

MIT Open Access Articles

Robust matching-integrated vehicle rebalancing in ride-hailing system with uncertain demand

The MIT Faculty has made this article openly available. **Please share**
how this access benefits you. Your story matters.

Citation: Guo, Xiaotong, Caros, Nicholas S and Zhao, Jinhua. 2021. "Robust matching-integrated vehicle rebalancing in ride-hailing system with uncertain demand." *Transportation Research Part B: Methodological*, 150.

As Published: 10.1016/J.TRB.2021.05.015

Publisher: Elsevier BV

Persistent URL: <https://hdl.handle.net/1721.1/140212>

Version: Original manuscript: author's manuscript prior to formal peer review

Terms of use: Creative Commons Attribution-NonCommercial-NoDerivs License



Robust Matching-Integrated Vehicle Rebalancing in Ride-Hailing System with Uncertain Demand

Xiaotong Guo^a, Nicholas S. Caros^a, Jinhua Zhao^{b,*}

^a*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

^b*Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Abstract

With the rapid growth of the mobility-on-demand (MoD) market in recent years, ride-hailing companies have become an important element of the urban mobility system. There are two critical components in the operations of ride-hailing companies: driver-customer matching and vehicle rebalancing. In most previous literature, each component is considered separately, and performances of vehicle rebalancing models rely on the accuracy of future demand predictions. To better immunize rebalancing decisions against demand uncertainty, a novel approach, the matching-integrated vehicle rebalancing (MIVR) model, is proposed in this paper to incorporate driver-customer matching into vehicle rebalancing problems to produce better rebalancing strategies. The MIVR model treats the driver-customer matching component at an aggregate level and minimizes a generalized cost including the total vehicle miles traveled (VMT) and the number of unsatisfied requests. For further protection against uncertainty, robust optimization (RO) techniques are introduced to construct a robust version of the MIVR model. Problem-specific uncertainty sets are designed for the robust MIVR model. The proposed MIVR model is tested against two benchmark vehicle rebalancing models using real ride-hailing demand and travel time data from New York City (NYC). The MIVR model is shown to have better performances by reducing customer wait times compared to benchmark models under most scenarios. In addition, the robust MIVR model produces better solutions by planning for demand uncertainty compared to the non-robust (nominal) MIVR model.

Keywords: Ride-hailing, Vehicle Rebalancing, Robust Optimization, Demand Uncertainty.

1. Introduction

Advanced wireless communication and cloud computing technologies coupled with the growing popularity of shared mobility have led to a fast-growing Mobility-on-Demand (MoD) market in recent years [1]. Ride-hailing companies, also known as Transportation Network Companies (TNCs), such as Uber and Lyft have become ubiquitous forms of MoD in most

*Corresponding author

Email addresses: xtguo@mit.edu (Xiaotong Guo), caros@mit.edu (Nicholas S. Caros), jinhua@mit.edu (Jinhua Zhao)

6 cities over the past decade. The number of worldwide active drivers for Uber grew from
7 almost zero in 2010 to over 3 million in 2017, while Lyft, a relative latecomer to the market,
8 had 1.4 million active drivers in the US and Toronto in 2017 [2]. Two of the primary
9 innovations that allowed them to capture a significant market share from their established
10 competitors, the taxi industry, were: 1) matching trip requests with drivers using a mobile
11 app rather than curbside hailing or an in-advance booking system, and 2) responding to
12 changes in demand by incentivizing or actively dispatching drivers to high-demand areas.
13 These innovations have been identified as two important ride-hailing operations problems in
14 the literature: the driver-customer matching problem and the vehicle rebalancing problem [3].

15 One of the key technological competence requirements for efficient operation of ride-
16 hailing platforms is the algorithmic approaches for optimally matching drivers and cus-
17 tomers in real-time [4]. Given a list of available vehicles and trips requested by customers,
18 the matching algorithm pairs drivers and customers according to specific objectives and fea-
19 sibility constraints. Moreover, matching decisions need to be made quickly, typically within
20 seconds. Researchers have been seeking solutions to improve the operational and computa-
21 tional performance of the on-demand driver-customer matching problem.

22 Because the spatial distributions of supply and demand in the ride-hailing system are
23 often unbalanced, platforms can improve the operational performance by actively rebalancing
24 idle vehicles to areas where the demand is expected to exceed supply based on estimates of
25 future demand. Algorithms for rebalancing idle vehicles have been proposed for ride-hailing
26 platforms to reduce wait times for customers [5, 6, 7, 8]. However, the performance of vehicle
27 rebalancing algorithms depends on the accurate future demand estimations. Rebalancing
28 decisions generated with inaccurate demand forecasts could have negative impacts on the
29 system performance. Incorporating robustness into the vehicle rebalancing algorithm is
30 one approach to protect solutions against demand uncertainty that arise from inaccurate
31 estimates of future demand [7].

32 While rebalancing and matching are often treated as separate operations in the litera-
33 ture [3], both problems relate to dispatching idle vehicles, either to pick up customers or to
34 increase supply in areas with high expected demand. A common objective for the driver-
35 customer matching problem is minimizing the vehicle miles traveled (VMT) and unsatisfied
36 requests [9, 10] while the primary objective for the vehicle rebalancing problem is minimizing
37 the VMT and a functional term measuring the system-wide service availability for future
38 demand [5, 6, 7].

