	40.76285	-73.969753
3	40.77507	-73.9438	3	40.74382	-73.7331	3	40.74004	-73.986493
4	40.78303	-73.9446	4	40.76253	-73.757	4	40.74762	-74.004151
5	40.79725	-73.9324	5	40.75851	-73.8515	5	40.77035	-73.9875
6	40.83492	-73.9359	6	40.77536	-73.924			
7	40.85633	-73.924	7	40.70011	-73.9619			
8	40.86915	-73.9297	8	40.70174	-73.9912			
9	40.829	-73.9508	9	40.66714	-73.9954			
10	40.75087	-74.0075	10	40.61386	-74.0371			
11	40.70291	-74.0132	11	40.58898	-73.9835			
12	40.71196	-73.9795	12	40.58975	-73.9217			
			13	40.62609	-73.9028			
			14	40.66892	-73.8424			
			15	40.66567	-73.7416			

A.1.1 Global Parameters

- *taxiDist* : Bi-Variate distribution to define the initial locations of the taxis and to select hot spots for the taxis to return to while searching
- *EVPercent* : The percent of the taxis that are electric vehicles

- *AVPercent* : The percent of the taxis that are autonomous vehicles
- *fleetSize* : The total number of taxis
- *taxiCapacityDistribution* : Distribution of the taxi capacity excluding the driver
- *busyPercent* : Percent of busy taxis at the start of the simulation
- *riderGroups* : Collection of all riders in the system
- *Taxis* : Collection of all taxis in the system
- *taxis_free_man*[*i*][*j*] :¹ Two dimensional grid collection of all the taxis that do not have any rider (idle) within the Manhattan area as defined by Table A.1. Resolution of the grid is 200×200
- *taxis_free_brook*[*i*][*j*] :¹ Two dimensional grid collection of all the taxis that do not have any rider (idle) within Brooklyn as defined by Table A.1. Resolution of the grid is 200×200
- *taxis_free_else* : Collection of all the taxis that do not have any rider (idle) within all other areas except Manhattan and Brooklyn as defined by Table A.1.
- *taxis_shr_man*[*i*][*j*] :¹ Two dimensional grid collection of all the taxis in sharing mode within Manhattan as defined by Table A.1. Resolution of the grid is 200×200
- *taxis_shr_brook*[*i*][*j*] :¹ Two dimensional grid collection of all the taxis in sharing mode within Brooklyn as defined by Table A.1. Resolution of the grid is 200×200

¹ *i* represents the index for the latitude and *j* represents the index for the longitude.

- *taxis_shr_else* : Collection of all the taxis in sharing mode in all other areas except Manhattan and Brooklyn as defined by Table A.1.
- N_{hot} : Count of hotspots for the taxis to pick at the start of the simulation.
- t : Main time counter
- $tEnd$: Simulation length

A.1.2 Main Algorithm

1. Load GIS region of NYC
2. Draw regions bounding Manhattan and Brooklyn
3. Initialize grid and arrays *taxis_free_man*, *taxis_free_brook*, *taxis_free_else*, *taxis_shr_man*, *taxis_shr_brook* and *taxis_shr_else*
4. Initialize *fleetSize* taxis in the environment as per *taxiDist* and set *Capacity* from *taxiCapacityDistribution*, *TaxiID*, *EVFlag* as per *EVPercent*, *AVFlag* as per *AVPercent*, and *State* (=0)
5. For each taxi do (the taxi parameters and variables are defined in Section A.3).
6. Add all taxis to the collection *Taxis*.
7. Initialize *busyPercent* taxis as busy (*State* = 0.1) and moving to a random location.
8. Select N_{hot} locations from *taxiDist* as hotspots and assign N_{hot_i} hotspots to each taxi i

9. Set $t = 0$
10. For each rider group in the database entering at time t , do
(# rider variables are defined in Section A.2.2)
11. Assign *riderID* sequentially
12. Set *State* = 0
13. set *r_indi*, *r_indj*, and *r_indk* as per location
14. set *canShare* and *deviationTolerance* as per distribution
15. For each rider group g in the system, execute function *riderAlgo*(Group g) (**# details of this function are in Section A.2.2**)
16. For each taxi x_c in the system Execute function *taxiAlgo*(Taxi x_c) textbf(# details of this function are in Section A.3)
17. If $t = t, End$ and Go to 19
18. Set $t = t + 1$ and Go to 10
19. Stop

A.2 Rider Algorithm and Data

The Rider Group Algorithm defines the actions the riders go through when they enter the system. The riders after initialization go through 3 main states (searching, waiting and riding) before exiting the system. The riders first search for rides using algorithm *searchForTaxis*. If they fail to find a ride within *waitLimit* the rider

group exits the system unserved. If the rider group does find a ride they waits for the taxi to pick them up at the pick up location ($pLong$, $pLat$), then ride with the taxi and exit the system when the taxi drops them off at the drop off location ($dLong$, $dLat$)

A.2.1 Parameters and Variables

The rider agents have the following individual parameters and variables.

- *riderID* : A unique identification number for each rider. These are assigned to the rider as it enters into the system sequentially.
- *riderType* : The type of rider that defines its searching strategy
- *State* : A state defining the following actions of the rider group. The possible states are 0. Initialization, 1. Searching taxi, 1.9 Waiting for response 2. Waiting for pick up 3. Riding
- *pDt* : The date and time of the riders entry in the system, taken from the data base.
- *pLong* : The pickup Longitude
- *pLat* : The pickup Latitude
- *dLong* : The drop off Longitude
- *dLat* : The drop off Latitude
- *passengers* : number of passengers in the group
- *waitLimit* : Maximum time a rider will wait before exiting the system

- *taxarr* : List of completely empty taxis under consideration
- *sharr* : List of taxis in sharing mode under consideration
- *t_l* : local time at that state in which it was used
- *Stats* : A record of the time the rider has changed their state
- *shrLimit* : The time for which a Type 4 rider would search for shared rides
- *myTaxi* : Taxi which has been assigned to the group
- *myTaxi_{reg}* : Empty Taxi which has been assigned to the group temporarily
- *myTaxi_{shr}* : Shared Taxi which has been assigned to the group temporarily
- *noTaxis* : Set of taxis that have refused service to the rider earlier
- *regLimit* : The time for which a Type 2 rider would search for non shared rides
- *r_indk* : Index defining region (0=Manhattan; 1=Brooklyn; 2 = Else)
- *r_indi* : Row index for rider group's location within the grid defined in Section A.1.1
- *r_indj* : Column index for rider group's location within the grid defined in Section A.1.1
- *tis* : Total waiting time of the rider group
- *tries* : Number of failed tries while searching for a rider group
- *deviationTolerance* : percent by which rider will tolerate deviation from path
- *stats* : a record of each time the rider has changed its *State*

A.2.2 Rider Algorithm - *riderAlgo(Group g)*

1. If $State = 0$ # Initialization
2. Set $State = 1$
3. Set local time $t_l = 0$
4. If $State = 1$ # Searching for regular ride
5. Execute Searching Algorithm *searchForTaxis*
6. If $State = 1.9$ # waiting for response (taxi will check shares using *preCheck* and *bestRoute*)
7. If message “Coming” received # Match is found
8. Set $myTaxi = sender$
9. Set $State = 2$
10. Go to 14
11. Else If $t_l = 15$
12. Set $State = 1$ and Goto 5
13. Else $t_l = t_l + 1$; $tis = tis + 1$ and Return to **Main** Step 15
14. If $State = 2$ # Waiting for Pickup
15. If $(myTaxi.Latitude, myTaxi.Longitude) = (PULat, PULong)$
16. Set $State = 3$

17. Return to **Main** Step 15
18. If State = 3 # Riding to destination
19. Set (*myTaxi.Latitude*, *myTaxi.Longitude*) → (*Latitude*, *Longitude*)
20. If (*Latitude*, *Longitude*) = (*DOLat*, *DOLong*)
21. Record Stats and exit the system
22. End
23. Return to Main
24. Return to **Main** Step 15

A.2.3 Searching Strategies - *searchForTaxis(Group g)*

1. Exit System if $t_l \geq waitLimit$ # Not found a taxi for time more than *waitLimit*
2. If $riderType = 1$ or $riderType = 3$ or ($riderType = 2$ and $tis < regLimit$) or ($riderType = 4$ and $tis \geq shareLimit$)
3. Send message “ride” to free taxis within *tries* index blocks of *r_indi* *r_indk*
r_indk # Search for regular taxis within *tries* index blocks of current location
4. *tries* = *tries* + 1 # Expand search area
5. If $riderType = 5$ or $riderType = 3$ or ($riderType = 2$ and $tis \geq regLimit$) or ($riderType = 4$ and $tis < shareLimit$)

6. Send message “Share” to Sharing taxis within 2 index blocks of r_indk
 r_indk r_indk except $noTaxis$ # Search for sharing taxis
7. Set $State = 1.9$ # Wait till response is received from taxis
8. Set $t_l = 0$
9. Go to *riderAlgo* line 6

A.3 Taxi Algorithm and Data

The Taxi Algorithm defined below sets rules for taxis to follow while searching for new rides (while idle) or while serving current rides and searching for potential shares.

The non-sharing requests from rider groups are sent to unoccupied taxis. The taxis give preferences to the rider groups that are closer to their current locations. Because each rider group may send the ride requests to more than one taxi, it is possible that a match has already been formed between another taxi and the preferred rider. So the taxi will check with the rider groups to only commit pick-up to the closest rider group that is still searching. A match is then formed between the taxi and the rider group. The rider groups will send out the ride requests first to taxis that are located in the same grid cell with them (I divided the map into grids). If no available taxis are found in the same grid cell with the rider group, then the search will expand to nearby grid cells.

The sharing requests are sent out by the rider groups that are willing to share rides (broadcasting a share request). The sharing requests are only sent to taxis that are currently in the process of serving other rider groups that are willing to share and can accept shares (taxis will stop accepting shares when the remaining distance

toleration of one or more rider groups currently in the car is used up and cannot accommodate more sharing). The taxis evaluate sharing requests following in a two-step process. First, to reduce computational intensity, a pre-screening is conducted to eliminate sharing requests that are unlikely to be feasible. I have a heuristic algorithm (*preCheck*) described in Section A.3.2 `preCheck()` to serve this purpose. For requests that pass the screening, the optimal way to insert the pickup and drop-off locations of the requested rider group to the current trip chain is then identified to form the candidate new sharing route. The deviation required from the candidate new sharing route is calculated and compared with the remaining tolerance deviation of rider groups this taxi is currently serving and that of the requested rider group. Only if the required deviation of the candidate new sharing route is acceptable to all rider groups, the sharing is considered feasible. All requests received by the taxi will be evaluated following this process and the matching will be formed for the best sharing chosen from all feasible requests.

A.3.1 Taxi Algorithm

Below are the parameters and variables defining the agent type *taxi*

- *taxiID* : A unique identification number for each taxi in the system
- *newLocation* : The next location that a taxi goes to while searching
- *capacity* : Capacity of the taxi when empty
- *curCapacity* : Number of empty seats in taxi
- *State* : State in which the taxi is in, defining current and future actions. The possible states that a taxi can take are 0. Initialization, 0.1. Busy Start, 1.

Searching, 1.5 Riding Empty 1.9 Evaluate candidate riders, 2. Occupied and Searching for shares 2.5 Searching for shares

- *newLocation* : The next location a taxi goes to while searching
- N_{hot_i} Number of hotspots for the taxi
- *rideList* : A collection of riders currently in the taxi
- *hotspots* Set of Hotspots for the taxi
- *myRoute* : A sequence of points the taxi must visit in order
- *newRoute* : A route proposed by the ride matching algorithm to enable taxi sharing
- *curRiders* : The riders to whom this taxi is currently assigned
- *flag* : a boolean variable to store the result of the precheck algorithm defined in Section A.3.2
- *index* : An array storing the coordinates of the bounding box defined in *prechk*
- *tempRoute* : A temporary variable used to permute through the points in *myRoute* to evaluate new shares.
- *extrad* : The extra distance that would be deviated on account of sharing
- *allow* : The maximum allowable deviation on account of sharing
- *df* : A factor between 0 to 1 that is used to compare the shares amongst each other. It is the product of the ratio of the total deviation with the maximum allowable deviation for each rider in *curRiders* and also the rider for which we are evaluating the share. Smaller is better.

- *shragent* : A collection of rider groups under evaluation for sharing (up to 3 riderGroups) not including those in the taxi
- t_l : local time in state
- t_indk : Index defining region (0=Manhattan; 1=Brooklyn; 2 = Else)
- t_indi : Row index for taxi's location within the grid defined in Section A.1.1
- t_indj : Column index for taxi's location within the grid defined in Section A.1.1

A.3.1.1 *taxiAlgo(Taxit)*

1. If $State = 0$ # Initialization
2. Clear *curRiders* and *myRoute*; *myRider* = *NULL*; *curCapacity* = *Capacity*
3. Update grid position in *free_taxi*
4. Set $State = 1$; $t_l = 0$ and Goto 7
5. If $State = 0.1$ # Initially Busy taxis set at simulation start
6. If taxi has reached random location set state = 0 Go to 1 # Taxi is now available for service
7. If $State = 1$ # Searching for the first rider group
8. If message "ride" is received from rider # non shared message received from rider
9. Set $State = 1.9$ Goto 23

10. Set $t_l = t_l + 1$
11. If $t_l \geq idleTime$ # Taxis move from one hotspot to another while searching
12. Select $newLocation \in hotspots$ and begin moving to $newLocation$
13. Set $t_l = 0$; $State = 1.5$
14. Else Return to Main Step 16
15. If $State = 1.5$ # Riding Empty
16. If message “ride” is recieved from a rider $\in Groups$
17. Set $State = 1.9$ Goto 23
18. Else If taxi at $newLocation$ Goto 3 # Taxi was moving and reached new location
19. Else If $t_l \geq 30$
20. Update location in grid
21. Set $t_l = 0$
22. Else Set $t_l = t_l + 1$ and Return to Main Step 16
23. If $State = 1.9$ # Evaluate all riders and select the closest one
24. Set $myRider = rider$ which sent the message
25. Add $myRider$ to $curRiders$

26. Add $(myRider.PULat, myRider.PULong)$ and $(myRider.DOLat, myRider.DOLong)$ to $myRoute$
27. De-register from free taxi grid
28. Move to $myRoute[0]$ #move at speed defined by road towards $myRoute[0]$ and update location at each time step
29. Set $State = 2$
30. If $State = 2$ # Riding
31. If $(Latitude, Longitude) = myRoute[0]$ Goto 45 # Taxi has reached a destination point
32. If $canShare = TRUE \ \forall \ curRiders$ # Search for shares only if all rider groups allow sharing. A rider group which has started sharing may refuse new requests if they have deviated too far from the original route
33. Update location in $shareTaxi$ grid
34. Set $t_l = 0$ and $State = 2.5$
35. Goto 37
36. Else Return to Main Step 16
37. If $State = 2.5$ #Searching for shares
38. Set $t_l = t_l + 1$
39. if $t_l \geq 10$ or $shcount = 3$

40. Run $flag \leftarrow precheck(Groupg, Taxithis, myRiders)$ # The *preCheck* algorithm filters impossible shares and returns TRUE if the ride should be evaluated by *bestRoute*
41. If $flag = TRUE$ Run $newRoute \leftarrow bestRoute$
42. If *newRoute* is acceptable set $myRoute \leftarrow newRoute$ and Goto 28
43. Else If $(Latitude, Longitude) = myRoute[0]$ Goto 45 # At a point in *myRoute*
44. Else Return to Main Step 16
45. If $State = 3$
46. If $myRoute[0]$ is a Drop off point
47. remove the rider whose drop of point is $myRoute[0]$
48. Remove $myRoute[0]$
49. If *myRoute* is empty Set $state = 0$ and Return to Main Step 16
50. Set $State = 1.9$ and Goto 28

A.3.2 $precheck(Group\ g, Taxi\ t, Array< Groups > A)$

The evaluation of whether a share is feasible or not is computationally expensive ($O(n^2)$ where n are the number of points involved in the share). Additionally each time a share is evaluated the routes between each of these combinations of points need to be evaluated which is highly computationally intensive. To prevent repeated

evaluations of shares that are not feasible, I introduce the below precheck algorithm to filter out the shares. Figure A.1 offers a pictorial representation of this algorithm.

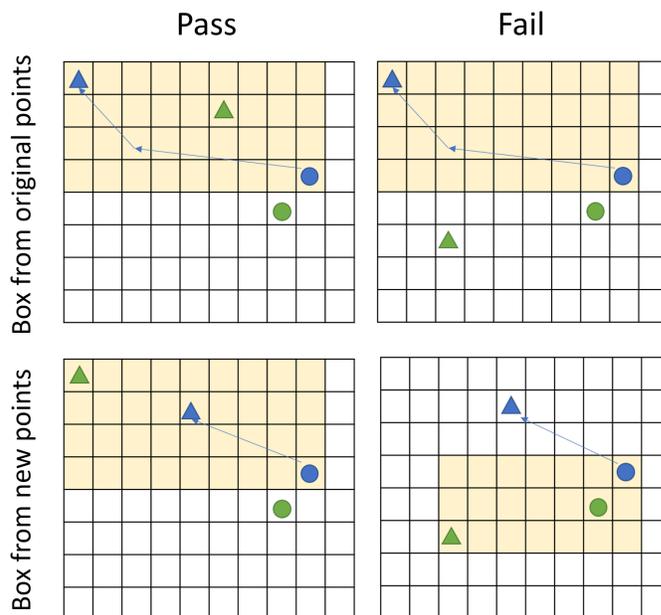


Figure A.1. Graphical illustration of the *PreCheck* algorithm - T : The current location of the taxi, P: The pick up point of the potential share, and D: The drop off point of the potential share. The yellow rectangle represents the largest bounding box formed by the points on the original trip chain and the current location of the taxi. The gray box is the bounding box formed with the drop off location of the potential share and the current location of the taxi. To pass the *preCheck*, the drop off point of the potential share should either lie in the yellow box or form a gray bounding box that covers the yellow box.

I use two bounding boxes to indicate the “direction” of the trips. The *preCheck* algorithm is a two-step process. Step 1: the first biggest bounding box is formed using all the pick up and drop off points that are yet to be visited by the taxi (the

current trip chain) and the current location of the taxi (shown as the yellow colored cells in Figure A.1). The drop off point of the candidate share is then evaluated relative to this bounding box. If the drop off point of the candidate share is within the bounding box, then this candidate share passes the *preCheck* and will be further evaluated by the *bestRoute* algorithm (Figure A.1a). Step 2: If the drop off point of the candidate share is outside of this bounding box, a second bounding box will be formed using the drop off point of the candidate share and the current location of the taxi (shown as the gray shaded boxes in Figure A.1c and d). If the second bounding box fully covers the first bounding box (figure A.1c), then this candidate share also passes the *preCheck* and will be further evaluated by the *bestRoute* algorithm. The second bounding box makes sure that we don't reject candidate trips that travel further than the existing trip chain. Because the pick up location of the candidate will be close to the current location of the taxi (sharing requests are only broadcast to nearby taxis), only the drop off location need to be evaluated.

The two biggest bounding boxes would encompass most of the streets that the taxi would travel on and hence allow those candidate sharing trips, that appear on most possible turns that the taxi would take, to pass the *preCheck*. We did consider the possible cases that the drop off location of a candidate trip falling slightly outside of the bounding boxes. So we tested scenarios with extended bounding boxes (enlarged by one or two grid cells in all directions). No significant differences are observed in the results.