39 In the vehicle rebalancing problem, the overall goal for improving the system-wide service
40 availability for incoming customers is to minimize the number of unsatisfied requests, which
41 coincides with the objective of the driver-customer matching problem. The functional term in
42 the objective of the vehicle rebalancing problem can therefore be treated as an approximation
43 to represent the number of unsatisfied requests. However, the maximum system-wide service
44 availability for incoming customers does not necessarily lead to the minimum number of
45 unsatisfied requests if there are inaccurate future demand estimates. To immunize vehicle
46 rebalancing decisions against the inherent demand uncertainty, we introduce the driver-
47 customer matching component into the vehicle rebalancing problem in order to explicitly
48 model the number of unsatisfied requests.

49 Nonetheless, there is a methodological difference between driver-customer matching prob-
50 lems and vehicle rebalancing problems. The driver-customer matching problem is typically

51 solved by an agent-based model, where each driver and customer are considered individually.
 52 For the vehicle rebalancing problem, most methods divide the study area into several sub-
 53 regions and the vehicle rebalancing problem is solved at an aggregate level, where vehicles
 54 are rebalanced between sub-regions.

55 To resolve this methodological difference, we propose the matching-integrated vehicle
 56 rebalancing (MIVR) model where the area partitioning method is retained and the matching
 57 component is modeled at an aggregate level. The objective of the MIVR model is to minimize
 58 the total VMT and the number of unsatisfied requests. The aggregate matching component
 59 of the MIVR model provides a satisfying approximation of the vehicle pickup distance and
 60 the number of unsatisfied requests when using small regions.

61 Figure 1 provides a toy example to illustrate the benefits of the MIVR model compared to
 62 an independent vehicle rebalancing (VR) model, where the service availability is represented
 63 by the absolute difference between estimated future demand and supply. Compared to the
 64 independent rebalancing scenario, the matching-integrated rebalancing scenario dispatches
 65 the idle vehicle to a location near sub-regions with estimated future demand. This “smart”
 66 rebalancing decision compensates for inaccurate future demand estimation by harmonizing
 67 vehicle pickup distance across different demand profiles.

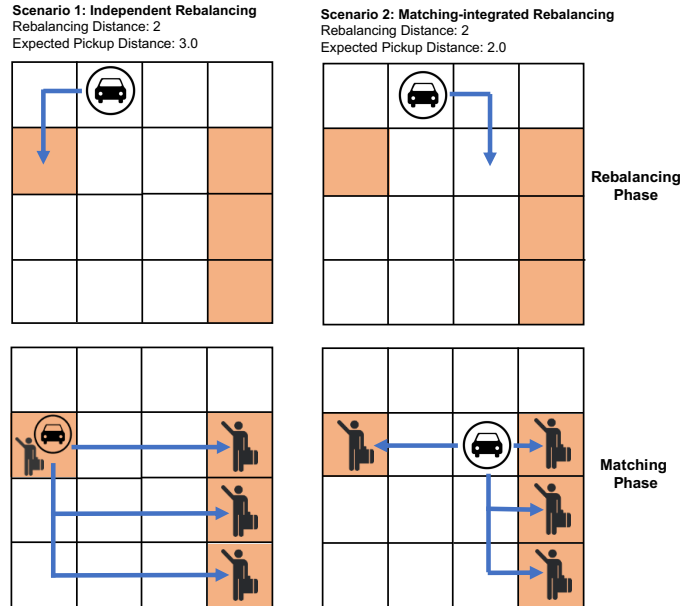


Figure 1: Example scenarios comparing regular VR decisions and the MIVR decisions. There are 16 unit squares (sub-regions) and a trip request is equally likely to appear in any of the orange sub-regions in the next time interval. For the independent rebalancing scenario, the rebalancing distance is 2 and the expected pick-up distance is 3 (four possible pick-up distances 0, 3, 4, 5 with $\frac{1}{4}$ probability on each case). For the matching-integrated rebalancing scenario, the rebalancing distance is 2 and the expected pick-up distance is 2.0 (four possible pick-up distances 1, 2, 2, 3 with $\frac{1}{4}$ probability on each case).

68 To further protect the vehicle rebalancing decisions against demand uncertainty, we in-
 69 troduce robust optimization (RO) techniques to construct a robust MIVR model. Problem-
 70 specific uncertainty sets are established to better reflect the uncertainty within ride-hailing
 71 demand.

72 In short, the ride-hailing matching process and RO techniques can be incorporated into
73 the rebalancing procedure to produce better vehicle rebalancing decisions for platforms when
74 facing demand uncertainty. The contributions of this paper can be summarized as follows:

- 75 • Proposing the MIVR model to incorporate driver-customer matching information to
76 improve vehicle rebalancing problems with explicit modeling of unsatisfied requests for
77 the first time, to the best of authors' knowledge.
- 78 • Proposing the robust MIVR model to consider demand uncertainty and designing
79 problem-specific uncertainty sets to better reflect the inherent demand uncertainty in
80 the ride-hailing system.
- 81 • Using simulations to show performance improvements of the MIVR model compared
82 to an independent VR model and a state-of-the-art empty-car routing policy with real
83 demand data and travel times from New York City (NYC). In high supply scenarios,
84 a *Pareto* improvement can be found for the MIVR model when compared to the VR
85 model at aggregate level regarding the overall VMT, the average customer wait time
86 and the number of unsatisfied requests.
- 87 • Comparing the nominal MIVR and the robust MIVR under multiple uncertain scenar-
88 ios by solving a driver-customer matching problem with realized demand and vehicle
89 distributions after rebalancing. The robust MIVR model is shown to perform better
90 under demand uncertainty, especially in conditions of high supply relative to demand.