Variables:

- *index*[] : An array to store the bounding box for the taxis
- *g* : The rider group that is being evaluated as a potential share
- *r* : A counter for the for loop in Line 8

- x_c : The taxi that calls the precheck algorithm
- A : A collection of all the rider groups involved in the share (that is all the riders in *riderList* and g)
- r_indi, r_indk, r_indk : The indexes of the rider group g as defined in Section A.2.2
- r_indi, r_indk, r_indk : The indexes of the rider group g as defined in Section A.2.2

A.3.2.1 Algorithm precheck(Group g , Taxi x_c , Array< Groups > A)

1. Set $index[1, 2] \leftarrow (x_c.Latitude, x_c.Longitude)$
2. If \forall Group $g \in A$ have the same r_indk
3. Set $index[3, 4]$ as the biggest bounding box that can be generated from all points $\in myRoute \cup (t.Latitude, t.Longitude)$ with $index[1, 2]$ as base
4. If $((r_indi - index[1]) \times (index[3] - index[1]) \geq -|(index[3] - index[1])|$ And $((r_indj - index[2]) \times (index[4] - index[2])) \geq -|(index[4] - index[2])|$
5. Return true
6. Else Return false
7. Else
8. For each Group $r \in A$
9. Set $d =$ distance from drop off point of g to drop off point of r

10. If $d >$ total distance of myRoute
11. Return false
12. Return true

A.3.3 bestRoute(Group a, myRoute)

The function of this algorithm is to build permutations for the points to be visited by the Taxi in *myRoute* in addition to the pick up and drop off points for the new share agent and also to select the best possible route amongst these permutations. The best possible route is that combination that minimizes the product of the deviations (A.1).

$$score = \prod_{i=1}^N \frac{\text{extra distance traveled by the rider}_i}{\text{extra distance allowed for the rider}_i} \quad (\text{A.1})$$

Where N is the total number of riders involved in the share

Instead of building all possible permutations for all these points as a heuristic we consider the original order of points to be visited in *myRoute* be fixed. We then sequentially try all other positions for the pick up and drop off points for r while keeping in mind that the pick up point must always be before the drop-off point. We thus reduce the time complexity from $O(n!)$ to $O(n^2)$ Variables:

- a : The rider group whose share is being evaluated
- $tempRoute$: A temporary variable to store route permutations
- $myRoute$: The route list of taxi x_c which calls the best route algorithm
- $distance[][]$: A matrix of distances from and to each point in *myRoute*

- d : Stores the extra distance of the current iteration
- df : Variable used to calculate the score
- df_s : Variable used to store the smallest score
- i, j, k : For loop counters
- $extrad$: The extra distance traveled by the rider group due to the share
- $allow$: The total distance allowed by the rider group after sharing
- rt : The new route for the share

Algorithm $\text{bestRoute}(\text{Group } a, \text{myRoute})$:

1. Set $\text{tempRoute} \leftarrow \text{myRoute}$
2. Add pick up and drop-off points of r to tempRoute
3. Set $\text{newRoute} \leftarrow \text{myRoute}$
4. Create Array $\text{distance}[\text{tempRoute.Size}][\text{tempRoute.Size}]$
5. For all $i, j \in 1$ to tempRoute.Size set $\text{distance}[i][j] \leftarrow$ distance between point i and point j **#create the distance matrix**
6. Set $d = 0, extrad = 0, df = 1, allow = 0, df_s = 1$
7. Set $\text{riders} \leftarrow$ all Groups in curRiders and a
8. For i in 0 to $\text{tempRoute.Size} - 1$ **#Outer loop to permute the pick up point of new request**
9. Add pick up point of a to newRoute at position i

10. For j in $i + 1$ to $tempRoute.Size$ **#Inner loop to permute the drop off point of new request**
11. Add drop off point of a to $newRoute$ at position j
12. Set $d =$ distance from taxi to $newRoute[0]$
13. Set $factors \leftarrow deviationTolerance \forall riders$
14. For k in 0 to $newRoute.Size - 1$ **#Loop to check score of share request**
15. set $d = d +$ distance of $newRoute[k]$ to $newRoute[k + 1]$
16. If $newRoute[k+1]$ is a drop off point **Check each riders distance deviation**
17. $extrad =$ Remaining distance for $newRoute[k+1].passenger$
18. $allow = (newRoute[k + 1].deviationTolerance + 1) \times extrad$
19. If $d > allow$, go to 26
20. Else
21. $df = df \times \frac{d}{allow}$
22. $factor[i] = factor[i] \times (1 - \frac{d}{allow})$ **#Calculate the score of the current iteration**
23. If $df_s > df$ **#Find the route with the highest score**
24. set $df_s > df$

25. Set $rt \leftarrow newRoute$
26. remove $newRoute[j]$
27. remove $newRoute[i]$
28. if $df < 1$, Return rt , Else return null **#Send the proposed route back to the taxi**

A.3.4 Shift Start and End

In order to model traditional taxis which are driven by drivers taking shifts, I implemented the below algorithm. I assign the initial status (in-shift or out-shift) for the taxis according to the historical availability ratio based on the model start time (normally midnight) and the day I am modeling (weekday or weekend). According to the availability ratio, a certain number of taxis are randomly selected to be in-service while the rest be out-service. We assume that each shift will last at least 8 hours. So for the taxis started with the in-service status, I randomly assign a shift start time within the previous eight hours according to the distribution of shift start time. For the taxis that are in-shift, at the end of their 8-hour shift, the unoccupied taxis will become off-shift immediately while the occupied taxis will become off-shift after finishing the current committed rides. For the off-shift taxis, a certain number of them will be randomly selected to start shift based on the shift start time distribution curve. To avoid having taxis become in-shift again immediately at the end of a previous shift, we implemented a “rest time” of 0 to 3 hours (uniform distribution) before it can be considered for being selected to start a shift. Only in-shift taxis can receive requests from the rider groups.

A.3.4.1 Variables

- *shiftStart* : Distribution of shift starts with time. Generated using the data obtained from the New York Taxi Cab fact book Figure A.2
- *OnShift* : True if taxi is in operation, False if not
- *inTaxis* : Collection of taxis in operation. All taxis in *inTaxis* execute taxiAlgo
- *outTaxis* : Collection of taxis not in shift.
- *shift* : The number of taxis to start shift
- *sleepTime* : The minimum time between two shifts for a taxi.
- *shiftTime* : The minimum time a taxi stays in shift
- x_c : A counter that runs through all the taxis

A.3.4.2 Algorithm *shiftControl*

At Start:

1. Set *OnShiftPC* as the percent of taxis on shift at model start time as obtained from *shiftStart*
2. For each taxi in *taxis*
3. Generate a uniform random number *rand*; If $rand > OnShiftPC$ set $OnShift = TRUE$ else $OnShift = FALSE$
4. If $OnShift = TRUE$

5. Select shift start time based on previous days entry distribution *shiftStart* and set shift end time to occur 8 hours later
6. add taxi to *inTaxis*
7. Else
8. randomly draw a sleep time *sleepTime* from uniform distribution between 0 to 3 hours.
9. After sleep time expired, add to *outTaxis*

During run time: This algorithm runs inside the main function every 15 minutes

1. Draw number of taxis (*shift*) to start shift from *shiftStart*
2. Send message to random *shift* taxis from *outTaxis* to begin shift
3. For all $x_c \in inTaxis$
4. If time taxi is in *inTaxis* $>$ *shiftTime*
5. If $x_c.State = 2$
6. remove t from *taxis_shr_man*, *taxis_shr_brook*, *taxis_shr_else*
7. set *canShare* = *False*
8. Else
9. remove x_c from *inTaxis*
10. add x_c to *outTaxis*

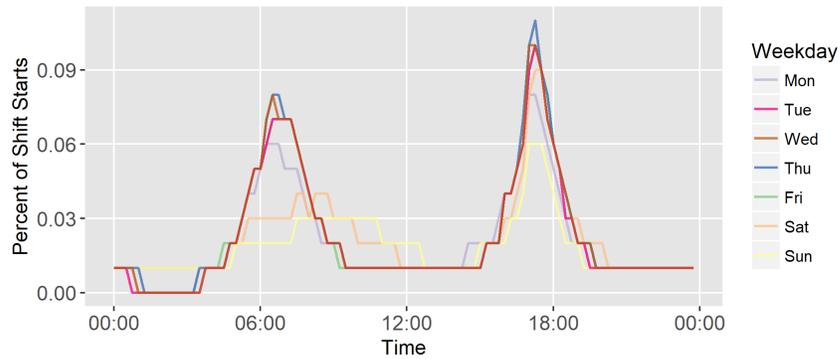


Figure A.2. Percent of taxis which start their shift at different times of the day (NYC DOT, 2014)

A.3.5 Electric Vehicles Charging

All electric vehicles in the system require charging as per their range. In order to facilitate this I have added chargers to the model from the Department of Energy (2018) data set. The electric taxis go towards the charger if one of the two conditions are met: 1) The charge level is below the minimum operating level as defined by *batteryThreshold* or 2) the charging utility is positive. The charging utility is a score that considers battery remaining and closeness to a charger.

A.3.5.1 Variables

- *batteryRemaining* : current state of charge in percentage
- *distance* : distance traveled since last update
- *fullChargeDistance* : distance that the EV can travel on a full charge
- *utility* : the utility function that requires

- *fullChargeTime* : Time required for full charge
- *chargeStation* : the charge station selected for the taxi
- *b* : a set of 5 coefficients to weigh the charging decision. *b*[0] : intercept *b*[1] : distance factor *b*[2] : distance offset *b*[3] : battery percent weight *b*[4]: Battery threshold
- X_1 : distance of nearest charger in terms of number of grid cells
- X_2 : = *batteryRemaining*

A.3.5.2 Algorithm

1. For every EV taxi *t* in *Taxis*
2. If taxi is moving update $batteryRemaining = batteryRemaining - \frac{distance}{fullChargeDistance}$
3. If *state* \neq 7 and *state* \neq 8
4. If *chargingUtility*() $>$ 0
5. If *state* \neq 2.5 find nearest charging Station as *chargeStation*
6. Move to *chargeStation*, remove self from *sharr* and *taxiArr* and set *State* = 7 (Move to Charging)
7. Else Set *canShare* = *FALSE* and *State* = 2
8. Else accept changed *myRoute*
9. If *state* = 7

10. If taxi has reached charging station Set $chargeTime = (1 - batteryRemaining) \times fullChargeTime$
11. Set $State = 8$ (Charging)
12. If $state = 8$
13. If $tis \geq chargeTime$ set $state = 1$
14. Else $tis = tis + 1$
15. $t = t + 1$

A.3.5.3 chargingUtility()

1. If $batteryLevel < b[4]$ $utility \leftarrow -1$
2. Else $utility \leftarrow b[0] + b[1] \times (X_1 - b[2]) + b[3] \times (X_1 - b[4])$
3. Return utility

A.4 Speed Estimation

The average speed is needed to track the travel time for the shared trips, especially for the deviated part of the trip due to sharing, for which we do not have information directly from the historical trip data (as illustrated in Figure A.4(a)). To estimate the travel speed and time more accurately, we could use the collective information of all trips to learn about the traveling speed on the road segments. Therefore, I estimated the speed for each road segment traveled by the taxis by simulating the trips from the historical taxi trip data. For each trip, based on the pick up and drop

off locations, I used OpenStreetMaps to identify the shortest path between these two points. I used the distance and the original trip time to calculate the average speed of this trip. The taxi will then "drop" points along the selected path as it travels (Figure A.4(b)). So each road segment will then store the speed points from this taxi/trip. I ran this for all trips in one day (May 7, 2014, the same day our analysis is based on). Then each road segment will then store speed data from different trips. I then used the average speed as the speed for this road segment.

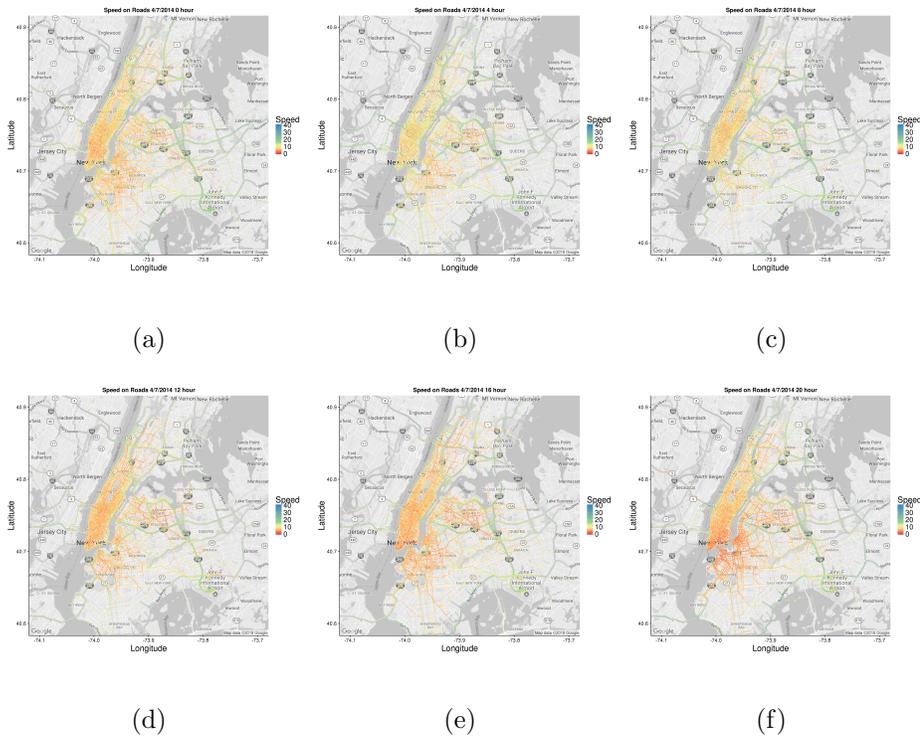


Figure A.3. Speeds on NYC roads as estimated from the data set

Table A.2.
Speeds on NYC roads (m/s)

Hour	Times Square	Manhattan	Brooklyn	HHP	JFK	Belt Pkwy	Bridges
0	7.76	8.08	9.85	11.10	16.70	12.29	11.04
1	8.04	8.42	10.09	11.25	16.73	11.98	11.18
2	8.46	8.76	9.85	11.70	18.74	12.02	11.05
3	8.94	9.25	10.22	12.04	15.96	12.77	11.95
4	9.58	10.85	12.84	13.48	18.45	15.64	14.68
5	9.22	10.49	14.52	12.83	17.99	15.22	14.06
6	8.77	9.79	14.21	12.81	17.70	14.56	13.64
7	8.21	9.20	13.64	11.83	17.19	13.99	12.88
8	7.55	8.49	12.57	11.22	17.29	13.62	12.44
9	7.28	8.03	12.05	10.52	16.72	13.10	11.56
10	6.83	7.60	11.52	10.21	16.28	12.93	11.37
11	6.69	7.36	10.57	9.95	15.38	11.99	11.22
12	6.35	7.28	10.66	9.68	14.48	11.92	12.37
13	6.00	6.80	9.76	9.22	13.38	11.45	10.34
14	5.97	6.93	9.66	9.24	14.13	11.19	10.80
15	6.40	7.01	9.70	8.74	15.33	11.64	10.35
16	6.83	7.26	10.63	8.93	16.15	11.99	10.64
17	6.72	7.10	11.44	9.04	16.08	11.72	11.50
18	6.77	6.92	9.62	8.84	15.80	10.46	10.03
19	6.92	6.62	8.88	8.49	15.26	9.24	9.68
20	6.99	6.76	9.78	8.88	15.16	9.44	10.04
21	6.14	6.51	7.51	8.80	14.96	8.55	8.94
22	6.16	6.71	7.64	9.20	14.67	9.39	9.50
23	6.94	7.57	9.70	12.32	16.15	12.07	10.55
Mean	6.62	6.92	9.35	9.31	15.59	10.61	10.14
(Peak Time)							

APPENDIX B

SUPPLEMENTARY ANALYSIS FOR STUDYING RIDE SHARING IN SHARED AUTONOMOUS VEHICLES AND TRADITIONAL TAXI CABS

B.1 Sensitivity of Busy Start Percent

We tested different values for the initial fraction of busy taxis ranging from 0 to 1. We found out that the initial fraction of busy taxis does not affect the output of the simulation. Figure B.1 shows that both the fraction of served riders and the fraction of occupied distance stay the same regardless of the different initial fractions of busy taxis (same for the other outputs, so we didn't show all figures). As shown in Figure 2.4(a) in the manuscript, the demand from midnight to early morning is very low. Hence, the system has an opportunity to “self balance” no matter how many busy taxis we started the day with. The busy taxis will have plenty of time before the morning peak to become available.

B.2 Explanation of Spatial Coverage Changes

Our results showed that as a result of sharing and autonomous vehicles, there was a reduction in service coverage in the suburban regions. In order to further explore the change of taxi movement in these scenarios, I calculated the radius of gyration and the corresponding center of mass of each individual taxis (Gonzalez et al., 2008;

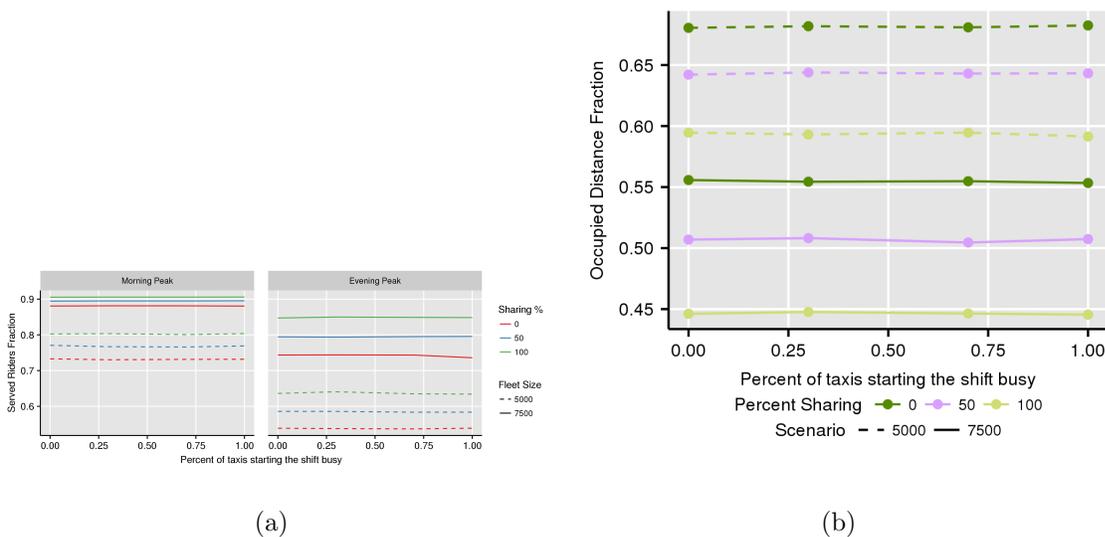
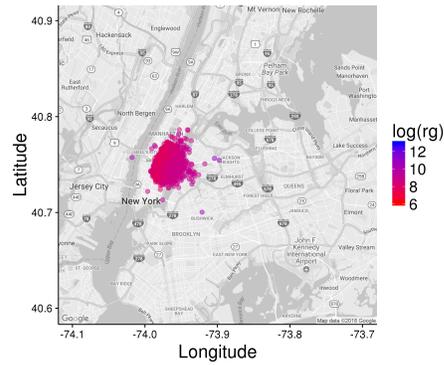


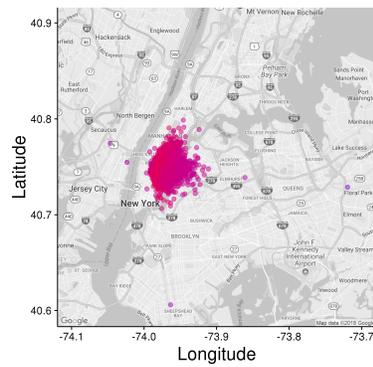
Figure B.1. Sensitivity of busy start percent

Cai et al., 2016) (the center of mass represents the center location of each taxi's trajectory and the radius of gyration represents the weighted average of the distance the taxi travel away from the center of mass). From Figure B.2, we can see that the spread for the center of masses follow the order ($A < S < B$), in the same ranking of the fleet size from the smallest to the largest.