91 The remainder of the paper is organized as follows. Section 2 reviews the relevant litera-
92 ture. Section 3 describes the nominal and robust MIVR models and the robust counterpart.
93 Section 4 includes the empirical study design and descriptions for data used in this paper.
94 Benchmark comparisons, scenario testing results and robust solution performances are de-
95 scribed in Section 5. Finally, Section 6 recaps the main contributions of this work, outlines
96 the limitations and provides future research directions.

97 **2. Existing Literature**

98 *2.1. Ride-hailing Matching and Rebalancing*

99 Ride-hailing matching is a variant of the classical Dial a Ride Problem, where customer
100 trips are matched with vehicles such that generalized costs are minimized. These costs can
101 include VMT, customer wait time, and penalties for poor service quality. Development of
102 new algorithms for this problem is a very active field of research and the methods have been
103 used by platform operators in practice [11]. Agatz et al. [4] provided a comprehensive survey
104 of literature related to optimization of driver-to-passenger for dynamic ride sharing between
105 travelers with similar itineraries. In a more recent survey, Mourad et al. [12] reviews research
106 related to optimization of shared mobility systems more broadly, which includes ride-hailing.
107 The authors identify demand uncertainty as a critical issue in modeling shared mobility sys-
108 tems, and identify stochastic programming and multi-scenario optimization as two possible
109 modeling techniques. Finally, Ho et al. [13] presents an overview of recent research relating to

110 the general Dial a Ride Problem. While this survey is focused on applications such as para-
111 transit and demand-responsive transit, the taxonomy and solution techniques are applicable
112 to ride-hailing problems. Like Mourad et al. [12], the authors find that the development
113 of models and solution methods that include stochastic demand is an important research
114 direction.

115 Ride-hailing is one type of on-demand service platform, which is characterized by the
116 waiting time sensitivity of customers and service providers without fixed work schedules.
117 Other on-demand service platforms include food and goods delivery services such as Door-
118 Dash and Uber Eats, and ride-pooling platforms. Several recent papers have examined the
119 dynamics of on-demand service platforms. It has been shown that customers' sensitivity
120 to delay has a significant impact on optimal pricing and wage setting [14]. Another paper
121 determines optimal prices and wages under different levels of demand, and calibrates param-
122 eters using actual ride-hailing data [15]. Cachon et al. [16] develops a model for dynamic
123 pricing in on-demand platforms, demonstrating that such policies benefit stakeholders by
124 expanding access to service during periods of peak demand. Theoretical relationships be-
125 tween pricing, demand and detour policies within ride-pooling platforms, which are similar
126 to ride-hailing but with the possibility of shared trips, have also been investigated [17].

127 Given the size and dynamic nature of the ride-hailing matching problem in large cities,
128 many approaches involve metaheuristic methods to generate sub-optimal solutions [18, 19].
129 Recently, researchers have investigated the role of matching radii and matching time periods
130 on the optimal solution [20]. Lyu et al. [21] develops an online matching algorithm that
131 considers multiple objectives, and provides a theoretical optimality guarantee for the online
132 solution. Xu et al. [11] proposed a dynamic programming approach to matching that seeks to
133 optimize matching decisions over a long time horizon. Their method, which did not consider
134 demand uncertainty, has been adopted by a leading ride-hailing platform.

135 Optimal rebalancing of idle ride-hailing vehicles has shown to substantially improve sys-
136 tem performance. Typical considerations in designing a rebalancing algorithm are the dura-
137 tion of the decision period and the costs included in the objective function. Chen and Levin
138 [22] proposed a simple linear programming (LP) model to select vehicle rebalancing flows
139 that minimize travel cost for five minute periods. Zhang et al. [23] showed that a stable
140 predictive control algorithm could be used for dispatching and rebalancing an autonomous
141 ride-hailing fleet in a discrete time system. At each decision period, a mixed-integer linear
142 programming (MILP) is solved to minimize rebalancing travel time. Their method pro-
143 duced significant reductions in peak wait times compared to the no rebalancing scenario.
144 Similarly, Iglesias et al. [24] proposed a model predictive control algorithm for operating the
145 ride-hailing system in real-time by leveraging short-term demand forecasts. They utilized
146 the Long Short-Term Memory (LSTM) neural networks to forecast future customer demand
147 for each origin and destination pair and their proposed algorithm outperformed a state-of-
148 the-art rebalancing strategy by reducing up to 89.6% of the average customer wait time.
149 Wallar et al. [5] developed an online vehicle rebalancing algorithm that discretized an area
150 into optimal rebalancing sub-regions, resulting in an average wait time reduction of 37%
151 compared to the scenario without rebalancing idle vehicles. Braverman et al. [25] formulates
152 a fluid-based optimization model for idle vehicle rebalancing in ride-hailing systems. The
153 authors use a nine-region network and real-life ride-hailing data to show how the fluid-based
154 model results in a higher fraction of passengers served compared to benchmark models. We