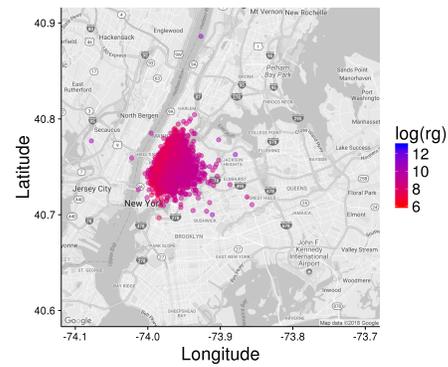
It is also notable that Scenario B has a few vehicles whose center of mass lies within Brooklyn and has a fairly large radius of gyration (which means large area of coverage). Even though the service level (the total number of rider groups served) is maintained, the reduced fleet focuses its service more in the Manhattan area. Potential policies to remedy this impact include providing price intensive for serving suburban area or dedicate a portion of the fleet for the suburban area.



(a)



(b)



(c)

Figure B.2. Radius of gyration

B.3 Statistical Significance Tests

For each of the scenarios presented in Table 3.2 we ran three additional runs to verify the statistical significance of the results that we obtained. The statistical significance tests for these results are presented in this section.

- The service level (percent of riders that were served by the system) are derived by counting the number of riders that were served and dividing it by the total number of riders. Since the rider groups can either be served or not served (0 or 1) a Pearsons Chi-Squared test is used to test for statistical significance.
 - B
 - Pearson's Chi-squared test
 - data: CM X-squared = 13.458, df = 2, p-value = 0.1196
 - A
 - Pearson's Chi-squared test
 - data: CM X-squared = 2.9532, df = 2, p-value = 0.2284
 - A2
 - Pearson's Chi-squared test
 - data: CM X-squared = 1.05, df = 2, p-value = 0.5916
 - A3
 - Pearson's Chi-squared test
 - data: CM X-squared = 0.42664, df = 2, p-value = 0.8079
 - S
 - Pearson's Chi-squared test
 - data: CM X-squared = 13.892, df = 2, p-value = 0.09625
- For all other results (the waiting time, riding time, the total distance travelled and the distance fraction, we use an Anova to test whether the scenario group is a significant factor and whether the random simulation seed is a significant factor. These results are presented in Figure B.3.

```

> summary(WaitingTimeAOV)
      Df Sum Sq Mean Sq F value Pr(>F)
Scenario 4 2.911e+09 727802854 5794.828 <2e-16 ***
Seeds    2 1.895e+05   94740    0.754  0.47
Residuals 2528671 3.176e+11 125595
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> RidingTimeAOV<-aov(RidingTime-Taxis+Seeds, data = distancesAgg)
> summary(RidingTimeAOV)
      Df Sum Sq Mean Sq F value Pr(>F)
Scenario 4 1.835e+11 4.587e+10 27748.887 <2e-16 ***
Seeds    2 6.085e+06 3.043e+06    1.841  0.159
Residuals 2528671 4.180e+12 1.653e+06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(a)

```

summary(TotalDistanceSumAOV)
      Df Sum Sq Mean Sq F value Pr(>F)
Scenario 4 1.611e+09 402793645 10387.180 <2e-16 ***
Seeds    2 1.888e+05   94394    2.434  0.0877 .
Residuals 119993 4.653e+09 38778
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
>
summary(dtDiffAOV)
      Df Sum Sq Mean Sq F value Pr(>F)
Scenario 4 135809 33952 4.268e+05 <2e-16 ***
Seed     1 0 0 4.810e-01 0.488
Residuals 6500359 517107 0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(b)

Figure B.3. ANOVA to test for the statistical significance of the results with the Scenario and the random simulation Seeds (Seed) for (a) waiting time and riding time (b) distance travelled and deviation tolerance difference.

We can see that in each of the cases (the Pearsons Chi Squared test and the Anova) the Seed variable (which refers to the separate simulation runs) are not statistically significant at the 95% confidence level.