155 include the Braverman et al. [25] model as a benchmark to test the results of our own model.
156 Al-Kanj et al. [26] combined the matching and vehicle rebalancing into a single dynamic
157 programming method for autonomous electric vehicles. Their approach employs incentives
158 rather than centralized control to rebalance vehicles, meaning that rebalancing decisions
159 made by the platform are subject to some amount of non-responsiveness by the passenger
160 or vehicle. Dandl et al. [27] also solves for matching and rebalancing decisions in a single
161 optimization model to inform a simulation that tests how demand forecasting accuracy affects
162 the system performance. The authors use an agent-based model, where the objective function
163 is a combination of penalties and rewards for matching and for reducing demand-supply
164 imbalances. Their simulation assumes that all requests are served and customers will wait
165 indefinitely for pickup. In contrast, our method includes matching information and explicitly
166 models customer wait time and unsatisfied requests in order to make rebalancing decisions.

167 In addition to optimization methods, machine learning (ML) approaches have been pro-
168 posed to predict demand in rebalancing vehicles [6, 28]. There has also been considerable
169 work on other practical methods, beyond explicit vehicle rebalancing, to achieve greater bal-
170 ance between supply and demand in ride-hailing systems. These methods include dynamic
171 pricing [16, 29], providing more information to drivers [30], reward schemes [31], alternative
172 market structures [32] and carpooling incentives [33].

173 *2.2. Robust Optimization*

174 RO is a common approach to handle data uncertainty in optimization problems. The
175 general approach is to specify a range for an uncertain parameter (the “uncertainty set”), and
176 optimize over the worst-case realizations within the bounded uncertainty set. The method
177 is therefore well suited to applications where there is considerable uncertainty related to the
178 model input parameters, and when data uncertainties can lead to significant penalties or
179 infeasibility in practice. The solution method for robust optimization problems involves gen-
180 erating a deterministic equivalent, called the robust counterpart. Computational tractability
181 of the robust counterpart has been a major practical difficulty [34]. A variety of uncertainty
182 sets have been identified for which the robust counterpart to a robust optimization problem
183 is reasonably tractable [35].

184 The RO field has grown substantially over the past two decades. Seminal papers in the
185 late 1990s [36, 37] and early 2000s [38] established the field. Comprehensive surveys on the
186 early literature were done by Ben-Tal et al. [34] and Bertsimas et al. [35]. The development
187 of the robust optimization technique has allowed researchers to tackle problems with data
188 uncertainty in a range of fields. Examples can be found for renewable energy network design
189 [39], supply chain operations [40] and health care logistics [41].

190 *2.3. Applications of RO in Ride-hailing Operations*

191 In recent years, robust optimization applications in transportation, and ride-hailing rebal-
192 ancing more specifically, have attracted considerable research attention. Liu et al. [42] con-
193 sidered uncertain local demand in their matching algorithm for ridesharing operations. Miao
194 et al. [43] proposed an RO model for the taxi dispatching problem and tested it using NYC
195 taxi data. They also proposed a data-driven approach to construct the uncertainty set based
196 on historical demand data with a probability guarantee, building on previous data-driven
197 RO theory proposed by Bertsimas et al. [44]. He et al. [45] tackled the robust ride-hailing

198 rebalancing problem using linear decision rules (LDR) to create a multi-period adaptive RO
 199 (ARO) model. Their ARO-based approach is heavily based upon theory developed by Bert-
 200 simas et al. [46]. To the best of authors' knowledge, no existing papers have incorporated
 201 matching component and robust optimization techniques into vehicle rebalancing problems.
 202 This research gap is important to address given the prominent role of ride-hailing in urban
 203 transportation. Demonstrating how robustness and matching-integrated rebalancing can be
 204 combined in ride-hailing operations, and evaluating whether this combination of methods is
 205 advantageous, can help to improve future ride-hailing operations.

206 3. Methodology

207 3.1. Problem Description

208 Given an operation period \mathcal{T} , we first divide it into Ω identical time intervals indexed by
 209 $k = 1, 2, \dots, \Omega$, where the length of each time interval is Δ^1 . Figure 2 displays the framework
 210 of the MIVR model. The MIVR model is solved in a rolling-horizon manner, where decision
 211 variables are determined repeatedly at the beginning of each time interval. At the beginning
 212 of time interval k , κ future time intervals are incorporated in the MIVR model, and only
 213 the vehicle rebalancing decisions of the current time interval k are implemented. When
 214 proceeding to the next time interval, vehicle locations are observed and updated as the
 215 input for the MIVR model. Let $(k, k + 1, \dots, k + \kappa - 1)$ represent time intervals considered
 216 at time k , to simplify the notation, these time intervals are indexed by $k = 1, 2, \dots, \kappa$. The
 217 study region is partitioned into n sub-regions, each sub-region i has an estimated demand
 218 $r_i^k \geq 0$ at time k . We define the following two sets: $N = \{1, \dots, n\}$ representing the set
 219 of sub-regions and $K = \{1, \dots, \kappa\}$ representing the set of time intervals considered in the
 220 problem.