APPENDIX C
SCENARIOS GENERATED BY SIMULATION
OPTIMIZATION

Table C.1.: All runs for case study in Section 4.4

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
1	5000	1.00	1.00	2	0.25	1	
2	6000	1.00	0.80	2	0.25	1	
3	5000	0.80	0.80	2	0.25	1	
4	5000	0.80	1.00	2	0.75	1	
5	6000	0.80	1.00	2	0.25	1	
6	5000	1.00	0.80	2	0.75	1	
7	6000	1.00	1.00	2	0.75	1	
8	6000	0.80	0.80	2	0.75	1	
9	5000	1.00	1.00	4	0.25	1	
10	6000	1.00	0.80	4	0.25	1	
11	5000	0.80	0.80	4	0.25	1	
12	5000	0.80	1.00	4	0.75	1	
13	6000	0.80	1.00	4	0.25	1	✓
14	5000	1.00	0.80	4	0.75	1	
15	6000	1.00	1.00	4	0.75	1	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
16	6000	0.80	0.80	4	0.75	1	
17	5000	1.00	1.00	6	0.25	1	
18	6000	1.00	0.80	6	0.25	1	
19	5000	0.80	0.80	6	0.25	1	
20	5000	0.80	1.00	6	0.75	1	
21	6000	0.80	1.00	6	0.25	1	✓
22	5000	1.00	0.80	6	0.75	1	
23	6000	1.00	1.00	6	0.75	1	✓
24	6000	0.80	0.80	6	0.75	1	
25	5500	0.90	1.00	2	0.25	1	
26	6500	0.90	0.80	2	0.25	1	
27	5500	0.60	0.80	2	0.25	1	
28	5500	0.60	1.00	2	0.75	1	
29	6500	0.60	1.00	2	0.25	1	
30	5500	0.90	0.80	2	0.75	1	
31	6500	0.90	1.00	2	0.75	1	
32	6500	0.60	0.80	2	0.75	1	
33	5500	0.90	1.00	4	0.25	1	
34	6500	0.90	0.80	4	0.25	1	✓
35	5500	0.60	0.80	4	0.25	1	
36	5500	0.60	1.00	4	0.75	1	
37	6500	0.60	1.00	4	0.25	1	✓
38	5500	0.90	0.80	4	0.75	1	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
39	6500	0.90	1.00	4	0.75	1	✓
40	6500	0.60	0.80	4	0.75	1	
41	5500	0.90	1.00	6	0.25	1	
42	6500	0.90	0.80	6	0.25	1	✓
43	5500	0.60	0.80	6	0.25	1	
44	5500	0.60	1.00	6	0.75	1	
45	6500	0.60	1.00	6	0.25	1	✓
46	5500	0.90	0.80	6	0.75	1	
47	6500	0.90	1.00	6	0.75	1	
48	6500	0.60	0.80	6	0.75	1	✓
49	6500	0.40	1.00	2	0.25	1	
50	7500	0.40	0.80	2	0.25	1	
51	6500	0.10	0.80	2	0.25	1	
52	6500	0.10	1.00	2	0.75	1	
53	7500	0.10	1.00	2	0.25	1	
54	6500	0.40	0.80	2	0.75	1	
55	7500	0.40	1.00	2	0.75	1	✓
56	7500	0.10	0.80	2	0.75	1	
57	6500	0.40	1.00	4	0.25	1	
58	7500	0.40	0.80	4	0.25	1	✓
59	6500	0.10	0.80	4	0.25	1	
60	6500	0.10	1.00	4	0.75	1	
61	7500	0.10	1.00	4	0.25	1	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
62	6500	0.40	0.80	4	0.75	1	
63	7500	0.40	1.00	4	0.75	1	✓
64	7500	0.10	0.80	4	0.75	1	✓
65	6500	0.40	1.00	6	0.25	1	✓
66	7500	0.40	0.80	6	0.25	1	✓
67	6500	0.10	0.80	6	0.25	1	
68	6500	0.10	1.00	6	0.75	1	
69	7500	0.10	1.00	6	0.25	1	✓
70	6500	0.40	0.80	6	0.75	1	
71	7500	0.40	1.00	6	0.75	1	
72	7500	0.10	0.80	6	0.75	1	✓
73	11500	0.30	0.30	2	0.25	1	✓
74	13500	0.30	0.00	2	0.25	1	✓
75	11500	0.00	0.00	2	0.25	1	✓
76	11500	0.00	0.30	2	0.75	1	✓
77	13500	0.00	0.30	2	0.25	1	✓
78	11500	0.30	0.00	2	0.75	1	✓
79	13500	0.30	0.30	2	0.75	1	✓
80	13500	0.00	0.00	2	0.75	1	✓
81	11500	0.30	0.30	4	0.25	1	
82	13500	0.30	0.00	4	0.25	1	
83	11500	0.00	0.00	4	0.25	1	✓
84	11500	0.00	0.30	4	0.75	1	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
85	13500	0.00	0.30	4	0.25	1	
86	11500	0.30	0.00	4	0.75	1	
87	13500	0.30	0.30	4	0.75	1	
88	13500	0.00	0.00	4	0.75	1	
89	11500	0.30	0.30	6	0.25	1	
90	13500	0.30	0.00	6	0.25	1	
91	11500	0.00	0.00	6	0.25	1	
92	11500	0.00	0.30	6	0.75	1	
93	13500	0.00	0.30	6	0.25	1	
94	11500	0.30	0.00	6	0.75	1	
95	13500	0.30	0.30	6	0.75	1	
96	13500	0.00	0.00	6	0.75	1	
97	7500	1.00	0.30	2	0.25	1	
98	8500	1.00	0.00	2	0.25	1	
99	7500	0.80	0.00	2	0.25	1	
100	7500	0.80	0.30	2	0.75	1	
101	8500	0.80	0.30	2	0.25	1	✓
102	7500	1.00	0.00	2	0.75	1	
103	8500	1.00	0.30	2	0.75	1	✓
104	8500	0.80	0.00	2	0.75	1	
105	7500	1.00	0.30	4	0.25	1	✓
106	8500	1.00	0.00	4	0.25	1	✓
107	7500	0.80	0.00	4	0.25	1	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
108	7500	0.80	0.30	4	0.75	1	✓
109	8500	0.80	0.30	4	0.25	1	✓
110	7500	1.00	0.00	4	0.75	1	
111	8500	1.00	0.30	4	0.75	1	
112	8500	0.80	0.00	4	0.75	1	✓
113	7500	1.00	0.30	6	0.25	1	✓
114	8500	1.00	0.00	6	0.25	1	✓
115	7500	0.80	0.00	6	0.25	1	
116	7500	0.80	0.30	6	0.75	1	✓
117	8500	0.80	0.30	6	0.25	1	
118	7500	1.00	0.00	6	0.75	1	
119	8500	1.00	0.30	6	0.75	1	
120	8500	0.80	0.00	6	0.75	1	✓
121	13500	1.00	0.00	2	0.25	2	✓
122	13500	0.95	0.00	2	0.73	2	✓
123	13500	0.00	0.00	4	0.25	2	
124	13500	0.00	0.00	4	0.73	2	
125	13500	0.70	1.00	2	0.25	2	✓
126	13500	0.60	1.00	2	0.61	2	✓
127	13500	0.22	0.84	2	0.75	2	✓
128	6000	0.97	0.95	6	0.25	2	✓
129	6000	0.97	0.95	6	0.25	2	✓
130	6000	1.00	0.91	6	0.75	2	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
131	6000	1.00	0.91	6	0.75	2	✓
132	6003	1.00	0.91	6	0.75	2	✓
133	6003	1.00	0.91	6	0.75	2	✓
134	6911	1.00	1.00	2	0.75	2	✓
135	6000	0.96	0.94	6	0.75	2	✓
136	6000	0.96	0.94	6	0.75	2	✓
137	7000	1.00	0.99	2	0.25	2	✓
138	7000	1.00	0.99	2	0.25	2	✓
139	10000	0.00	0.00	6	0.25	2	✓
140	9997	0.00	0.00	6	0.25	2	✓
141	10000	0.00	0.00	6	0.25	2	✓
142	6000	1.00	1.00	4	0.25	2	✓
143	6000	1.00	1.00	4	0.25	2	✓
144	9939	0.00	0.00	6	0.25	2	✓
145	9939	0.00	0.00	6	0.25	2	✓
146	9939	0.00	0.00	6	0.25	2	✓
147	9939	0.00	0.00	6	0.25	2	✓
148	9939	0.00	0.00	6	0.25	2	✓
149	6000	0.86	1.00	6	0.75	2	✓
150	6000	0.86	1.00	6	0.75	2	✓
151	10000	1.00	0.00	2	0.25	2	✓
152	10000	0.00	0.00	6	0.75	2	✓
153	6004	0.86	1.00	6	0.75	2	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
154	10000	0.00	0.00	6	0.75	2	✓
155	10000	1.00	0.00	2	0.25	2	✓
156	6000	1.00	1.00	4	0.75	2	✓
157	6000	1.00	1.00	4	0.75	2	✓
158	6005	0.86	1.00	6	0.75	2	✓
159	6005	0.86	1.00	6	0.75	2	✓
160	11000	0.92	0.00	2	0.25	2	✓
161	6094	1.00	1.00	4	0.75	2	✓
162	6094	1.00	1.00	4	0.75	2	✓
163	6094	1.00	1.00	4	0.75	2	✓
164	6094	1.00	1.00	4	0.75	2	✓
165	6094	1.00	1.00	4	0.75	2	✓
166	7000	1.00	0.96	2	0.75	2	✓
167	7000	1.00	0.96	2	0.75	2	✓
168	7003	1.00	0.96	2	0.75	2	✓
169	10000	0.99	0.00	2	0.75	2	✓
170	10000	0.99	0.00	2	0.75	2	✓
171	7000	0.05	1.00	6	0.75	2	✓
172	7000	0.06	1.00	6	0.25	2	✓
173	7000	0.06	1.00	6	0.25	2	✓
174	7000	0.99	0.97	2	0.75	2	✓
175	7000	0.99	0.97	2	0.75	2	✓
176	13500	0.22	0.35	2	0.75	2	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
177	11000	0.91	0.00	2	0.73	2	✓
178	11000	0.91	0.00	2	0.73	2	✓
179	11000	0.00	0.00	4	0.25	2	✓
180	11000	0.00	0.00	4	0.25	2	✓
181	10998	0.00	0.00	4	0.25	2	✓
182	7000	0.93	1.00	2	0.75	2	✓
183	7000	0.93	1.00	2	0.75	2	✓
184	7004	0.92	1.00	2	0.75	2	✓
185	7004	0.92	1.00	2	0.75	2	✓
186	7004	0.92	1.00	2	0.75	2	✓
187	7005	0.92	1.00	2	0.75	2	✓
188	13500	0.33	0.77	2	0.36	2	✓
189	11000	0.00	0.00	4	0.75	2	✓
190	11000	0.00	0.00	4	0.75	2	✓
191	10740	0.00	0.00	4	0.25	2	✓
192	10740	0.00	0.00	4	0.25	2	✓
193	10740	0.00	0.00	4	0.25	2	✓
194	10740	0.00	0.00	4	0.25	2	✓
195	8000	1.00	0.07	6	0.75	2	✓
196	8000	0.03	1.00	2	0.75	2	✓
197	8000	1.00	0.09	6	0.25	2	✓
198	8000	1.00	0.09	6	0.25	2	✓
199	10717	0.00	0.01	4	0.25	2	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
200	8000	0.05	1.00	2	0.26	2	✓
201	8000	0.05	1.00	2	0.26	2	✓
202	8000	1.00	0.09	6	0.75	2	✓
203	6000	1.00	0.90	4	0.75	2	✓
204	6003	1.00	0.90	4	0.75	2	✓
205	13500	0.26	0.43	2	0.34	2	✓
206	7000	0.17	1.00	6	0.75	2	✓
207	6000	0.65	1.00	6	0.27	2	✓
208	6000	0.65	1.00	6	0.27	2	✓
209	6004	0.65	1.00	6	0.27	2	✓
210	11000	0.00	0.22	2	0.25	2	✓
211	11000	0.00	0.22	2	0.25	2	✓
212	8000	0.00	0.86	6	0.25	2	
213	11000	0.06	0.14	2	0.75	2	✓
214	11000	0.06	0.14	2	0.75	2	✓
215	7003	1.00	0.79	2	0.75	2	✓
216	8000	0.14	1.00	2	0.75	2	✓
217	10499	0.09	0.00	4	0.25	2	✓
218	11000	0.07	0.18	2	0.26	2	✓
219	11000	0.07	0.18	2	0.26	2	✓
220	9000	0.79	0.00	6	0.25	2	✓
221	9000	0.95	0.00	4	0.25	2	✓
222	8000	0.00	0.95	4	0.25	2	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
223	5000	1.00	1.00	2	1.00	2	
224	6500	0.90	0.80	4	1.00	2	✓
225	7500	0.40	0.80	2	1.00	2	
226	7500	0.10	1.00	4	1.00	2	✓
227	11500	0.30	0.30	2	1.00	2	✓
228	13500	0.30	0.00	2	1.00	2	✓
229	13500	0.30	0.00	4	1.00	2	
230	7500	1.00	0.30	2	1.00	2	
231	8500	0.80	0.30	2	1.00	2	✓
232	8500	0.80	0.30	4	1.00	2	✓
233	5000	1.00	1.00	2	0.00	2	
234	6500	0.90	0.80	4	0.00	2	✓
235	7500	0.40	0.80	2	0.00	2	
236	7500	0.10	1.00	4	0.00	2	✓
237	11500	0.30	0.30	2	0.00	2	✓
238	13500	0.30	0.00	2	0.00	2	✓
239	13500	0.30	0.00	4	0.00	2	
240	7500	1.00	0.30	2	0.00	2	
241	8500	0.80	0.30	2	0.00	2	
242	8500	0.80	0.30	4	0.00	2	✓
243	9000	0.00	0.38	6	0.00	3	✓
244	13000	0.85	1.00	2	1.00	3	✓
245	13000	0.00	0.77	2	0.05	3	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
246	8500	0.50	0.68	2	1.00	3	✓
247	9000	0.08	0.30	6	1.00	3	✓
248	10000	0.00	0.75	2	0.00	3	✓
249	8000	0.31	0.76	2	0.04	3	
250	9998	0.33	0.00	6	0.00	3	✓
251	10000	0.32	0.00	6	0.97	3	✓
252	7501	0.63	0.79	2	0.52	3	✓
253	10500	0.49	0.47	2	0.50	3	✓
254	11000	0.00	0.60	2	0.98	3	✓
255	8000	0.65	0.34	6	0.00	3	✓
256	9000	0.00	0.55	4	0.00	3	✓
257	7500	0.53	0.55	4	0.50	3	✓
258	8000	0.00	0.74	6	0.00	3	✓
259	13000	0.00	0.00	4	1.00	3	
260	9500	0.50	0.42	2	0.00	3	✓
261	7006	0.68	1.00	2	1.00	3	✓
262	7006	0.67	1.00	2	0.11	3	✓
263	10000	0.02	0.19	4	1.00	3	✓
264	8000	0.29	0.67	4	1.00	3	✓
265	8000	1.00	0.45	2	0.02	3	✓
266	10000	0.00	0.24	4	0.00	3	✓
267	8000	0.31	0.52	4	0.02	3	✓
268	7000	1.00	0.37	6	1.00	3	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
269	7000	1.00	0.47	4	1.00	3	✓
270	7000	0.40	0.85	6	1.00	3	✓
271	6501	0.71	0.84	4	0.52	3	✓
272	8000	1.00	0.62	2	1.00	3	✓
273	9000	0.49	0.38	4	0.96	3	
274	10000	0.72	0.19	2	0.00	3	✓
275	13000	0.00	0.00	4	0.00	3	
276	6999	0.84	0.62	6	0.00	3	✓
277	7999	0.69	0.60	4	0.97	3	
278	9000	0.00	0.75	2	0.98	3	✓
279	9000	0.51	0.21	4	0.00	3	✓
280	13000	0.09	0.25	2	1.00	3	✓
281	8999	0.63	0.00	6	0.00	3	✓
282	8000	0.25	0.44	6	0.02	3	
283	7500	0.50	0.58	6	1.00	3	✓
284	11002	0.57	0.36	2	0.95	3	✓
285	8000	0.29	0.95	2	1.00	3	✓
286	9000	0.93	0.00	4	1.00	3	✓
287	13000	0.00	0.31	2	0.55	3	✓
288	13000	0.08	0.24	2	0.00	3	✓
289	10000	0.63	0.00	4	0.00	3	✓
290	10000	0.29	0.23	4	0.96	3	
291	13000	0.96	0.00	2	1.00	3	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
292	10000	0.36	0.00	4	0.00	3	✓
293	13000	0.95	0.00	2	0.00	3	✓
294	13000	0.85	1.00	2	0.04	3	✓
295	8001	0.25	0.44	6	1.00	3	✓
296	7001	0.39	1.00	4	1.00	3	✓
297	6501	0.66	0.77	6	1.00	3	✓
298	10000	0.00	0.49	2	0.00	3	✓
299	10000	1.00	0.05	2	0.00	3	✓
300	10000	0.01	0.02	6	1.00	3	✓
301	10000	1.00	0.02	2	1.00	3	✓
302	5737	1.00	1.00	6	1.00	3	✓
303	8000	0.98	0.70	2	0.00	3	✓
304	10192	0.00	0.00	6	0.00	3	✓
305	7000	0.04	1.00	6	1.00	3	
306	7000	1.00	1.00	2	1.00	3	✓
307	8000	0.03	1.00	2	1.00	3	✓
308	8000	1.00	0.06	6	1.00	3	✓
309	6000	0.97	0.96	6	0.00	3	✓
310	9000	0.95	0.00	4	0.00	3	✓
311	8000	1.00	0.09	6	0.00	3	✓
312	7000	0.06	1.00	6	0.01	3	
313	8000	0.05	1.00	2	0.01	3	✓
314	12500	0.51	0.00	2	0.00	3	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
315	7000	1.00	1.00	2	0.01	3	✓
316	10000	0.00	0.43	4	0.97	3	
317	11000	0.69	0.00	2	0.00	3	✓
318	10000	0.07	0.42	2	1.00	3	✓
319	7000	0.48	0.82	4	0.04	3	✓
320	8000	1.00	0.20	4	1.00	3	✓
321	6000	1.00	1.00	4	0.00	3	✓
322	6000	1.00	1.00	4	1.00	3	✓
323	9000	0.82	0.64	2	0.97	3	✓
324	9000	0.18	0.46	2	0.02	3	✓
325	10499	0.16	0.00	4	0.00	3	✓
326	9000	0.84	0.42	2	0.00	3	✓
327	9000	1.00	0.29	2	1.00	3	✓
328	10000	0.70	0.45	2	0.96	3	✓
329	9001	0.16	0.26	4	0.01	3	✓
330	7003	0.48	0.81	4	1.00	3	✓
331	8000	0.00	0.86	6	0.99	3	✓
332	9499	0.29	0.18	6	0.00	3	✓
333	8500	0.50	0.40	4	0.00	3	✓
334	9000	0.16	0.26	4	1.00	3	✓
335	12002	0.00	0.63	2	0.00	3	✓
336	9000	0.46	0.10	6	0.00	3	✓
337	9999	0.29	0.23	4	0.00	3	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
338	9001	0.16	0.66	2	1.00	3	✓
339	6000	0.82	1.00	6	1.00	3	✓
340	12000	0.05	0.14	2	1.00	3	✓
341	6999	0.82	0.73	6	0.99	3	
342	8000	0.95	0.90	2	0.00	3	✓
343	7999	0.72	0.44	4	0.00	3	✓
344	12002	0.00	0.18	2	0.00	3	✓
345	11000	0.00	0.28	2	0.00	3	✓
346	6003	1.00	0.80	6	0.04	3	
347	9001	0.14	0.42	4	1.00	3	✓
348	8000	0.00	0.90	4	0.99	3	✓
349	11000	0.08	0.20	2	1.00	3	✓
350	8002	0.31	0.52	4	1.00	3	✓
351	12000	0.89	0.00	2	1.00	3	✓
352	9500	0.38	0.33	4	0.00	3	✓
353	12000	0.90	0.00	2	0.00	3	✓
354	6000	0.67	1.00	6	1.00	3	✓
355	11000	0.94	0.00	2	0.00	3	✓
356	11000	0.93	0.00	2	1.00	3	✓
357	9000	0.00	0.79	2	0.00	3	✓
358	7000	1.00	0.86	2	1.00	3	✓
359	6000	0.86	1.00	4	1.00	3	✓
360	12000	0.00	0.00	4	0.00	3	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
361	12000	0.00	0.00	4	1.00	3	
362	11000	0.00	0.00	4	1.00	3	✓
363	11000	0.00	0.00	4	0.00	3	✓
364	8000	0.16	1.00	2	1.00	3	✓
365	8000	0.00	0.94	4	0.00	3	✓
366	8998	0.44	0.24	6	0.00	3	✓
367	12800	1.00	0.65	2	1.00	3	✓
368	12700	1.00	0.55	2	0.00	3	✓
369	8300	0.30	0.35	6	1.00	3	✓
370	12600	0.00	1.00	2	0.00	4	✓
371	8900	0.00	0.45	4	0.00	4	✓
372	9900	0.50	0.55	2	0.60	4	✓
373	12700	1.00	0.55	2	0.50	4	✓
374	10300	0.05	0.65	2	0.45	4	✓
375	10100	0.50	0.00	4	0.00	4	✓
376	10200	0.00	0.65	2	1.00	4	✓
377	13500	0.50	0.60	2	1.00	4	✓
378	10000	0.55	0.05	2	0.45	4	✓
379	12600	0.00	1.00	2	0.45	4	✓
380	12800	1.00	1.00	2	0.55	4	✓
381	9100	0.75	0.70	2	0.05	4	✓
382	12700	0.50	1.00	2	1.00	4	✓
383	8400	0.40	0.30	4	0.15	4	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
384	10600	0.60	0.00	2	1.00	4	✓
385	6900	1.00	0.55	6	1.00	4	✓
386	8900	0.75	0.70	2	1.00	4	✓
387	8900	0.00	0.45	4	0.40	4	✓
388	8300	0.30	0.35	6	0.60	4	✓
389	10500	0.00	0.70	2	0.00	4	✓
390	12800	1.00	1.00	2	0.00	4	✓
391	12800	0.70	0.30	2	0.45	4	✓
392	12700	0.65	0.10	2	0.10	4	✓
393	12600	0.00	1.00	2	1.00	4	✓
394	7600	0.60	0.60	4	0.00	4	✓
395	6900	1.00	0.55	6	0.55	4	✓
396	13500	0.55	0.00	2	0.65	4	✓
397	7200	0.65	0.55	6	0.45	4	✓
398	6800	1.00	0.55	6	0.00	4	✓
399	8400	0.10	0.40	6	0.00	4	
400	12400	0.45	0.60	2	0.10	4	✓
401	12200	0.70	0.35	2	0.80	4	✓
402	8400	0.40	0.20	6	0.00	4	
403	7400	0.55	0.50	6	0.00	4	✓
404	10200	1.00	0.30	2	0.45	4	✓
405	12800	1.00	1.00	2	1.00	4	✓
406	7400	1.00	0.60	4	0.55	4	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
407	13500	0.15	0.55	2	0.00	4	✓
408	13400	0.00	0.25	2	1.00	4	✓
409	9000	0.00	0.50	6	1.00	4	✓
410	9600	0.00	0.30	2	1.00	4	
411	10000	0.55	0.00	4	1.00	4	✓
412	8900	1.00	0.60	2	0.45	4	✓
413	8400	0.30	0.35	4	0.60	4	
414	10200	0.55	0.00	4	0.40	4	✓
415	10800	0.95	0.20	2	1.00	4	✓
416	10000	0.55	0.00	2	0.05	4	
417	9200	0.00	0.25	4	1.00	4	
418	12000	0.00	0.65	2	0.75	4	✓
419	12900	0.15	0.15	2	0.50	4	✓
420	10600	0.20	0.00	2	0.00	4	
421	8900	0.15	0.50	2	0.10	4	
422	10600	1.00	0.25	2	0.00	4	✓
423	7100	1.00	0.50	4	0.95	4	✓
424	7100	1.00	0.50	4	0.00	4	✓
425	8700	0.50	0.90	2	0.50	4	✓
426	13000	0.00	0.10	2	0.00	4	✓
427	9100	1.00	0.00	4	1.00	4	✓
428	7100	0.60	1.00	4	1.00	4	✓
429	7000	0.25	0.90	6	0.00	4	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
430	7000	0.70	0.85	6	0.00	4	✓
431	8000	0.90	0.15	4	0.50	4	✓
432	12700	0.65	0.85	2	0.00	4	✓
433	11900	0.00	0.00	2	1.00	4	✓
434	7200	0.75	0.90	4	0.50	4	✓
435	8700	0.05	0.45	4	0.75	4	
436	9400	0.50	0.20	2	0.85	4	
437	6600	0.70	0.95	6	1.00	4	
438	8200	0.00	1.00	4	0.55	4	✓
439	7900	0.75	0.60	4	0.85	4	✓
440	8000	0.85	0.15	6	0.00	4	✓
441	7000	0.70	0.55	6	1.00	4	✓
442	8400	0.30	0.35	4	0.90	4	
443	12200	0.05	0.00	4	1.00	4	✓
444	8300	0.85	0.15	6	0.50	4	
445	8300	0.65	0.95	2	0.90	4	✓
446	9400	0.50	0.20	4	0.80	4	✓
447	10100	0.05	0.15	6	0.00	4	✓
448	11200	0.60	0.00	2	0.70	4	✓
449	8100	0.95	0.90	2	0.50	4	✓
450	7000	0.50	0.90	6	0.50	4	✓
451	8800	0.30	1.00	2	0.10	4	✓
452	11600	0.15	0.00	4	0.50	4	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
453	13500	0.45	0.25	2	1.00	4	✓
454	7900	0.00	0.90	6	1.00	4	✓
455	12800	1.00	0.85	2	0.30	4	✓
456	8600	0.55	0.30	6	0.80	4	
457	10300	0.20	0.05	2	1.00	4	
458	12500	0.25	0.80	2	1.00	4	✓
459	12600	0.50	0.80	2	0.80	4	✓
460	8200	0.00	0.90	2	0.00	4	
461	7100	0.70	1.00	4	0.00	4	✓
462	12500	0.00	0.00	4	0.00	4	✓
463	13400	0.75	0.90	2	1.00	4	✓
464	12700	0.45	1.00	2	0.40	4	✓
465	9900	0.20	0.20	4	0.80	4	✓
466	11000	0.25	0.05	4	0.00	4	✓
467	10000	0.55	0.55	2	0.10	4	✓
468	7000	0.25	0.85	6	1.00	4	✓
469	8700	0.95	0.00	6	1.00	4	
470	10600	0.75	0.35	2	0.35	4	✓
471	7900	0.95	1.00	2	0.00	4	✓
472	9500	0.40	0.20	2	0.15	4	
473	8000	1.00	0.10	4	0.00	4	
474	8900	0.55	0.40	4	0.35	4	✓
475	8000	0.25	0.90	4	0.50	4	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
476	9400	1.00	0.45	2	0.95	4	✓
477	9400	0.20	0.80	2	0.95	4	✓
478	10700	0.00	0.05	2	0.50	4	
479	9400	0.25	0.55	2	0.85	4	✓
480	9100	0.00	0.20	6	0.60	4	
481	7200	0.00	0.90	6	0.50	4	
482	6900	0.35	0.90	4	1.00	4	
483	12100	0.70	0.35	2	0.15	4	✓
484	9300	0.25	0.85	2	0.45	4	✓
485	8900	0.00	0.75	4	0.50	4	✓
486	13400	0.00	1.00	2	0.70	4	✓
487	12400	0.80	0.10	2	1.00	4	✓
488	9600	0.25	0.00	4	0.50	4	
489	7400	0.20	1.00	2	1.00	4	
490	13400	0.75	0.50	2	0.30	4	✓
491	8950	0.43	0.03	6	0.43	5	✓
492	11250	0.73	0.78	2	0.48	5	✓
493	8950	0.03	0.58	6	0.53	5	
494	11650	0.33	0.93	2	0.08	5	✓
495	9650	0.18	0.33	2	0.53	5	
496	10950	0.28	0.98	2	0.73	5	✓
497	13450	0.53	0.53	2	0.63	5	✓
498	10850	0.93	0.63	2	0.23	5	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
499	13450	0.03	0.68	2	0.48	5	✓
500	8750	0.18	0.23	6	0.33	5	
501	11950	0.28	0.63	2	0.58	5	✓
502	13450	0.98	0.33	2	0.88	5	✓
503	9350	0.73	0.93	2	0.28	5	✓
504	7750	0.38	0.48	6	0.33	5	✓
505	8450	0.58	0.03	6	0.98	5	✓
506	12350	0.98	0.28	2	0.28	5	✓
507	9050	0.23	0.63	4	0.28	5	✓
508	11750	0.43	0.23	2	0.58	5	✓
509	8050	0.73	0.63	4	0.28	5	✓
510	8450	0.73	0.48	2	0.53	5	✓
511	13450	0.08	0.58	2	0.98	5	✓
512	12450	0.83	0.03	2	0.48	5	✓
513	8650	0.33	0.68	4	0.68	5	✓
514	13450	0.48	0.33	2	0.03	5	✓
515	11150	0.43	0.63	2	0.98	5	✓
516	10350	0.03	0.23	4	0.48	5	✓
517	7950	0.48	0.58	6	0.68	5	
518	8450	0.58	0.18	4	0.53	5	✓
519	13350	0.73	0.58	2	0.03	5	✓
520	9450	0.83	0.08	2	0.48	5	✓
521	8450	0.78	0.43	4	0.53	5	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
522	9950	0.48	0.83	2	0.78	5	✓
523	9750	0.03	0.68	2	0.73	5	✓
524	8650	0.03	0.73	2	0.28	5	
525	12150	0.13	0.63	2	0.28	5	✓
526	8450	0.33	0.68	2	0.63	5	
527	12350	0.53	0.08	2	0.38	5	✓
528	9050	0.28	0.53	2	0.38	5	
529	11950	0.98	0.28	2	0.68	5	✓
530	9750	0.23	0.28	2	0.83	5	
531	9850	0.73	0.33	2	0.68	5	✓
532	11350	0.83	0.73	2	0.73	5	✓
533	9350	0.63	0.98	2	0.03	5	✓
534	9150	0.03	0.68	4	0.93	5	✓
535	6850	0.88	0.53	6	0.28	5	✓
536	13250	0.58	0.03	2	0.93	5	✓
537	6750	0.78	0.88	6	0.48	5	
538	9550	0.03	0.43	2	0.33	5	
539	9650	0.78	0.03	4	0.53	5	✓
540	8750	0.58	0.08	4	0.23	5	✓
541	7550	0.68	0.68	6	0.23	5	
542	8050	0.38	0.68	6	0.03	5	
543	6650	0.48	0.98	4	0.53	5	✓
544	11950	0.48	0.18	2	0.83	5	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
545	8950	0.48	0.43	2	0.98	5	✓
546	13450	0.23	0.98	2	0.53	5	✓
547	7550	0.73	0.58	6	0.73	5	
548	13450	0.03	0.03	2	0.33	5	✓
549	13450	0.28	0.03	2	0.68	5	✓
550	9650	0.73	0.08	2	0.23	5	✓
551	8550	0.03	0.78	2	0.53	5	
552	11050	0.33	0.68	2	0.33	5	✓
553	9250	0.18	0.23	4	0.33	5	✓
554	11150	0.53	0.23	2	0.08	5	✓
555	7450	0.83	0.38	6	0.48	5	✓
556	13450	0.73	0.68	2	0.63	5	✓
557	8350	0.63	0.13	4	0.98	5	✓
558	8250	0.18	0.68	6	0.98	5	✓
559	7050	0.73	0.63	4	0.63	5	✓
560	9350	0.78	0.08	2	0.98	5	
561	6550	0.98	0.83	4	0.53	5	✓
562	6750	0.23	0.98	6	0.48	5	
563	7750	0.23	0.98	2	0.53	5	
564	7550	0.68	0.98	2	0.68	5	✓
565	7850	0.33	0.98	4	0.08	5	✓
566	7050	0.98	0.68	4	0.78	5	✓
567	12150	0.23	0.13	2	0.13	5	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
568	9250	0.33	0.03	4	0.88	5	
569	11750	0.38	0.53	2	0.78	5	✓
570	11350	0.33	0.03	2	0.38	5	✓
571	6750	0.48	0.98	6	0.03	5	✓
572	6650	0.98	0.48	6	0.78	5	✓
573	9950	0.18	0.28	2	0.13	5	
574	13150	0.73	0.13	2	0.78	5	✓
575	7550	0.88	0.38	6	0.03	5	✓
576	8950	0.98	0.23	2	0.13	5	✓
577	8350	0.53	0.63	2	0.03	5	✓
578	5550	0.98	0.98	6	0.48	5	✓
579	9150	0.98	0.03	4	0.63	5	
580	13450	0.33	0.98	2	0.08	5	✓
581	10850	0.28	0.03	4	0.98	5	✓
582	13250	0.03	0.23	2	0.78	5	✓
583	8550	0.38	0.98	2	0.33	5	✓
584	9550	0.98	0.78	2	0.23	5	✓
585	13350	0.48	0.23	2	0.28	5	✓
586	13450	0.28	0.98	2	0.93	5	✓
587	9850	0.03	0.23	6	0.83	5	✓
588	11250	0.87	0.78	2	0.08	5	✓
589	8350	0.02	0.98	4	0.78	5	✓
590	10250	0.42	0.03	2	0.88	5	