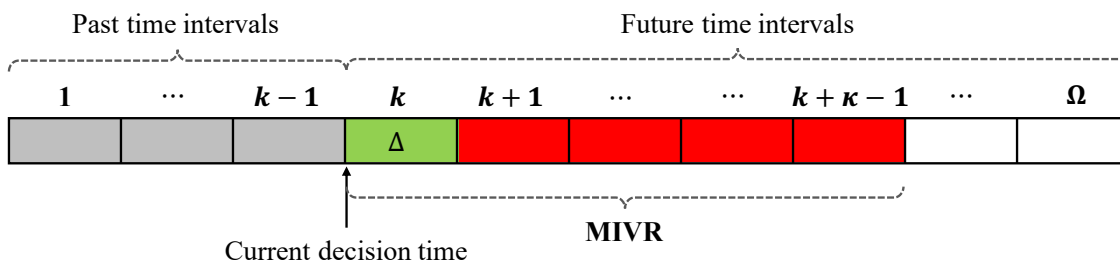


Figure 2: MIVR model framework. Each time interval has length Δ . Grey intervals indicate past time intervals that have been optimized. The green interval represents the current decision time interval and red intervals stand for look-ahead time within the MIVR model.

221 The MIVR model introduces the driver-customer matching component into the vehicle re-
 222 balancing problem by considering interzonal matchings based on estimated demand. Within
 223 a time interval k , the vehicle rebalancing phase happens at the beginning of the interval and
 224 the driver-customer matching phase is conducted at the end of the interval. In the vehicle
 225 rebalancing phase, decision variables are represented by $x_{ij}^k \in \mathbb{N}$ denoting the number of idle

¹The choice of Δ should depend on the size of sub-regions.

226 vehicles rebalanced from sub-region i to sub-region j at time k . Let $S_i^k \in \mathbb{N}$ indicate the
227 number of available vehicles in sub-region i at time k for the matching phase. Let d_{ij}^k, w_{ij}^k
228 denote the travel distance and time from sub-region i to sub-region j at time k , respectively,
229 which can be approximated by the distance and travel time between the centroids of two
230 sub-regions. We define a parameter $a_{ij}^k \in \{0, 1\}$ denoting whether an idle vehicle can be
231 rebalanced from sub-region i to sub-region j at time k , where $a_{ij}^k = 0$ if rebalancing between
232 sub-regions i, j is feasible at time k . The vehicle rebalancing from sub-region i to sub-region
233 j at time k is feasible if $w_{ij}^k \leq \Delta$, stipulating that the vehicle can be rebalanced to the des-
234 tination sub-region j within time interval k . Then the feasibility constraint of rebalancing
235 between sub-regions is given by:

$$a_{ij}^k \cdot x_{ij}^k = 0 \quad \forall i, j \in N, \forall k \in K. \quad (1)$$

236 This constraint does not prevent long-distance rebalancing decisions that occur over
237 several time periods, but rather limits the movement of rebalancing vehicles within a single
238 time period to zones that are reachable within that time period.

239 In the driver-customer matching phase, matching is considered between sub-regions with-
240 out considering actual demand and detailed locations of customers and vehicles. Let $y_{ij}^k \in \mathbb{N}$
241 denote the number of customers in sub-region i matched with vehicles in sub-region j at
242 time k . It is worth mentioning that decision variables y_{ij}^k of the matching component only
243 serve as auxiliary variables in the MIVR model, which focuses on computing the rebalanc-
244 ing decisions. When vehicle are rebalanced and requests are collected, the driver-customer
245 matching problem can then be solved by a separate driver-customer matching problem given
246 the realized demand. Let $T_i^k \in \mathbb{N}$ denote the number of unsatisfied requests in sub-region i
247 at time k . Then constraints related to the matching phase are:

$$\sum_{j=1}^n y_{ji}^k \leq S_i^k \quad \forall i \in N, \forall k \in K \quad (2a)$$

$$\sum_{j=1}^n y_{ij}^k \leq r_i^k \quad \forall i \in N, \forall k \in K \quad (2b)$$

$$T_i^k = r_i^k - \sum_{j=1}^n y_{ij}^k \quad \forall i \in N, \forall k \in K \quad (2c)$$

248 Constraints (2a) and (2b) restrict the interzonal matching decisions by the number of
249 available vehicles S_i^k and estimated demand r_i^k . Constraints (2c) define the number of un-
250 satisfied requests, which is equivalent to the number of customers who have not been as-
251 signed drivers within the current matching phase. When matching customers and drivers,
252 a maximum pickup time constraint is imposed to guarantee that customers do not experi-
253 ence excessive wait times. Let \bar{w} denote customers' maximum pickup time and parameter
254 $b_{ij}^k \in \{0, 1\}$ denote whether customers in sub-region i can be matched with drivers in sub-
255 region j at time k , where $b_{ij}^k = 0$ indicates a feasible interzonal matching. The matching
256 between customers in sub-region i and drivers in sub-region j at time k is feasible if $w_{ji}^k \leq \bar{w}$,
257 which enforces the maximum pickup time constraint. The matching feasibility constraint is

258 then

$$b_{ij}^k \cdot y_{ij}^k = 0 \quad \forall i, j \in N, \forall k \in K. \quad (3)$$

259 Next, we establish the connection between the two phases. Let $V_i^k, O_i^k \in \mathbb{N}$ represent the
 260 number of vacant and occupied vehicles for sub-region i at the beginning of time interval
 261 k , respectively. The initial vehicle locations, $V_i^1, O_i^1, \forall i \in N$, are inputs for the MIVR
 262 model. Other inputs to the model are regional transition matrices P^k, Q^k , which describe
 263 the dynamics of occupied vehicles. The entry (i, j) for P^k, P_{ij}^k , denotes the probability that
 264 an occupied vehicle located in sub-region i at time k will be in sub-region j and stay occupied
 265 at time $k + 1$. The entry (i, j) for Q^k, Q_{ij}^k , indicates the probability that an occupied vehicle
 266 starting in sub-region i at time k will be in sub-region j and become vacant at time $k + 1$.

267 In reality, the regional transition matrices depend on the spatio-temporal demand flows
 268 as well as the operator's dispatching and rebalancing strategies. The matching and rebalanc-
 269 ing decisions in the MIVR model are defined at interzonal level, and the regional transition
 270 matrices formulated with interzonal level decision variables are approximations to the real
 271 matrices. To reduce the model complexity, we further approximate the real regional transi-
 272 tion matrices with static matrices estimated from the historical data. [The impact of utilizing](#)
 273 [static transition matrices will be elaborated in the results section.](#) These matrices must
 274 satisfy the following constraints:

$$\sum_{j=1}^n (P_{ij}^k + Q_{ij}^k) = 1, \quad \forall i \in N, \forall k \in K.$$

275 Then, we specify the following relationships between S_i^k, V_i^k, O_i^k and decision variables
 276 x_{ij}^k, y_{ij}^k :

$$\sum_{j=1}^n x_{ij}^k \leq V_i^k \quad \forall i \in N, \forall k \in K \quad (4a)$$

$$S_i^k = V_i^k + \sum_{j=1}^n x_{ji}^k - \sum_{j=1}^n x_{ij}^k \quad \forall i \in N, \forall k \in K \quad (4b)$$

$$V_i^{k+1} = S_i^k - \sum_{j=1}^n y_{ji}^k + \sum_{j=1}^n Q_{ji}^k O_j^k \quad \forall i \in N, \forall k \in K \setminus \{\kappa\} \quad (4c)$$

$$O_i^{k+1} = \sum_{j=1}^n y_{ji}^k + \sum_{j=1}^n P_{ji}^k O_j^k \quad \forall i \in N, \forall k \in K \setminus \{\kappa\} \quad (4d)$$

277 Where constraints (4a) ensure that the number of vehicles in sub-region i that can be
 278 rebalanced to other sub-regions is bounded by the number of vacant vehicles. Constraints
 279 (4b) show that the available vehicles in sub-region i at time k consist of vacant and rebalanced
 280 vehicles. Similarly, constraints (4c) indicate that the set of vacant vehicles in sub-region i
 281 at time $k + 1$ is comprised of currently vacant vehicles at time k and currently occupied
 282 vehicles that become vacant in the next time interval. The number of unmatched vehicles at

time k , denoted by $S_i^k - \sum_{j=1}^n y_{ji}^k$, is equal to the difference between the number of available vehicles and the number of vehicles dispatched for interzonal matching. The number of occupied vehicles at time k that become vacant at time $k + 1$ in sub-region i is represented by $\sum_{j=1}^n Q_{ji}^k O_j^k$. Constraints (4d) state that occupied vehicles in sub-region i at time $k + 1$ are comprised of currently vacant vehicles that become occupied in the next interval as well as currently occupied vehicles at time k . The number of vacant vehicles that become occupied in sub-region i at time $k + 1$ because of interzonal matching at time k is indicated by $\sum_{j=1}^n y_{ji}^k$. The number of occupied vehicles at time k that stay occupied at time $k + 1$ in sub-region i is enforced by $\sum_{j=1}^n P_{ji}^k O_j^k$.

The objective for the MIVR model is minimizing the number of unsatisfied requests and the total vehicle travel distance, which consists of vehicle rebalancing distance and vehicle pickup distance. To construct the objective function as the generalized VMT for ride-hailing operations, we assume γ to be a parameter indicating the penalty VMT induced by each unsatisfied request. Let β represent a parameter that defines the relative weighting of rebalancing distance and pickup distance. The parameter β controls the trade-off between the total non-occupied VMT (from the system perspective) and the service quality (from the customer perspective). A larger β indicates a higher priority on minimizing the vehicle pickup distance, which leads to better service quality with a smaller customer wait time. When $\beta = 1$, the MIVR model purely minimizes the total VMT and the number of unsatisfied requests without explicitly putting any weight on the customer wait times².

$$(MIVR) \quad \min \quad Z = \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n x_{ij}^k d_{ij}^k + \beta \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n y_{ij}^k d_{ji}^k + \gamma \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n T_i^k \quad (5a)$$

s.t. Constraints (1), (2a) – (2c), (3), (4a) – (4d)

$$x_{ij}^k, y_{ij}^k \in \mathbb{N} \quad \forall i, j \in N, \forall k \in K \quad (5b)$$

$$S_i^k, V_i^k, O_i^k, T_i^k \in \mathbb{N} \quad \forall i \in N, \forall k \in K \quad (5c)$$

The MIVR model is an integer linear programming (ILP) problem with integer variables $x_{ij}^k, y_{ij}^k, S_i^k, V_i^k, O_i^k$ and T_i^k . ILP problems of this size and complexity can be difficult to solve in a reasonable time frame. To improve the computational performance of our model while producing satisfying results, we relax all integer variables in the problem to positive real numbers \mathbb{R}^+ . The rebalancing decisions used for implementations can be generated by rounding down the solutions generated by the relaxed model. The approximated rebalancing decisions are guaranteed to be feasible regarding to constraints (4a), which impose an upper-bound on the number of vehicles that can be rebalanced.