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
591	7350	0.72	0.83	4	0.93	5	✓
592	9750	0.77	0.78	2	0.68	5	✓
593	9350	0.97	0.73	2	0.73	5	✓
594	7650	0.72	0.83	2	0.03	5	✓
595	5850	0.97	0.78	6	0.98	5	✓
596	9950	0.62	0.03	4	0.68	5	✓
597	12050	0.07	0.03	2	0.63	5	✓
598	13450	0.42	0.83	2	0.98	5	✓
599	10850	0.47	0.88	2	0.18	5	✓
600	10450	0.97	0.63	2	0.03	5	✓
601	13450	0.07	0.03	2	0.98	5	✓
602	10350	0.92	0.23	2	0.23	5	✓
603	8150	0.62	0.28	4	0.13	5	✓
604	11050	0.27	0.78	2	0.68	5	✓
605	13450	0.42	0.23	2	0.48	5	✓
606	13450	0.82	0.73	2	0.13	5	✓
607	6350	0.47	0.98	6	0.88	5	✓
608	13450	0.97	0.63	2	0.23	5	✓
609	10650	0.03	0.08	2	0.08	5	
610	7350	0.77	0.43	4	0.93	5	✓
611	9850	0.32	0.13	4	0.28	5	✓
612	9350	0.02	0.68	4	0.73	5	✓
613	7350	0.97	0.23	6	0.88	5	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
614	7850	0.32	0.98	4	0.33	5	✓
615	8050	0.82	0.13	4	0.93	5	✓
616	5950	0.77	0.98	6	0.03	5	
617	7550	0.52	0.98	4	0.13	5	✓
618	6850	0.82	0.63	4	0.98	5	✓
619	11350	0.82	0.18	2	0.13	5	✓
620	8550	0.97	0.03	6	0.58	5	
621	7750	0.47	0.98	2	0.98	5	✓
622	8750	0.92	0.18	4	0.33	5	✓
623	7350	0.97	0.33	4	0.88	5	✓
624	8850	0.67	0.58	2	0.38	5	✓
625	8150	0.07	0.53	6	0.38	5	✓
626	8850	0.62	0.38	4	0.63	5	✓
627	13450	0.67	0.23	2	0.93	5	✓
628	8950	0.87	0.03	6	0.08	5	
629	10550	0.02	0.18	4	0.83	5	✓
630	13450	0.72	0.03	2	0.38	5	✓
631	8450	0.47	0.63	4	0.78	5	✓
632	8550	0.77	0.58	2	0.68	5	✓
633	13250	0.97	0.28	2	0.13	5	✓
634	13250	0.97	0.98	2	0.18	5	✓
635	9650	0.47	0.23	4	0.53	5	✓
636	10750	0.52	0.33	2	0.23	5	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
637	13450	0.52	0.28	2	0.63	5	✓
638	12350	0.97	0.18	2	0.53	5	✓
639	7950	0.17	0.98	4	0.68	5	✓
640	13450	0.77	0.03	2	0.03	5	✓
641	9550	0.17	0.03	4	0.98	5	
642	10850	0.57	0.73	2	0.98	5	✓
643	8050	0.62	0.28	4	0.88	5	✓
644	8750	0.42	0.13	4	0.98	5	
645	7650	0.97	0.18	4	0.68	5	✓
646	12050	0.02	0.03	4	0.63	5	✓
647	6250	0.82	0.98	4	0.08	5	✓
648	8450	0.67	0.53	2	0.03	5	✓
649	11250	0.97	0.43	2	0.73	5	✓
650	5850	0.92	0.98	4	0.53	5	✓
651	11550	0.12	0.03	4	0.13	5	✓
652	6650	0.72	0.93	6	0.83	5	
653	6950	0.97	0.98	2	0.43	5	✓
655	8250	0.87	0.03	6	0.38	5	✓
656	7650	0.12	0.93	6	0.88	5	✓
657	9250	0.12	0.38	6	0.33	5	✓
658	13450	0.97	0.28	2	0.38	5	✓
659	7250	0.47	0.83	4	0.53	5	✓
660	8750	0.52	0.48	2	0.83	5	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
661	7050	0.47	0.73	6	0.03	5	✓
662	6650	0.92	0.83	6	0.53	5	
663	10050	0.52	0.88	2	0.03	5	✓
664	9050	0.37	0.78	2	0.73	5	✓
665	13450	0.97	0.78	2	0.08	5	✓
666	12050	0.67	0.03	2	0.38	5	✓
667	8150	0.82	0.58	2	0.03	5	✓
668	9350	0.02	0.23	4	0.58	5	
669	11350	0.52	0.38	2	0.03	5	✓
670	8450	0.92	0.58	2	0.73	5	✓
671	8950	0.27	0.63	4	0.43	5	✓
672	8850	0.22	0.58	2	0.58	5	
673	5950	0.72	0.98	6	0.53	5	✓
674	12050	0.02	0.03	4	0.83	5	✓
675	11350	0.22	0.88	2	0.18	5	✓
676	7850	0.02	0.98	6	0.38	5	
677	7250	0.57	0.53	6	0.73	5	✓
678	9350	0.32	0.03	4	0.73	5	
679	6450	0.32	0.98	6	0.88	5	
680	9150	0.97	0.03	4	0.48	5	✓
681	13450	0.97	0.33	2	0.73	5	✓
682	7150	0.12	0.98	4	0.58	5	
683	13450	0.97	0.23	2	0.53	5	✓

Table C.1 continued from previous page

No.	<i>fleetSize</i>	<i>Sharing</i>	<i>percentAV</i>	<i>taxiCapacity</i>	<i>dtMode</i>	DOE Step	In band
684	9950	0.37	0.23	4	0.43	5	✓
685	11750	0.47	0.03	2	0.28	5	✓
686	8150	0.02	0.63	4	0.98	5	
687	8950	0.57	0.03	4	0.63	5	✓
688	8050	0.87	0.43	4	0.88	5	✓
689	7650	0.32	0.98	2	0.43	5	✓
690	7450	0.77	0.53	6	0.23	5	✓

APPENDIX D

SCENARIOS FROM MIXTURE EXPERIMENTS

Table D.1.: Mixture design for case study in Section 5.3

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
6000	0.1	0.6	0.1	0.1	0.1	1
6000	0	1	0	0	0	1
6000	0	0	0.5	0.25	0.25	1
6000	0.25	0.25	0.5	0	0	1
6000	0	0	0.25	0.5	0.25	1
6000	0.25	0.25	0.25	0	0.25	1
6000	0	0.25	0.25	0	0.5	1
6000	0.25	0	0	0.25	0.5	1
6000	0.5	0.25	0.25	0	0	1
6000	0	0.25	0.25	0.25	0.25	1
6000	0	0	0	0.5	0.5	1
6000	1	0	0	0	0	1
6000	0.5	0	0.25	0.25	0	1
6000	0	0.5	0	0.5	0	1
6000	0.5	0	0	0.25	0.25	1
6000	0.5	0	0.5	0	0	1
6000	0.25	0	0.5	0	0.25	1

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
6000	0.75	0	0.25	0	0	1
6000	0	0	0	0.25	0.75	1
6000	0	0	0.25	0.75	0	1
6000	0.25	0	0.25	0	0.5	1
6000	0.25	0.5	0	0.25	0	1
6000	0.25	0.5	0	0	0.25	1
6000	0	0.75	0	0.25	0	1
6000	0	0	0.75	0.25	0	1
6000	0	0.25	0.5	0.25	0	1
6000	0	0	0.5	0	0.5	1
6000	0.1	0.1	0.6	0.1	0.1	1
6000	0	0	0	0	1	1
6000	0.25	0	0	0.75	0	1
6000	0.6	0.1	0.1	0.1	0.1	1
6000	0.25	0	0.75	0	0	1
6000	0.25	0.75	0	0	0	1
6000	0	0	1	0	0	1
6000	0.25	0.25	0.25	0.25	0	1
6000	0.5	0.25	0	0.25	0	1
6000	0	0.5	0	0.25	0.25	1
6000	0	0	0	0.75	0.25	1
6000	0	0.25	0.25	0.5	0	1
6000	0.75	0.25	0	0	0	1

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
6000	0	0.5	0.25	0.25	0	1
6000	0.25	0.25	0	0	0.5	1
6000	0.5	0.25	0	0	0.25	1
6000	0.25	0	0	0.5	0.25	1
6000	0.25	0	0.25	0.5	0	1
6000	0.5	0	0.25	0	0.25	1
6000	0.25	0.5	0.25	0	0	1
6000	0.5	0.5	0	0	0	1
6000	0.25	0.25	0	0.25	0.25	1
6000	0	0.25	0	0.25	0.5	1
6000	0	0.5	0.25	0	0.25	1
6000	0	0	0.25	0.25	0.5	1
6000	0.2	0.2	0.2	0.2	0.2	1
6000	0.75	0	0	0.25	0	1
6000	0.25	0	0	0	0.75	1
6000	0	0.25	0	0.75	0	1
6000	0.1	0.1	0.1	0.1	0.6	1
6000	0	0.25	0.5	0	0.25	1
6000	0.5	0	0	0	0.5	1
6000	0	0.75	0.25	0	0	1
6000	0	0.5	0.5	0	0	1
6000	0	0	0.75	0	0.25	1
6000	0	0.5	0	0	0.5	1

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
6000	0.1	0.1	0.1	0.6	0.1	1
6000	0	0	0	1	0	1
6000	0.75	0	0	0	0.25	1
6000	0	0	0.5	0.5	0	1
6000	0.5	0	0	0.5	0	1
6000	0	0.25	0	0.5	0.25	1
6000	0.25	0	0.5	0.25	0	1
6000	0	0.25	0	0	0.75	1
6000	0.25	0.25	0	0.5	0	1
6000	0	0	0.25	0	0.75	1
6000	0	0.75	0	0	0.25	1
6000	0	0.25	0.75	0	0	1
6000	0.25	0	0.25	0.25	0.25	1
6000	0	0.25	0.5	0.25	0	0
6000	0.1	0.6	0.1	0.1	0.1	0
6000	0.1	0.1	0.6	0.1	0.1	0
6000	0.2	0.2	0.2	0.2	0.2	0
6000	0.5	0	0.25	0	0.25	0
6000	0	0.5	0	0.5	0	0
6000	0.5	0	0	0.25	0.25	0
6000	0	0.25	0.25	0.5	0	0
6000	0	0	0.25	0.5	0.25	0
6000	0.25	0	0.25	0.25	0.25	0

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
6000	0	0.5	0.25	0.25	0	0
6000	0	0	0	0.25	0.75	0
6000	0.75	0	0	0	0.25	0
6000	0.25	0	0	0.5	0.25	0
6000	0	0.25	0	0.75	0	0
6000	0	0.75	0	0	0.25	0
6000	0	0.25	0	0	0.75	0
6000	0	0	0	0	1	0
6000	0	0.25	0.75	0	0	0
6000	0	0.25	0.25	0.25	0.25	0
6000	0.25	0	0.25	0	0.5	0
6000	0.5	0.25	0.25	0	0	0
6000	0.5	0	0.25	0.25	0	0
6000	0	0	0.5	0	0.5	0
6000	0	0	0.75	0	0.25	0
6000	0	0.75	0.25	0	0	0
6000	0	0.25	0.5	0	0.25	0
6000	0.25	0	0	0	0.75	0
6000	0.25	0.5	0	0	0.25	0
6000	0.25	0	0.5	0.25	0	0
6000	0.1	0.1	0.1	0.6	0.1	0
6000	0.25	0.75	0	0	0	0
6000	0	0	0	0.5	0.5	0

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
6000	0.5	0.25	0	0	0.25	0
6000	0	0.25	0	0.25	0.5	0
6000	0.25	0.25	0	0.25	0.25	0
6000	0	0	0.75	0.25	0	0
6000	1	0	0	0	0	0
6000	0	0.5	0	0	0.5	0
6000	0	0.25	0.25	0	0.5	0
6000	0.25	0.25	0	0.5	0	0
6000	0.5	0	0.5	0	0	0
6000	0.6	0.1	0.1	0.1	0.1	0
6000	0.5	0.5	0	0	0	0
6000	0	0.75	0	0.25	0	0
6000	0.75	0	0.25	0	0	0
6000	0.25	0.5	0.25	0	0	0
6000	0.5	0	0	0	0.5	0
6000	0	0	0.5	0.5	0	0
6000	0.25	0	0.75	0	0	0
6000	0	0.5	0	0.25	0.25	0
6000	0.25	0.25	0.25	0	0.25	0
6000	0	0	0.25	0	0.75	0
6000	0.25	0	0	0.25	0.5	0
6000	0	0	1	0	0	0
6000	0.25	0	0.5	0	0.25	0

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
6000	0.1	0.1	0.1	0.1	0.6	0
6000	0	0.25	0	0.5	0.25	0
6000	0.75	0	0	0.25	0	0
6000	0.25	0	0.25	0.5	0	0
6000	0	0	0	0.75	0.25	0
6000	0.5	0.25	0	0.25	0	0
6000	0.25	0.25	0.5	0	0	0
6000	0.5	0	0	0.5	0	0
6000	0	0.5	0.5	0	0	0
6000	0	0	0.25	0.25	0.5	0
6000	0	0	0	1	0	0
6000	0	0.5	0.25	0	0.25	0
6000	0	0	0.5	0.25	0.25	0
6000	0.75	0.25	0	0	0	0
6000	0	1	0	0	0	0
6000	0.25	0.25	0	0	0.5	0
6000	0.25	0.5	0	0.25	0	0
6000	0	0	0.25	0.75	0	0
6000	0.25	0	0	0.75	0	0
6000	0.25	0.25	0.25	0.25	0	0
4000	0.25	0.25	0	0	0.5	1
4000	0.25	0	0.25	0	0.5	1
4000	0.25	0.5	0	0.25	0	1

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
4000	0.5	0	0.25	0.25	0	1
4000	0.5	0	0.25	0	0.25	1
4000	0	0	0	0.75	0.25	1
4000	0.25	0.25	0.5	0	0	1
4000	0	0.5	0.25	0.25	0	1
4000	0.5	0.5	0	0	0	1
4000	0.25	0.25	0	0.25	0.25	1
4000	0.25	0.5	0.25	0	0	1
4000	0.5	0.25	0	0.25	0	1
4000	0	0.5	0	0.25	0.25	1
4000	0	0.25	0	0	0.75	1
4000	1	0	0	0	0	1
4000	0.75	0.25	0	0	0	1
4000	0.25	0	0	0.25	0.5	1
4000	0	0.25	0.25	0	0.5	1
4000	0.5	0	0	0.25	0.25	1
4000	0.75	0	0	0	0.25	1
4000	0	1	0	0	0	1
4000	0	0.25	0.5	0.25	0	1
4000	0.1	0.1	0.1	0.6	0.1	1
4000	0	0	0	0	1	1
4000	0.25	0.25	0	0.5	0	1
4000	0	0	0.25	0.25	0.5	1

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
4000	0	0.5	0	0	0.5	1
4000	0.6	0.1	0.1	0.1	0.1	1
4000	0.25	0	0.25	0.5	0	1
4000	0.5	0	0.5	0	0	1
4000	0	0.5	0.25	0	0.25	1
4000	0	0	1	0	0	1
4000	0.25	0.5	0	0	0.25	1
4000	0.2	0.2	0.2	0.2	0.2	1
4000	0	0	0.25	0.5	0.25	1
4000	0	0.25	0	0.75	0	1
4000	0.25	0	0.25	0.25	0.25	1
4000	0.75	0	0.25	0	0	1
4000	0	0.25	0.25	0.25	0.25	1
4000	0	0.5	0.5	0	0	1
4000	0.25	0	0	0.75	0	1
4000	0.25	0.25	0.25	0	0.25	1
4000	0	0	0	0.5	0.5	1
4000	0	0.75	0	0	0.25	1
4000	0.5	0.25	0.25	0	0	1
4000	0	0	0.25	0	0.75	1
4000	0	0.25	0	0.25	0.5	1
4000	0	0.75	0	0.25	0	1
4000	0	0	0	1	0	1

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
4000	0.25	0	0.5	0	0.25	1
4000	0.1	0.1	0.6	0.1	0.1	1
4000	0	0.75	0.25	0	0	1
4000	0	0.25	0.25	0.5	0	1
4000	0.5	0	0	0	0.5	1
4000	0	0	0.75	0.25	0	1
4000	0.25	0	0.75	0	0	1
4000	0	0	0.5	0	0.5	1
4000	0.25	0.75	0	0	0	1
4000	0	0	0.5	0.25	0.25	1
4000	0.5	0	0	0.5	0	1
4000	0	0	0.5	0.5	0	1
4000	0	0.5	0	0.5	0	1
4000	0	0	0	0.25	0.75	1
4000	0	0.25	0	0.5	0.25	1
4000	0.25	0.25	0.25	0.25	0	1
4000	0.25	0	0	0.5	0.25	1
4000	0.1	0.1	0.1	0.1	0.6	1
4000	0.25	0	0.5	0.25	0	1
4000	0	0	0.75	0	0.25	1
4000	0.75	0	0	0.25	0	1
4000	0.1	0.6	0.1	0.1	0.1	1
4000	0.5	0.25	0	0	0.25	1

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
4000	0	0.25	0.75	0	0	1
4000	0.25	0	0	0	0.75	1
4000	0	0.25	0.5	0	0.25	1
4000	0	0	0.25	0.75	0	1
4000	0	0	0.25	0.75	0	0
4000	0.1	0.1	0.6	0.1	0.1	0
4000	0.25	0.25	0	0.5	0	0
4000	0	0	0.5	0.25	0.25	0
4000	0.25	0	0.25	0.25	0.25	0
4000	0	0.5	0	0.25	0.25	0
4000	0	0.5	0	0	0.5	0
4000	0	0.25	0.25	0.25	0.25	0
4000	0.25	0.25	0	0	0.5	0
4000	0	0.25	0	0	0.75	0
4000	0	0	0.25	0.5	0.25	0
4000	0	0.25	0	0.75	0	0
4000	0.75	0	0.25	0	0	0
4000	0	0	0	1	0	0
4000	0.5	0.5	0	0	0	0
4000	0.1	0.6	0.1	0.1	0.1	0
4000	0.25	0.75	0	0	0	0
4000	0	0	0	0.5	0.5	0
4000	0.5	0	0.25	0.25	0	0

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
4000	0.2	0.2	0.2	0.2	0.2	0
4000	0	0	0.5	0.5	0	0
4000	0.25	0	0	0.25	0.5	0
4000	0	0	0.5	0	0.5	0
4000	0.5	0	0	0.5	0	0
4000	0.5	0	0.5	0	0	0
4000	0	0	0.25	0.25	0.5	0
4000	0	0.75	0	0.25	0	0
4000	0	0.25	0.5	0.25	0	0
4000	0.5	0.25	0	0.25	0	0
4000	0	0	0	0.75	0.25	0
4000	0	0.5	0.25	0.25	0	0
4000	0.25	0.25	0.25	0.25	0	0
4000	0.75	0	0	0.25	0	0
4000	0	0.5	0	0.5	0	0
4000	0	0.25	0	0.5	0.25	0
4000	0.25	0	0	0	0.75	0
4000	0.25	0	0.5	0.25	0	0
4000	0.1	0.1	0.1	0.1	0.6	0
4000	0	0	0.75	0	0.25	0
4000	0.6	0.1	0.1	0.1	0.1	0
4000	0.25	0.5	0.25	0	0	0
4000	1	0	0	0	0	0

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
4000	0	0	0.25	0	0.75	0
4000	0	0	0	0.25	0.75	0
4000	0	1	0	0	0	0
4000	0.25	0	0.5	0	0.25	0
4000	0.25	0	0.75	0	0	0
4000	0	0.25	0.75	0	0	0
4000	0	0	0	0	1	0
4000	0.75	0.25	0	0	0	0
4000	0	0	0.75	0.25	0	0
4000	0.25	0.5	0	0.25	0	0
4000	0.75	0	0	0	0.25	0
4000	0	0.5	0.25	0	0.25	0
4000	0.25	0.5	0	0	0.25	0
4000	0.1	0.1	0.1	0.6	0.1	0
4000	0.5	0	0.25	0	0.25	0
4000	0.25	0	0	0.5	0.25	0
4000	0	0.25	0.5	0	0.25	0
4000	0	0.5	0.5	0	0	0
4000	0.5	0.25	0	0	0.25	0
4000	0.25	0.25	0.5	0	0	0
4000	0.5	0	0	0	0.5	0
4000	0	0.25	0.25	0	0.5	0
4000	0	0.25	0.25	0.5	0	0

Table D.1 continued from previous page

fleetSize	riderTypes1	riderTypes2	riderTypes3	riderTypes4	riderTypes5	dtMode
4000	0.25	0	0.25	0	0.5	0
4000	0.25	0.25	0.25	0	0.25	0
4000	0	0.75	0.25	0	0	0
4000	0.25	0.25	0	0.25	0.25	0
4000	0	0	1	0	0	0
4000	0	0.25	0	0.25	0.5	0
4000	0.5	0	0	0.25	0.25	0
4000	0.25	0	0.25	0.5	0	0
4000	0.5	0.25	0.25	0	0	0
4000	0.25	0	0	0.75	0	0
4000	0	0.75	0	0	0.25	0

APPENDIX E
REGRESSION MODELS FOR MIXTURE
EXPERIMENTS

Table E.6.: Significant terms in the least squares regression

Term	Coef	SE Coef	T-Value	P-Value
x_1	0.8853	0.04964	17.833	¡ 2e-16
x_2	0.4634	0.04964	9.335	0.000000000000000287
x_3	0.1906	0.04964	3.84	0.000189
x_4	-0.08693	0.03384	-2.569	0.011298
x_5	0.8641	0.04964	17.407	¡ 2e-16
$x_1 : fleetSize$	-0.000002498	0.000009115	-0.274	0.784457
$x_2 : fleetSize$	0.00008297	0.000009115	9.103	0.000000000000000108
$x_3 : fleetSize$	0.0001236	0.000009115	13.558	¡ 2e-16
$x_4 : fleetSize$	0.0001586	0.000006526	24.297	¡ 2e-16
$x_5 : fleetSize$	0.00003731	0.000009115	4.093	0.0000732
$x_1 : dtMode$	0.1907	0.06573	2.902	0.004341
$x_2 : dtMode$	-0.3285	0.06573	-4.998	0.00000178
$x_3 : dtMode$	-0.04346	0.06573	-0.661	0.509649
$x_4 : dtMode$	0.1266	0.04706	2.689	0.008074

continued on next page

Table E.6.: *continued*

Term	Coef	SE Coef	T-Value	P-Value
$x_5 : dtMode$	-0.1573	0.06573	-2.393	0.01809
$x_1 : x_2$	-0.03211	0.05985	-0.536	0.592511
$x_1 : x_3$	0.008988	0.05985	0.15	0.880858
$x_1 : x_4$	0.06819	0.05172	1.318	0.18963
$x_1 : x_5$	-0.303	0.05985	-5.062	0.00000134
$x_2 : x_3$	-0.06213	0.05985	-1.038	0.301069
$x_2 : x_4$	-0.1225	0.05172	-2.368	0.019289
$x_2 : x_5$	-0.1841	0.05985	-3.075	0.002549
$x_3 : x_4$	-0.2617	0.05172	-5.059	0.00000136
$x_3 : x_5$	-0.1697	0.05985	-2.835	0.005294
$x_4 : x_5$	-0.4489	0.05172	-8.678	0.00000000000000119
$x_1 : fleetSize : dtMode$	-0.0000287	0.00001289	-2.226	0.027659
$x_2 : fleetSize : dtMode$	0.00005156	0.00001289	4	0.000104
$x_3 : fleetSize : dtMode$	0.000008136	0.00001289	0.631	0.529005
$x_4 : fleetSize : dtMode$	-0.00001759	0.00000923	-1.906	0.058853
$x_5 : fleetSize : dtMode$	0.00002446	0.00001289	1.897	0.059926

Table E.1.
Regression model for percent shared rides $RMSE = 0.055$ and $R^2 = 0.92$

term	β	Std. Error	t value	p value
x_1	9.66E-02	6.00E-02	1.611	$1.08E - 01$
x_2	6.07E-01	5.76E-02	10.542	$< 2e - 16$
x_3	1.13E-01	5.76E-02	1.954	0.051654
x_4	-1.29E-02	5.76E-02	-0.224	0.822868
x_5	-7.44E-02	5.91E-02	-1.258	0.209434
$dtMode$	8.75E-02	8.29E-03	10.555	$< 2e - 16$
$x_1 : x_5$	1.48E-01	8.07E-02	1.833	$6.78E - 02$
$x_2 : x_5$	5.23E-01	7.89E-02	6.634	$1.64E - 10$
$x_3 : x_5$	1.34E+00	7.89E-02	16.998	$< 2e - 16$
$x_4 : x_5$	1.30E+00	7.89E-02	16.52	$< 2e - 16$
$x_1 : x_4$	2.70E-01	7.89E-02	3.428	0.000699
$x_1 : x_3$	2.18E-01	7.89E-02	2.761	0.006145
$x_1 : x_2$	1.44E-01	7.89E-02	1.827	0.068678
$x_1 : fleetSize$	-2.76E-05	1.11E-05	-2.49	0.013347
$x_2 : fleetSize$	-3.36E-05	1.11E-05	-3.036	0.002616
$x_3 : fleetSize$	8.54E-05	1.11E-05	7.719	$2.00E - 13$
$x_4 : fleetSize$	1.17E-04	1.11E-05	10.581	$< 2e - 16$
$x_5 : fleetSize$	2.27E-05	1.11E-05	2.052	0.041062
$x_1 : dtMode$	-7.02E-02	2.65E-02	-2.647	0.008569

Table E.10.: Significant terms in the least squares regression

Term	Coef	SE Coef	T-Value	P-Value
x_1	435.929322	15.837079	27.526	$< 2e - 16$

continued on next page

Table E.10.: *continued*

Term	Coef	SE Coef	T-Value	P-Value
x_2	616.451413	15.837079	38.925	$< 2e - 16$
x_3	609.632324	10.796164	56.467	$< 2e - 16$
x_4	337.813236	15.837079	21.331	$< 2e - 16$
x_5	597.731992	15.837079	37.743	$< 2e - 16$
$x_1 : fleetSize$	-0.037512	0.002908	-12.899	$< 2e - 16$
$x_2 : fleetSize$	-0.070801	0.002908	-24.347	$< 2e - 16$
$x_3 : fleetSize$	-0.062999	0.002082	-30.257	$< 2e - 16$
$x_4 : fleetSize$	-0.022487	0.002908	-7.733	0.000000000000224
$x_5 : fleetSize$	-0.067013	0.002908	-23.044	$< 2e - 16$
$x_1 : dtMode$	35.606955	20.970065	1.698	0.091831
$x_2 : dtMode$	109.357914	20.970065	5.215	0.00000068
$x_3 : dtMode$	129.807213	15.014415	8.646	0.0000000000000143
$x_4 : dtMode$	115.618317	20.970065	5.513	0.000000175
$x_5 : dtMode$	171.997821	20.970065	8.202	0.000000000000017
$x_1 : x_2$	90.039926	19.09422	4.716	0.00000597
$x_1 : x_3$	65.446292	16.501312	3.966	0.000118
$x_1 : x_4$	-62.542581	19.09422	-3.275	0.001343
$x_1 : x_5$	340.904307	19.09422	17.854	$< 2e - 16$
$x_2 : x_3$	11.234398	16.501312	0.681	0.497161
$x_2 : x_4$	55.40913	19.09422	2.902	0.004338
$x_2 : x_5$	94.65239	19.09422	4.957	0.00000212
$x_3 : x_4$	35.926	16.501312	2.177	0.03122

continued on next page

Table E.10.: *continued*

Term	Coef	SE Coef	T-Value	P-Value
$x_3 : x_5$	18.573285	16.501312	1.126	0.262362
$x_4 : x_5$	130.140414	19.09422	6.816	0.00000000029
$x_1 : fleetSize : dtMode$	-0.00262	0.004113	-0.637	0.525231
$x_2 : fleetSize : dtMode$	-0.015205	0.004113	-3.697	0.000317
$x_3 : fleetSize : dtMode$	-0.019804	0.002945	-6.726	0.000000000462
$x_4 : fleetSize : dtMode$	-0.008156	0.004113	-1.983	0.049384
$x_5 : fleetSize : dtMode$	-0.021435	0.004113	-5.212	0.000000689

Table E.2.
 Regression model for average taxi capacity $RMSE = 0.095$ and $R^2 = 0.91$

term	β	Std. Error	t value	p value
x_1	1.36E+00	5.34E-02	25.502	$< 2e - 16$
x_2	2.98E+00	9.85E-02	30.248	$< 2e - 16$
x_3	2.33E+00	4.48E-02	52.022	$< 2e - 16$
x_4	2.33E+00	4.48E-02	51.846	$< 2e - 16$
x_5	1.94E+00	6.28E-02	30.924	$< 2e - 16$
$dtMode$	2.16E-01	1.43E-02	15.106	$< 2e - 16$
$fleetSize$	-3.75E-05	7.16E-06	-5.231	$3.28E - 07$
$x_1 : x_5$	-4.56E-01	1.56E-01	-2.922	0.003756
$x_2 : x_5$	5.29E-01	1.54E-01	3.444	0.00066
$x_3 : x_5$	1.45E+00	1.54E-01	9.42	$< 2e - 16$
$x_4 : x_5$	1.59E+00	1.54E-01	10.338	$< 2e - 16$
$x_1 : x_4$	6.41E-01	1.35E-01	4.757	$3.13E - 06$
$x_1 : x_3$	5.62E-01	1.35E-01	4.167	$4.10E - 05$
$x_1 : x_2$	3.82E-01	1.35E-01	2.837	0.004886
$x_2 : fleetSize$	-1.43E-04	2.27E-05	-6.266	$1.37E - 09$
$x_1 : dtMode$	-1.66E-01	4.55E-02	-3.657	0.000304

Table E.3.

Regression model for served fraction for rider type 1 $RMSE = 0.039$ and $R^2 = 0.96$

term	β	Std. Error	t value	p value
x_1	1.46E-01	5.20E-02	2.805	0.00571
x_2	-4.24E-01	7.14E-02	-5.936	$1.99E - 08$
x_3	-4.69E-01	7.14E-02	-6.568	$8.12E - 10$
x_4	-8.98E-01	7.14E-02	-12.574	$< 2e - 16$
x_5	8.16E-01	7.27E-02	11.228	$< 2e - 16$
$dtMode$	-1.57E-01	3.13E-02	-5.019	$1.47E - 06$
$dtMode : fleetSize$	2.82E-05	6.22E-06	4.539	$1.16E - 05$
$x_1 : fleetSize$	9.18E-05	1.02E-05	8.995	$1.05E - 15$
$x_2 : fleetSize$	2.32E-04	1.41E-05	16.407	$< 2e - 16$
$x_3 : fleetSize$	1.98E-04	1.39E-05	14.194	$< 2e - 16$
$x_4 : fleetSize$	2.62E-04	1.39E-05	18.804	$< 2e - 16$
$x_5 : fleetSize$	5.17E-05	1.39E-05	3.714	0.000289
$x_2 : x_5$	-3.18E-01	1.19E-01	-2.666	0.008538
$x_3 : x_5$	-3.31E-01	1.19E-01	-2.769	0.006334
$x_4 : x_5$	-8.01E-01	1.19E-01	-6.709	$3.90E - 10$
$x_2 : dtMode : fleetSize$	-2.35E-05	6.08E-06	-3.866	0.000165

Table E.4.

Regression model for Served fraction for rider type 2 $RMSE = 0.020$ and $R^2 = 0.96$

term	β	Std. Error	t value	p value
x_1	5.55E-01	3.79E-02	14.656	0.00571
x_2	1.91E-01	2.73E-02	7.013	$1.99E - 08$
x_3	7.92E-02	3.81E-02	2.077	$8.12E - 10$
x_4	3.32E-01	4.14E-02	8.014	$< 2e - 16$
x_5	6.46E-01	4.23E-02	15.265	$< 2e - 16$
$x_1 : dtMode$	1.64E-01	1.45E-02	11.346	$1.47E - 06$
$x_2 : dtMode$	4.81E-02	1.04E-02	4.643	$1.16E - 05$
$x_3 : dtMode$	1.06E-01	1.45E-02	7.336	$1.05E - 15$
$x_4 : dtMode$	7.98E-02	1.45E-02	5.521	$< 2e - 16$
$x_5 : dtMode$	2.03E-02	1.45E-02	1.406	$< 2e - 16$
$x_1 : fleetSize$	2.38E-05	7.23E-06	3.297	$< 2e - 16$
$x_2 : fleetSize$	1.02E-04	5.17E-06	19.653	0.000289
$x_3 : fleetSize$	1.11E-04	7.23E-06	15.298	0.008538
$x_4 : fleetSize$	7.97E-05	7.23E-06	11.029	0.006334
$x_5 : fleetSize$	6.72E-05	7.23E-06	9.297	$3.90E - 10$
$x_1 : x_5$	-4.38E-01	6.50E-02	-6.737	0.000165

Table E.5.

Regression model for served fraction for rider type 3 $RMSE = 0.016$ and $R^2 = 0.97$

term	β	Std. Error	t value	p value
x_1	0.7407	0.03236	22.888	$< 2e - 16$
x_2	0.3814	0.03236	11.786	$< 2e - 16$
x_3	0.0994	0.02118	4.694	0.00000621
x_4	0.5673	0.03236	17.529	$< 2e - 16$
x_5	0.5015	0.03236	15.497	$< 2e - 16$
$dtMode$	0.01431	0.002538	5.638	0.0000000888
$x_1 : fleetSize$	0.000008273	0.000005602	1.477	0.141968
$x_2 : fleetSize$	0.00008612	0.000005602	15.372	$< 2e - 16$
$x_3 : fleetSize$	0.000125	0.000004011	31.162	$< 2e - 16$
$x_4 : fleetSize$	0.00005772	0.000005602	10.303	$< 2e - 16$
$x_5 : fleetSize$	0.00006981	0.000005602	12.46	$< 2e - 16$
$x_1 : x_2$	0.1111	0.05202	2.136	0.034365
$x_1 : x_3$	0.08255	0.04496	1.836	0.06842
$x_1 : x_4$	0.3092	0.05202	5.943	0.0000000204
$x_1 : x_5$	-0.3018	0.05202	-5.802	0.0000000405
$x_2 : x_3$	-0.03064	0.04496	-0.682	0.49665
$x_2 : x_4$	-0.1189	0.05202	-2.286	0.023707
$x_2 : x_5$	-0.1162	0.05202	-2.234	0.027051
$x_3 : x_4$	-0.1565	0.04496	-3.481	0.000663
$x_3 : x_5$	-0.07512	0.04496	-1.671	0.096925
$x_4 : x_5$	-0.2916	0.05202	-5.605	0.000000104

Table E.7.

Regression model for served fraction for rider type 5 $RMSE = 0.015$ and $R^2 = 0.60$

term	β	Std. Error	t value	p value
x_1	0.02976	0.07977	0.373	0.7096
x_2	0.5131	0.07977	6.432	0.00000000144
x_3	0.5769	0.07977	7.232	0.0000000000196
x_4	0.6858	0.07977	8.597	0.0000000000000078
x_5	-0.1471	0.07095	-2.073	0.0398
<i>fleetSize</i>	0.00006797	0.00001183	5.747	0.0000000459
<i>dtMode</i>	0.04417	0.02365	1.867	0.0637

Table E.8.
 Regression model for waiting time for rider type 1 $RMSE = 13.4$ and $R^2 = 0.95$

term	β	Std. Error	t value	p value
x_1	368.675913	17.188815	21.449	$< 2e - 16$
x_2	542.972282	25.528429	21.269	$< 2e - 16$
x_3	483.382099	23.717387	20.381	$< 2e - 16$
x_4	412.129855	25.528429	16.144	$< 2e - 16$
x_5	298.754537	26.566712	11.245	$< 2e - 16$
$x_2 : fleetSize$	-0.06146	0.004603	-13.351	$< 2e - 16$
$x_3 : fleetSize$	-0.045669	0.004603	-9.921	$< 2e - 16$
$x_4 : fleetSize$	-0.032078	0.004603	-6.968	0.0000000000983
$x_5 : fleetSize$	-0.059095	0.004603	-12.837	$< 2e - 16$
$x_1 : fleetSize$	-0.0233	0.003296	-7.069	0.0000000000573
$x_1 : x_2$	71.748569	35.237472	2.036	0.04352
$x_1 : x_4$	97.957933	35.237472	2.78	0.00614
$x_1 : x_5$	102.962685	36.607174	2.813	0.00558
$x_2 : x_5$	359.871633	41.95543	8.577	0.00000000000012
$x_3 : x_5$	436.975141	41.338472	10.571	$< 2e - 16$
$x_4 : x_5$	543.36072	41.95543	12.951	$< 2e - 16$

Table E.9.

Regression model for waiting time for rider type 2 $RMSE = 24.17$ and $R^2 = 0.99$

term	β	Std. Error	t value	p value
x_1	579.5	51.2	11.318	$< 2e - 16$
x_2	800.4	32.91	24.319	$< 2e - 16$
x_3	758.1	44.67	16.969	$< 2e - 16$
x_4	504.9	44.67	11.302	$< 2e - 16$
x_5	588.6	55.69	10.569	$< 2e - 16$
$x_1 : dtMode$	167.7	25.18	6.658	0.000000000553
$x_2 : dtMode$	174.1	22.33	7.797	0.000000000012
$x_3 : dtMode$	177.2	25.18	7.036	0.000000000758
$x_4 : dtMode$	183.6	25.18	7.292	0.000000000192
$x_5 : dtMode$	218.2	25.18	8.665	0.000000000000881
$x_1 : fleetSize$	-0.018	0.009912	-1.816	0.07148
$x_2 : fleetSize$	-0.06303	0.006324	-9.967	$< 2e - 16$
$x_3 : fleetSize$	-0.05422	0.008556	-6.337	0.00000000287
$x_4 : fleetSize$	-0.01013	0.008556	-1.184	0.23849
$x_5 : fleetSize$	-0.1108	0.009912	-11.181	$< 2e - 16$
$dtMode : fleetSize$	-0.01827	0.003775	-4.84	0.00000332
$x_1 : x_5$	1720	370.7	4.638	0.00000786
$x_2 : x_5$	508.8	65.43	7.775	0.0000000000135
$x_3 : x_5$	817.8	74.84	10.927	$< 2e - 16$
$x_4 : x_5$	782.2	74.84	10.452	$< 2e - 16$
$x_1 : x_5 : fleetSize$	-0.2015	0.07262	-2.775	0.00626

Table E.11.
 Regression model for waiting time for rider type 4 $RMSE = 8.38$ and
 $R^2 = 0.99$

term	β	Std. Error	t value	p value
x_1	424.028391	16.894699	25.098	< 2e-16
x_2	885.065534	16.894699	52.387	< 2e-16
x_3	955.623022	21.256672	44.956	< 2e-16
x_4	358.990568	11.076043	32.411	< 2e-16
x_5	886.077558	16.894699	52.447	< 2e-16
$x_1 : fleetSize$	-0.029655	0.002884	-10.283	< 2e-16
$x_2 : fleetSize$	-0.110817	0.002884	-38.425	< 2e-16
$x_3 : fleetSize$	-0.105888	0.00387	-27.364	< 2e-16
$x_4 : fleetSize$	-0.015634	0.002065	-7.571	0.00000000000479
$x_5 : fleetSize$	-0.111885	0.002884	-38.795	< 2e-16
$x_1 : dtMode$	49.296442	5.767953	8.547	0.000000000000212
$x_2 : dtMode$	21.688875	5.767953	3.76	0.00025
$x_3 : dtMode$	118.154321	26.437289	4.469	0.0000162
$x_4 : dtMode$	60.977457	4.129812	14.765	< 2e-16
$x_5 : dtMode$	110.130221	5.767953	19.093	< 2e-16
$x_1 : x_2$	142.272316	26.779996	5.313	0.000000423
$x_1 : x_3$	23.759709	26.779996	0.887	0.3765
$x_1 : x_4$	4.308384	23.143395	0.186	0.85259
$x_1 : x_5$	450.825826	26.779996	16.834	< 2e-16
$x_2 : x_3$	-4.57097	26.779996	-0.171	0.86472
$x_2 : x_4$	52.318398	23.143395	2.261	0.02535
$x_2 : x_5$	82.353885	26.779996	3.075	0.00254
$x_3 : x_4$	-33.039489	23.143395	-1.428	0.15567
$x_3 : x_5$	-7.070741	26.779996	-0.264	0.79215
$x_4 : x_5$	140.152036	23.143395	6.056	0.000000125
$x_3 : fleetSize : dtMode$	-0.021259	0.00516	-4.12	0.000065

Table E.12.

Regression model for waiting time for rider type 5 $RMSE = 19.09$ and $R^2 = 0.90$

term	β	Std. Error	t value	p value
x_1	352.6	45.33	7.779	0.000000000000205
x_2	556.8	36.05	15.448	< 2e-16
x_3	838.1	38.78	21.609	< 2e-16
x_4	471.5	38.78	12.158	< 2e-16
x_5	469.9	32.26	14.565	< 2e-16
$x_1 : fleetSize$	-0.01589	0.008079	-1.967	0.051347
$x_2 : fleetSize$	-0.06104	0.006876	-8.877	0.00000000000000496
$x_3 : fleetSize$	-0.09667	0.006876	-14.059	< 2e-16
$x_4 : fleetSize$	-0.0436	0.006876	-6.341	0.00000000353
$x_5 : fleetSize$	-0.003335	0.005965	-0.559	0.577007
$x_1 : dtMode$	33.01	16.16	2.043	0.043092
$x_2 : dtMode$	24.67	13.75	1.793	0.075241
$x_3 : dtMode$	29.6	13.75	2.152	0.033264
$x_4 : dtMode$	88.41	13.75	6.428	0.00000000229
$x_5 : dtMode$	45.03	11.93	3.775	0.000243
$x_1 : x_5$	161.2	72.82	2.214	0.028569
$x_3 : x_4$	98.65	59.28	1.664	0.098487
$x_3 : x_5$	-524.4	58.7	-8.933	0.00000000000000363
$x_4 : x_5$	-241.6	58.7	-4.115	0.0000685

APPENDIX F

GENETIC ALGORITHM TO SITE EV CHARGING STATIONS

F.1 Genetic Algorithm - Details

In this Section, we expand on some of the details of the genetic algorithm which we discussed briefly in Section 6.3.3.1. The following sub-sections will explain the initial solution selection, the filtering procedure, and all of the genetic operators (selection, crossover, mutation, perturbation, fission, fusion, and correction) in detail.