By incorporating matching decisions within vehicle rebalancing problem, the model also considers future matching distances in addition to the rebalancing distance, leading to “smarter” rebalancing decisions. Essentially, the MIVR reduces the cost of inaccurate demand estimation when rebalancing idle vehicles. Meanwhile, the MIVR model is a forward-

²The MIVR model implicitly weights the customer wait times because of the correlation between the vehicle pickup distance and wait times.

315 looking model by incorporating κ future time intervals into the model.

316 3.2. Robust Optimization Model Formulation

317 The estimation of the future demand r_i^k is crucial for vehicle rebalancing problems in
 318 ride-hailing systems. Previous studies have assumed the number of customers in any sub-
 319 region followed a Poisson distribution [5, 6]. However, in most applications we have limited
 320 knowledge about the “true” distribution for the future demand. The assumption that com-
 321 plex customer behaviour can be described by a simple probability distribution might be too
 322 strong. Instead of imposing a probability distribution on the future demand, we introduce
 323 the robust optimization technique where the uncertain demand parameters are described by
 324 uncertainty sets rather than specific probability distributions. The uncertainty sets specify
 325 a range for the uncertain demand where the demand can lie anywhere in the range.

326 First, we define the uncertainty set for the robust MIVR model. For the uncertainty in
 327 the demand originating in sub-region i within time interval k , we construct an uncertainty
 328 set \mathcal{U} from the intersection of two different sets: a box uncertainty set $\tilde{\mathcal{U}}_i^k$ and a polyhedral
 329 uncertainty set $\bar{\mathcal{U}}^k$ which constrains the total variation in demand across all sub-regions.
 330 The uncertainty set \mathcal{U} was selected to reflect the actual range of demand variability across
 331 different sub-regions without producing solutions that are too conservative in practice.

332 The box uncertainty set imposes upper and lower bounds of ρ standard deviations be-
 333 tween estimated regional demand and the mean regional demand at each time interval k .
 334 The parameter ρ is set according to the operator’s level of risk tolerance, with a higher ρ rep-
 335 resenting a lower tolerance for risk. The mean μ_i^k and standard deviation σ_i^k of the demand
 336 in sub-region i during time k are estimated with the historical data. The box uncertainty
 337 set for estimated demand r_i^k is then

$$\tilde{\mathcal{U}}_i^k(\rho) = \left\{ r_i^k : \left| \frac{r_i^k - \mu_i^k}{\sigma_i^k} \right| \leq \rho \right\} \quad \forall i \in N, \forall k \in K.$$

338 The polyhedral uncertainty set limits the total offset in the sum of the demand during a
 339 time interval across all sub-regions. This second restriction is intuitive; within a given time
 340 interval, demand may be above or below the mean in one region, but the total demand across
 341 the entire service area could be expected to remain at a similar level compared to previous
 342 days under most scenarios. Sub-regions with unusually high demand should be offset by other
 343 nearby sub-regions of low demand. The polyhedral uncertainty set for estimated demand r_i^k
 344 is

$$\bar{\mathcal{U}}^k(\Gamma) = \left\{ (r_1^k, \dots, r_n^k) : \left| \sum_{i=1}^n (r_i^k - \mu_i^k) \right| \leq \Gamma \right\} \quad \forall k \in K,$$

345 where Γ is the parameter to control the level of uncertainty for the polyhedral uncertainty
 346 set. It is worth noting that the construction of the uncertainty set indicates how much
 347 uncertainty the operator would like to tolerate in the operation. In reality, there exists
 348 scenarios where the total demand at certain time intervals exceed the historical mean by far,
 349 for instance ride-hailing demand after concerts or large events. It is wise for the ride-hailing
 350 operator to not take such unusual demand scenarios into consideration.

351 The combined uncertainty set \mathcal{U} for the estimated demand r_i^k is:

$$\mathcal{U} = \left[\bigcap_{i=1}^n \bigcap_{k=1}^{\kappa} \tilde{\mathcal{U}}_i^k(\rho) \right] \cap \left[\bigcap_{k=1}^{\kappa} \mathcal{U}^k(\Gamma) \right]$$

352 By defining an uncertain parameter $\zeta \in \mathbb{R}^{n\kappa}$ and letting $r_i^k = \mu_i^k + \zeta_i^k \sigma_i^k$, we can write \mathcal{U}
 353 as follows:

$$\mathcal{U} = \{ \zeta : \|\zeta\|_{\infty} \leq \rho; |e^T(\zeta^k \circ \sigma^k)| \leq \Gamma, \forall k \in K \}, \quad (6)$$

354 where $\zeta^k, \sigma^k \in \mathbb{R}^n$ are vectors for a specific time interval k , $e \in \mathbb{R}^n$ is a vector with all
 355 entries equal to one, and $\zeta^k \circ \sigma^k$ indicates the element-wise product for vectors ζ^k and σ^k .
 356 The parameters ρ and Γ control the size of the uncertainty set for estimated demand, and
 357 can be adjusted based on the operators' risk tolerance or desired probability guarantee for
 358 constraints involving uncertain parameters. Increasing the value of ρ and Γ leads to more
 359 conservative rebalancing decisions for the robust model.