F.1.1 Generate initial solutions (parents)

We use algorithm 10 to generate the set of initial parents. The algorithm generates multiple starting solutions as initial parents used in the the genetic algorithm. In order to explore the possible number of new locations, we generate parents for 1 new location till *reps* new locations. The algorithm does this *nRand* times to generate multiple starting solutions with the same number of new locations. This is done to explore starting points with a different number of sites (controlled by *reps*), different site locations and different number of ports (controlled by *nRand*). For each randomization (outer for loop controlled by *i*) and each number of new locations (inner for loop controlled by *j*) we cluster the demand into *j* clusters. Then, since the cluster centers are unlikely to be exactly at a charging station location, we

snap then randomly to a candidate charging station (we snap the cluster center randomly to increase the diversity in the locations of the charging stations between each iteration of the outer loop). The probability with which a particular cluster center is snapped to a particular charging station is proportional to the distance between the two. Then, we assign the number of ports to each charging station to generate vector N (chromosome). The probability of having more charging ports on a charging station is proportional to the number of demands associated with the cluster center to which that charging station has been snapped to. Figure F.1 represents this process pictorially.

Below we describe the variables that we have used in algorithm 10.

- $nRand = 50$: The number of randomizations.
- $reps = 30$: The maximum number of charging stations included in the initial population. Both, $nRand$ and $reps$ are tuning parameters and they control the size of the initial population ($nRand \times reps$).
- i : Variable controlling the outer for loop and the randomization number
- j : Variable controlling the inner for loop. j also determines the number of charging stations in the chromosome generated by a single iteration of the two loops.
- $c.old$: The number of charging stations currently existing in the system.
- $curBudget$: The budget remaining for use in the current generation.
- $centers$: Set of cluster centers. Its size will depend upon the value of j

- $freq$: a vector of the same size of $centers$. An element of $freq$ refers to the number of $demand$ that are associated with the corresponding element of $centers$.
- k : Variable controlling the innermost for loop. At each iteration it refers to an element in $centers$
- $dist$: A vector of distances from k to each potential site in \mathcal{L} .
- p : A probability vector that is proportional to $1/dist$. It is set such that $\sum p = 1$
- X : A set of charging station locations selected
- $Disc(v, p)$: The empirical discrete probability distribution where each element of v has probability p . This is used to randomly select the charging station which a cluster center in $centers$ is snapped to.
- c : the charging station that is associated with the center k
- n : A vector that generates the number of charging stations.
- C_{vec} : A vector of available charging station locations. Each selected charging station site c and existing charging station sites $c.old$ is repeated K times where K is the remaining capacity at the charging station location.
- K_c : The maximum number of vehicles a charging site c can hold. $K = \{K_1, \dots, K_{length(\mathcal{L})}\}$

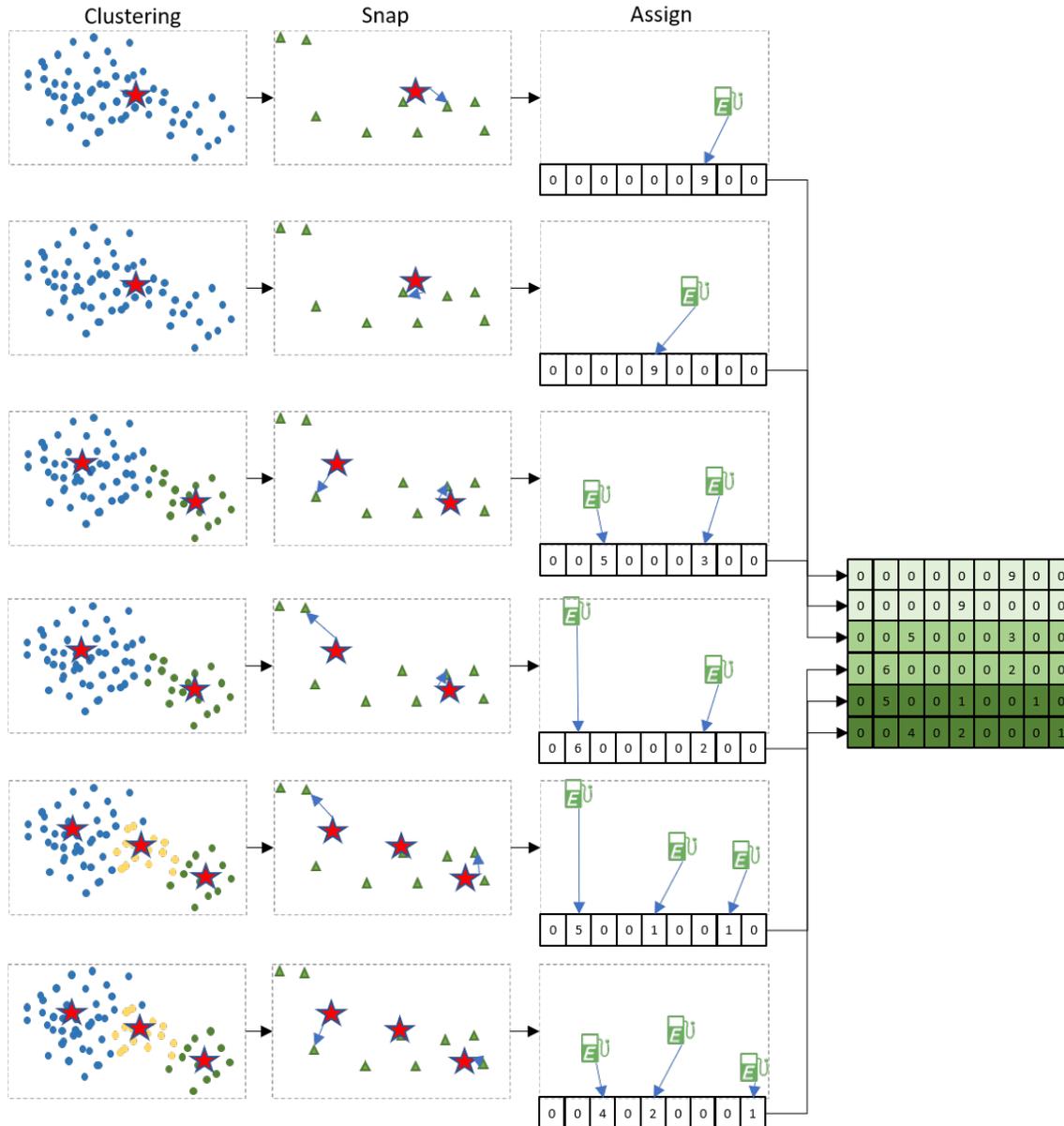


Figure F.1. Steps to generate initial population for $nRand = 2$ and $reps = 3$: **cluster** the demand (dots) into $j = 1, 2, 3$ centers (stars). The color represents the associated cluster center; **Snap** the cluster centers to potential charging station (triangle) locations with probability proportional to inverse of distance; **Assign** charging stations at the spots snapped in and generate the chromosome with all other potential charging station locations set to 0 and those selected set to the number of ports at the station randomly. The number of ports in the station are set randomly with probability proportional to the number of demands associated with that cluster.

Algorithm 10 Initial Parents

procedure INITIAL PARENTS(*riderGroup*)

 Get *nRand*, *reps*, and \mathcal{L} as input

for $i \in 1 \dots nRand$ **do**

 for $j \in 0 \dots reps$ **do**

 Set $c.old = \sum_{k \in \mathcal{L}} \mathbb{I}_{\hat{N}_j > 0}$
 $curBudget = B$

 Set X set of current charging stations

if $j > 0$ **then**

 Use k-means clustering to find j cluster centers (*centers*) and clustering vector counts *freq* for *demand* ▷ We use k-means to find locations where the demand would most likely benefit from additional charging station sites

 for $k \in centers$ **do** ▷ Snap clusters with probability in inverse proportion to distance

 find vector *dist* = distance from k to each potential charging location \mathcal{L} which has $K_c > freq$

 Compute probability vector $p \propto 1/dist$

 Select $c \sim Disc(\mathcal{L}, p)$ as the charging location and add to C (set of charging stations)

▷ Set the number of ports on each station randomly

 Generate vector n of length $\text{floor}(\frac{curBudget - C_n i}{C_u})$ by picking without replacement from vector C_{vec} of available charging station locations.

 Set vector of charging ports N from the frequency of n

F.1.2 Filter initial candidate set

Potentially in a big city such as New York City, there could be a large number of candidate charging stations. In such cases, in order to find potentially better solutions in shorter time, we can remove those candidate solutions that are unlikely to be chosen in the final solution. We define a few new variables that we will use in this section:

- $D = 5$: The number of divisions that we divide \mathcal{L}_{full} into.
 - \mathcal{L}_d : The $d \in 1 \dots D$ division for the divided set \mathcal{L}_{full} .
1. Divide the number of potential charging locations \mathcal{L}_{full} into D divisions randomly. We randomize the divisions so that each replication of the procedure that we run will have a different set of candidate charging stations.
 2. Run algorithm 10 for each of the D divisions to generate initial parents (chromosomes) from each of the \mathcal{L}_d subsets of charging stations.
 3. For each of the \mathcal{L}_d subsets of charging stations, evaluate the objective function using the method in section 6.3.4 and rank in decreasing order of objective function value
 4. In each set, from the best 100 chromosomes, select the genes (locations) which had a charging station and add to the reduced location set \mathcal{L} .

The reduced location set \mathcal{L} will contain the genes that are most likely to be included in the final solution.

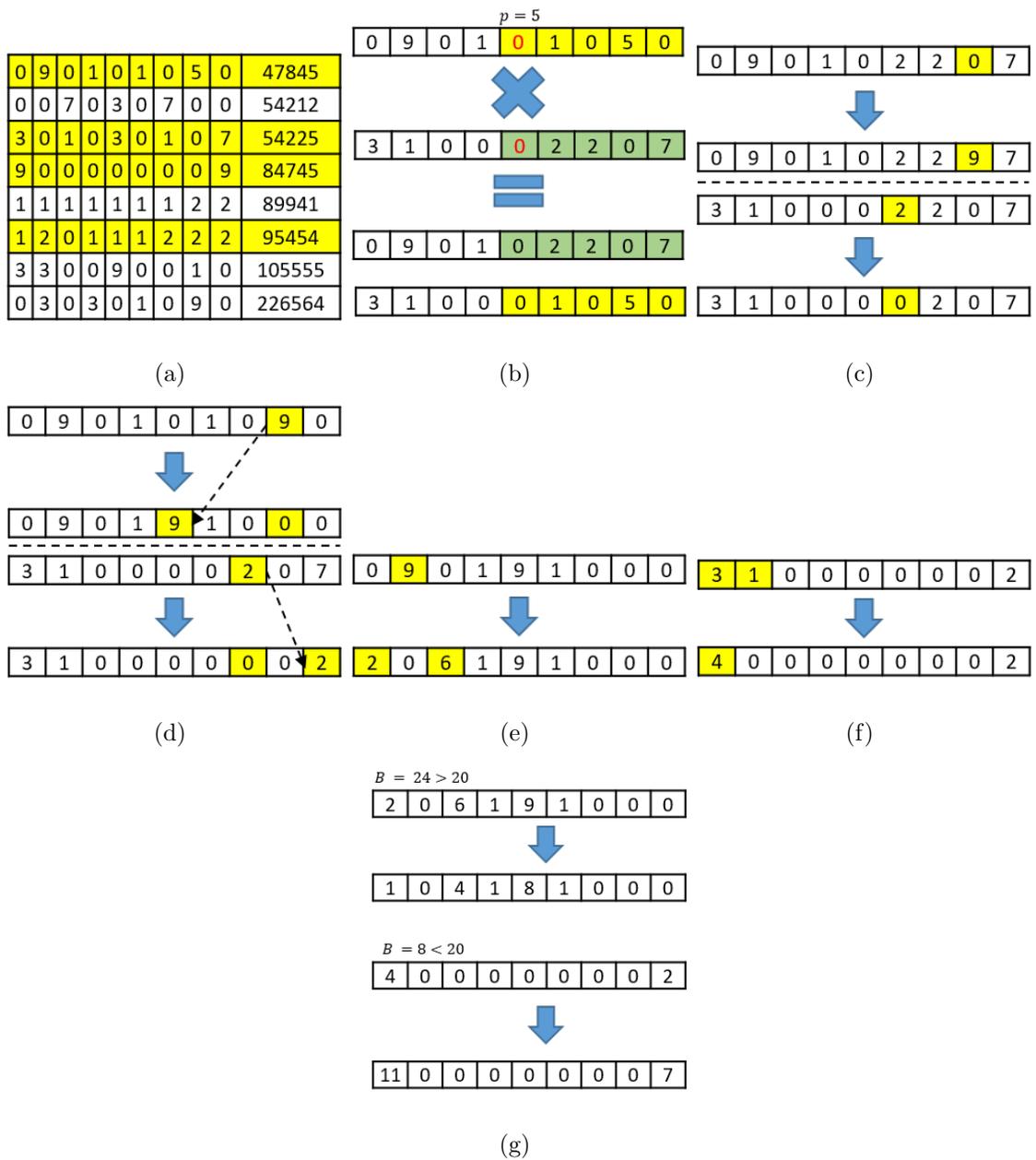


Figure F.2. Operations used in our genetic algorithm. The chromosomes here have a target budget of $B = 20$ for illustration: (a) Rank Selection: The $n_{best} = 4$ highlighted chromosomes are selected by the rank selection algorithm; (b) Crossover: The yellow genes from the first chromosome are interchanged with the green in the second; (c) Mutation: Two instances of mutation where the genes in yellow are switched; (d) Perturbation: The highlighted genes are randomly moved to a new location (e) Fission: The one highlighted gene is split into 2 (f) Fusion: The two highlighted genes are merged into one; (g) Correction: The non zero genes are edited to meet the budget requirements $B = 20$ while preserving the number of sites

F.1.3 Genetic operators

In this section we introduce the genetic operators rank selection, crossover, mutation, perturbation, fission, and fusion (figure F.2).

F.1.3.1 Rank Selection

The purpose of selection in a genetic algorithm is to select a good set of chromosomes that will be used as parents to produce off-springs in the genetic algorithm. We use rank selection as described in Sivanandam (2007) to perform selection from our chromosome set L . We introduce the variables and parameters that impact the selection of chromosomes below:

- L_1 : The set of chromosomes which will be used as parents to which genetic operations will be applied. The set L_1 will be transformed via genetic operations to create the next generation of chromosomes.
- $n_{best} = 20$: A parameter of the genetic algorithm which determines the number of parents that will be selected from set L
- $p_{choice} = 0.3$: A parameter of the genetic algorithm that determines how likely a chromosome is to be picked as a parent. A higher value means that the genetic algorithm will include a larger proportion of chromosomes which have higher objective function values (fitter) as parents and increase the probability of the algorithm finding a local minimum (exploitation).

The algorithm for the rank selection is below:

1. Rank the chromosomes L in increasing order of the objective function R .

2. Set $i = 1$
3. Repeat while $length(L_p) = n_{best}$ or $length(L) = i$
4. Add $L[i]$ to set L_p with probability p_{choice}
5. $i = i + 1$
6. Report L_p as the set of parents to perform genetic operators

F.1.3.2 Crossover

The purpose of the crossover operator is to generate new chromosomes from two parents. The intuition behind crossover is that a child composed of a cross between two good parents would share some of the qualities of both and hence may be fitter (lower wasted time) than both (Sivanandam, 2007). The crossover operator is shown in Figure F.2(b). We generate $N_{best} \times (N_{best} - 1)$ children using all of N_{best} parents by selecting 2 at a time.

Since our chromosomes are potentially very sparse (lots of potential charging stations with 0 ports), we choose a crossover gene $p_{crossover}$ between the first and last non zero gene. This increases the probability of generating an offspring that is different from the parent. Then, we generate the new children by interchanging all the genes that appear after $p_{crossover}$.

Specifically, the Algorithm 11 details the steps used for crossover.

F.1.3.3 Mutation

The purpose of the mutate operator is to have the algorithm generate chromosomes that differ in a single gene from the parents. This encourages the algorithm to

Algorithm 11 Crossover

```

1: procedure CROSSOVER(riderGroup)
2:   Get  $L_p$  as input from selection
3:    $k=1$ 
4:   for  $i \in 1 \dots \text{length}(L)$  do
5:     for  $j \in 1 \dots \text{length}(L)$  do
6:       if  $i \neq j$  then Select with uniform probability a point  $p_{crossover}$  between
         the first and last non zero genes in  $L[i]$  and  $L[j]$  Swap all genes after point
          $p_{crossover}$  to create two new genes  $L_1[k]$  and  $L_1[k + 1]$   $k = k + 1$ 

```

seek solutions that are different from the current set of accepted chromosomes (encourages exploration). The mutation operator has a single parameter $p_{mutate} = 0.2$ which controls the mutation probability. All the $N_{best} \times (N_{best} - 1)$ children go through the below mutation procedure:

1. With probability p_{mutate} , randomly select a non zero gene in the chromosome and set it to 0
2. With probability p_{mutate} , randomly select a zero gene in the chromosome and set it to $\min(\max(N), \text{capacity})$

F.1.3.4 Perturbation

Often in our problem the GA tends to find good solutions using the crossover and mutate operators. The purpose of the perturbation operator is to move the gene to another location randomly so that it can seek potentially better solutions

while making local moves. We use perturbation to move a non-zero gene by a small distance to test if a better location for a set number of charging stations exists.

The perturbation operator has two parameters:

All the $N_{best} \times (N_{best} - 1)$ children go through the below perturbation procedure:

- $p_{perturb} = 0.3$: The probability of perturbation
 - D : The distance (in number of genes) for perturbation.
 - G : The gene selected for perturbation
1. With probability $p_{perturb}$ select a non zero gene (G) randomly and displace it by a random distance between $[max(0, G - D), min(G + D, N)]$

F.1.3.5 Fission and Fusion

In our problem of charging station siting considering the budget restriction represented by Equation 6.2, one of the main trade-offs is to place 2 separate charging stations with less ports vs siting a single charging station with more ports. While the 2 charging stations would reduce the travel time if properly sited, a single charging station would be able to reduce the waiting time on account of it having an extra port. In order to explore these tradeoffs, we introduce the fission (Figure F.2(e)) and fusion (figure F.2(f)) operators.

The fission operator selects a single gene and splits it into two new genes, while the fusion operator selects two genes and joins them into a single gene. A chromosome is selected to either undergo fission or fusion (50% probability of each) .The algorithm for these operators is below:

1. With probability p_{ff}

2. With probability 0.5 (fusion), select two non zero genes from the chromosome (a and b) with inverse probability of distance from each other
3. Set gene $N_a = N_a + N_b + 1$ and $N_b = 0$
4. else Select one non-zero gene c with probability proportional to its total distance from other genes¹
5. Select two genes a and b such that $N_a = 0$ and $N_b = 0$ and set $N_a = N_b = \text{floor}(N_c/2)$ and $N_c = 0$

F.1.3.6 Correction

The budget constraint 6.2 is an equality constraint. After applying the genetic operators crossover, mutation, perturbation, fission, and fusion it is highly likely that the resulting chromosome would not satisfy constraint 6.2. Instead of discarding these chromosomes (as is the practise in traditional genetic algorithms (Sivanandam, 2007)) and reduce the efficiency of the genetic algorithm, we apply a correction step to force the chromosome N to adhere to the constraint 6.2. While performing correction we take care to preserve the number of locations as we noticed that in our case changing the number of locations results in a large change in objective function value. Additionally, all the genetic operators serve primarily to modify the number of new locations in the charging station configuration, and we wanted to preserve these changes while having the correction step modify the number of ports. We first compute the budget utilized (*curBudget*) by the chromosome. In-case the resulting chromosome is under-budget, we add charging ports to the non-zero genes.

¹We prefer genes that are isolated as it will be more likely that splitting an isolated gene would help reduce the travel distance for demands

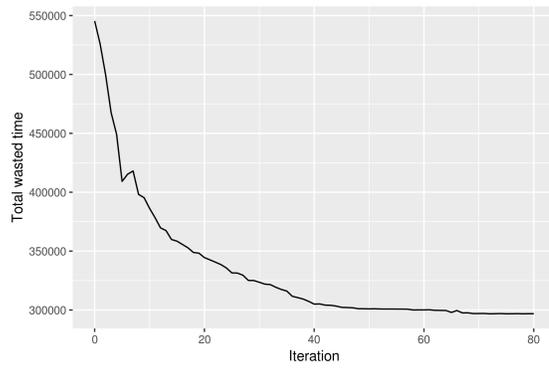
On the other hand if the resulting chromosome is over-budget, we remove charging ports from the non zero gene without decreasing the number of charging ports at any charging station to zero. In both cases, we attempt to preserve the proportion of ports across the charging stations, since the number of ports have already been randomized earlier by the other genetic operators. The algorithm for this correction step is below:

1. Compute $curBudget = C_n \sum_{i=1}^L \mathbb{I}_{N_i > 0} + C_u \sum_{i=1}^L (N_i - \mathbb{I}_{N_i > 0})$
2. If $curBudget < B$ (Underbudget)
3. Add $\text{floor} \frac{curBudget - B}{C_u}$ additional charging ports to genes randomly in the proportion of $K_c - N_c \quad \forall c \in n : N > 0$
4. else If $curBudget > B$ (Overbudget)
5. Remove $\text{floor} \frac{curBudget}{C_u}$ charging stations from the non zero genes with probability proportional to $N_c \quad \forall c \in n : N > 0$

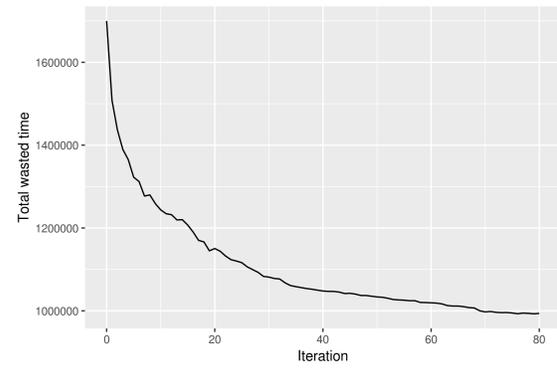
APPENDIX G

PROGRESSION OF OBJECTIVE FUNCTION IN GENETIC ALGORITHM

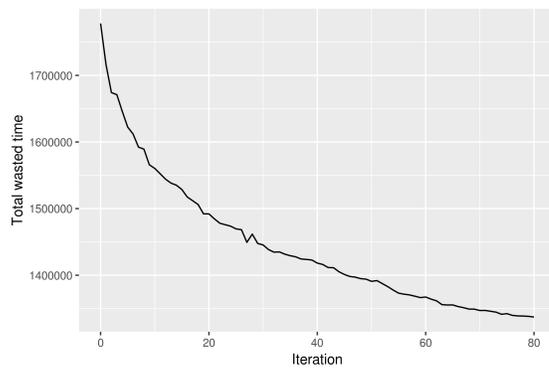
In each of the 3 cases mentioned in Section 6.4, we ran the genetic algorithm to site charging stations for $I = 80$ iterations as we believed that this number was large enough for the genetic algorithm to reach to a “good” solution. We kept the value of I fixed for all cases so that each case was comparable with each other. Additionally, since GA’s do not have any guarantees of decrease in objective function at each stage, we did not set our stopping criterion based on difference in objective function values. In each case, we plotted the progress of the best site selection at each iteration (Figure G.1, G.2 & G.3).



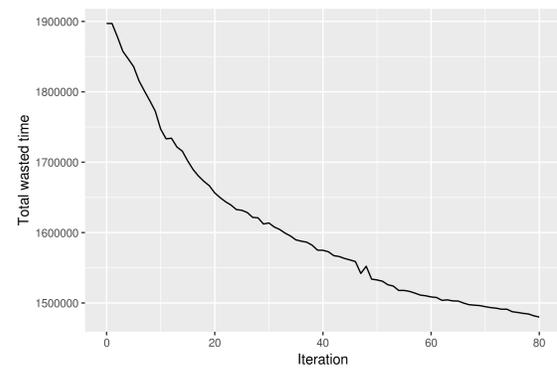
(a)



(b)

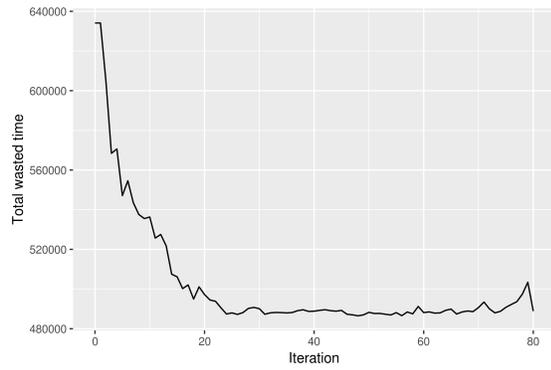


(c)

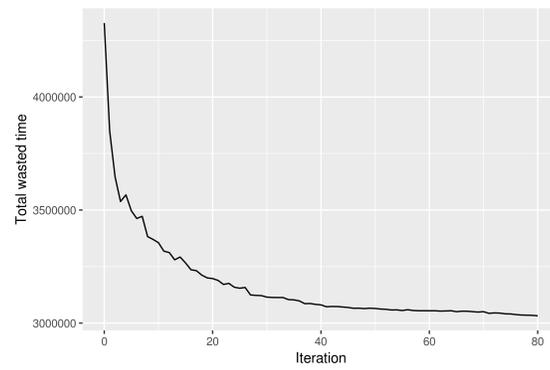


(d)

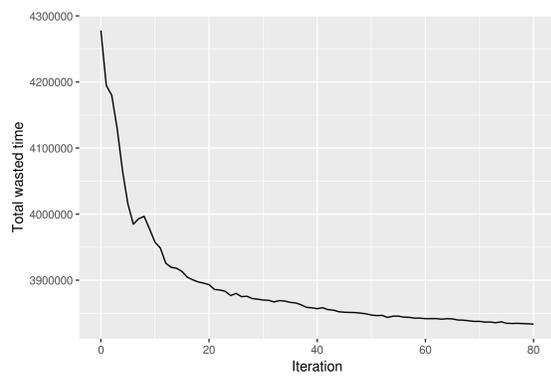
Figure G.1. Objective function progression for case F



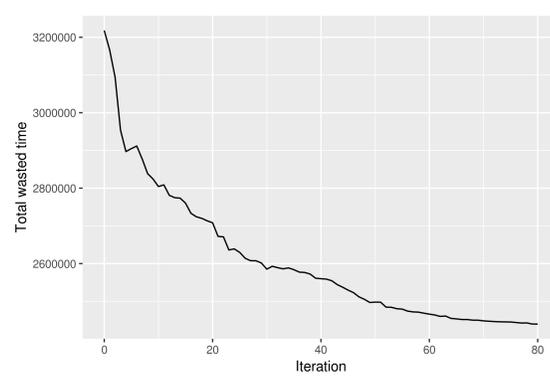
(a)



(b)



(c)



(d)

Figure G.2. Objective function progression for case P

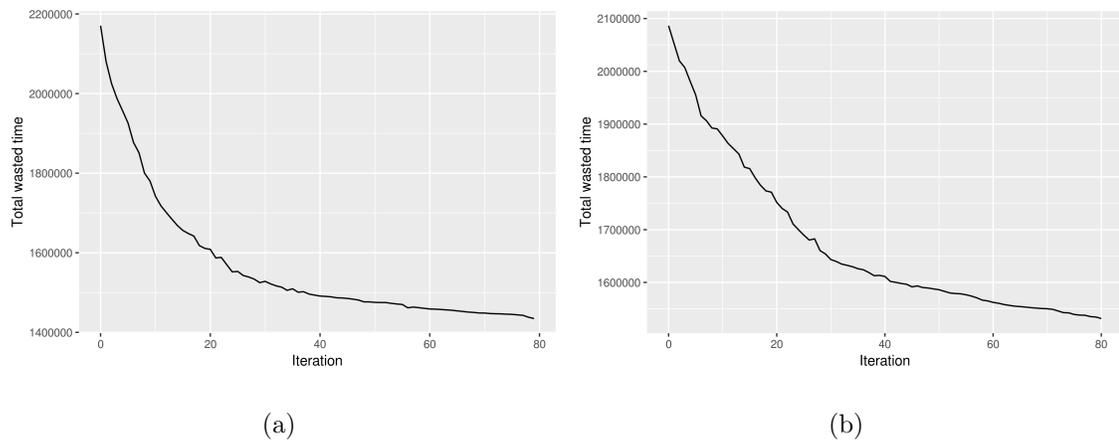


Figure G.3. Objective function progression for case M

APPENDIX H

SENSITIVITY ANALYSIS

H.1 Budget Sensitivity

In order to understand how the budget influences the objective function evaluation we evaluated the F and the P cases with different budgets as mentioned below:

- Case A1 : A10 ($B = 100$) \rightarrow A40 ($B = 300$) \rightarrow A70 ($B = 400$) \rightarrow A100 ($B = 800$)
- Case A2 : A10 ($B = 150$) \rightarrow A40 ($B = 350$) \rightarrow A70 ($B = 500$) \rightarrow A100 ($B = 1000$)
- Case A3 : A10 ($B = 200$) \rightarrow A40 ($B = 400$) \rightarrow A70 ($B = 600$) \rightarrow A100 ($B = 1200$)
- Case O1 : O10 ($B = 100$) \rightarrow O40 ($B = 300$) \rightarrow O70 ($B = 400$) \rightarrow O100 ($B = 800$)
- Case O2 : O10 ($B = 150$) \rightarrow O40 ($B = 350$) \rightarrow O70 ($B = 500$) \rightarrow O100 ($B = 1000$)
- Case O3 : O10 ($B = 200$) \rightarrow O40 ($B = 400$) \rightarrow O70 ($B = 600$) \rightarrow O100 ($B = 1200$)

Figure H.1 shows the objective function T (unscaled and scaled) along with its two components (waiting time T_w and travel time T_t). From Figure H.1 we see that

for the A scenarios, the additional budget is mainly used to reduce travel time by placing new charging stations, while in the O scenarios, the algorithm needs to find a balance between waiting time and travel time to lower the total objective function by co-locating charging stations and by placing new ones.

H.2 Sensitivity of C_N for $C_u = 1$

If we fix the cost of an additional port to be 1, the optimal charging station configuration could be influenced by varying values of C_N . For higher values of C_N , the model will prefer to add ports to existing charging stations, since adding an additional charging station is expensive, thus potentially increasing the total travel time. Also, as C_n increases, the total number of ports that the model can place is lower (because, by adding an additional charging station, the model uses a larger part of the budget). Hence it is important to understand how sensitive is the total wasted time to different values of C_N . So we found optimal charging station configurations for case F and Case P when $C_u = 1$ and $C_n = \{1.5, 2, 3\}$ in Figure H.2.

We see from Figure H.2 that in all the cases, the total wasted time is higher for larger values of C_N , by a small amount (either due to a higher waiting time or a higher travel time)

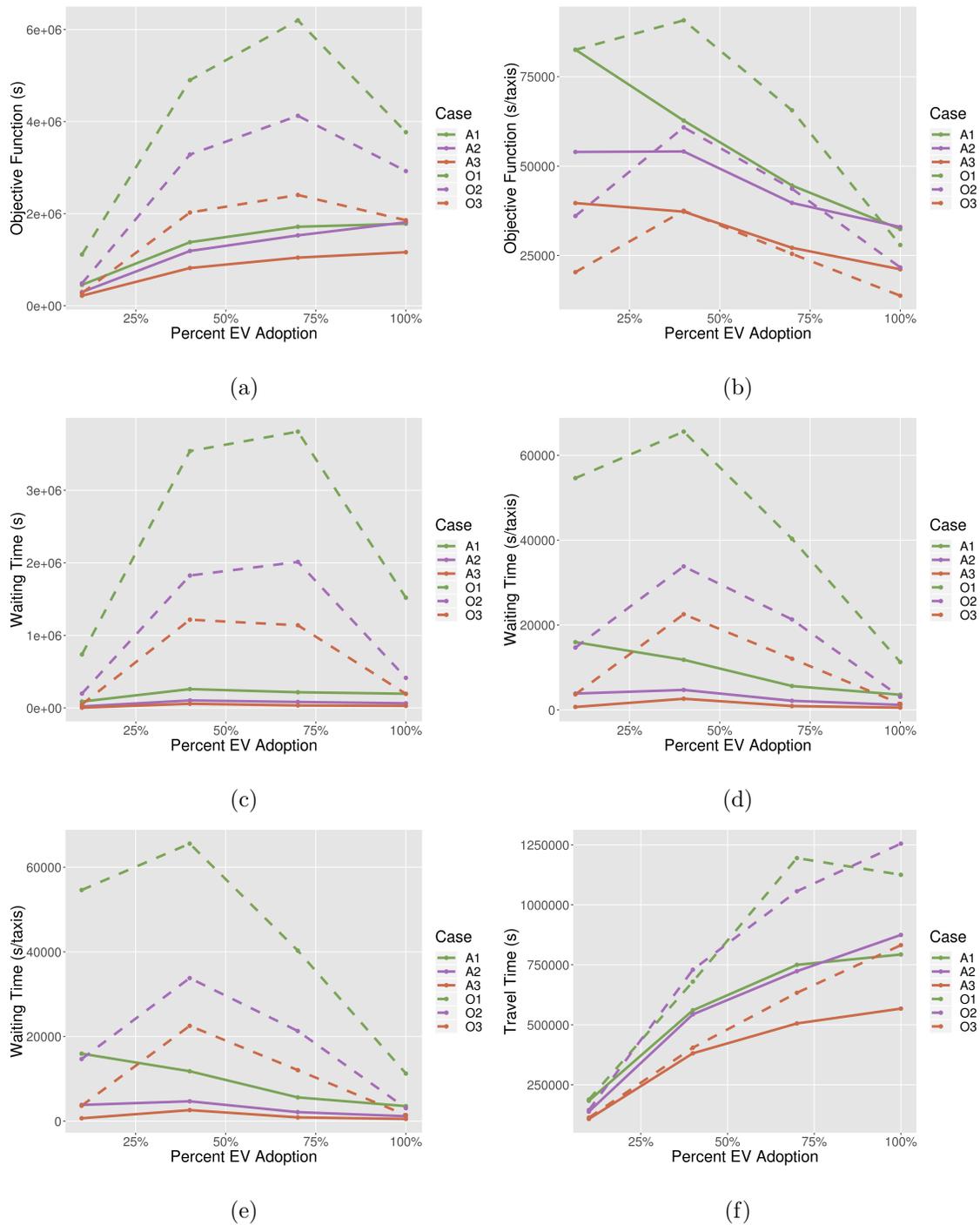


Figure H.1. Sensitivity of the (a) objective function (b) scaled objective function (c) waiting time (d) scaled waiting time (e) traveling time (f) scaled travel time for cases which defer by budget as per Section H.1 (color and line-type)

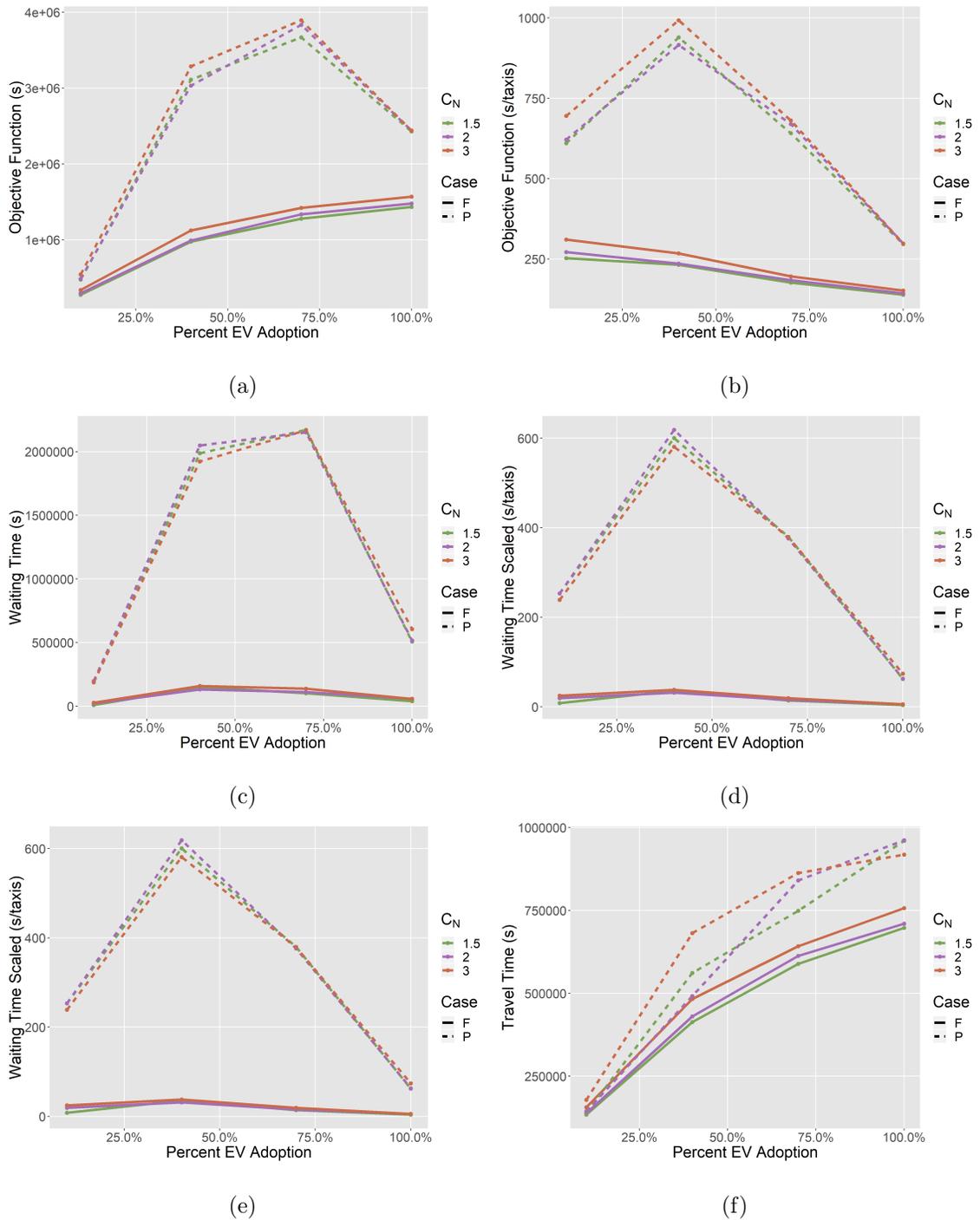


Figure H.2. Sensitivity of the (a) objective function (b) scaled objective function (c) waiting time (d) scaled waiting time (e) traveling time (f) scaled travel time for cases which defer by C_N as per Section H.2 (color and line-type)