360 Combining the MIVR model with the uncertainty set described above, we propose a
 361 robust MIVR model:

$$(P) \quad \min_{x_{ij}^k, y_{ij}^k} \quad Z = \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n x_{ij}^k d_{ij}^k + \beta \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n y_{ij}^k d_{ji}^k + \gamma \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n T_i^k \quad (7a)$$

$$\text{s.t.} \quad S_i^k = V_i^k + \sum_{j=1}^n x_{ji}^k - \sum_{j=1}^n x_{ij}^k \quad \forall i \in N, \forall k \in K \quad (7b)$$

$$V_i^{k+1} = S_i^k - \sum_{j=1}^n y_{ji}^k + \sum_{j=1}^n Q_{ji}^k O_j^k \quad \forall i \in N, \forall k \in K \setminus \{\kappa\} \quad (7c)$$

$$O_i^{k+1} = \sum_{j=1}^n y_{ji}^k + \sum_{j=1}^n P_{ji}^k O_j^k \quad \forall i \in N, \forall k \in K \setminus \{\kappa\} \quad (7d)$$

$$\sum_{j=1}^n x_{ij}^k \leq V_i^k \quad \forall i \in N, \forall k \in K \quad (7e)$$

$$\sum_{j=1}^n y_{ji}^k \leq S_i^k \quad \forall i \in N, \forall k \in K \quad (7f)$$

$$\sum_{j=1}^n y_{ij}^k \leq \mu_i^k + \zeta_i^k \sigma_i^k \quad \forall i \in N, \forall k \in K, \forall \zeta \in \mathcal{U} \quad (7g)$$

$$T_i^k = \mu_i^k + \zeta_i^k \sigma_i^k - \sum_{j=1}^n y_{ij}^k \quad \forall i \in N, \forall k \in K, \forall \zeta \in \mathcal{U} \quad (7h)$$

$$b_{ij}^k \cdot y_{ij}^k = 0 \quad \forall i \in N, \forall k \in K \quad (7i)$$

$$a_{ij}^k \cdot x_{ij}^k = 0 \quad \forall i \in N, \forall k \in K \quad (7j)$$

$$x_{ij}^k, y_{ij}^k \geq 0 \quad \forall i, j \in N, \forall k \in K \quad (7k)$$

$$S_i^k, V_i^k, O_i^k, T_i^k \geq 0 \quad \forall i \in N, \forall k \in K \quad (7l)$$

362 The problem (P) becomes infeasible even with a small value of ρ if the coefficient of
363 variation³ for uncertain demand is large for some sub-regions during certain time intervals.
364 Particularly, the problem (P) is infeasible if $\exists i \in N, \exists k \in K$ and $\rho \geq \frac{\mu_i^k}{\sigma_i^k}$. Because when
365 inequality $\rho \geq \frac{\mu_i^k}{\sigma_i^k}$ holds, the box uncertainty set $\tilde{\mathcal{U}}_i^k(\rho)$ allows ζ_i^k to take values smaller
366 than $-\frac{\mu_i^k}{\sigma_i^k}$, which leads to a negative uncertain demand, i.e., $r_i^k = \mu_i^k + \zeta_i^k \sigma_i^k < 0$. The
367 constraint (7g) is infeasible when the right-hand side is negative since the decision variable
368 y_{ij}^k is non-negative. To prevent infeasibility that can result from demand uncertainty, we add
369 restrictions on the uncertainty set in the problem (P) to guarantee that estimated demand
370 is non-negative:

$$\mu_i^k + \zeta_i^k \sigma_i^k \geq 0 \quad \forall i \in N, \forall k \in K, \forall \zeta \in \mathcal{U} \quad (8)$$

371 When modeling robust optimization problems, equality constraints with uncertain pa-
372 rameters should be avoided as much as possible since they dramatically shrink the feasible
373 region and often lead to infeasibility [47]. For the problem (P) with uncertain parameter
374 ζ , we must therefore reformulate equality constraints (7h). Equality constraints (7h) can
375 be avoided by eliminating variable T_i^k through substitution. After this variable elimination
376 step, objective function of problem (7a) becomes:

$$\min_{x_{ij}^k, y_{ij}^k} \left\{ \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n x_{ij}^k d_{ij}^k + \beta \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n y_{ij}^k d_{ji}^k + \max_{\zeta \in \mathcal{U}} \left[\gamma \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n (\mu_i^k + \zeta_i^k \sigma_i^k - \sum_{j=1}^n y_{ij}^k) \right] \right\}. \quad (9)$$

377 The objective function (9) with min-max formulation can be reformulated by introducing
378 an auxiliary variable ω :

$$\min \quad Z = \omega \quad (10a)$$

$$\text{s.t.} \quad \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n x_{ij}^k d_{ij}^k + \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n (\beta \cdot d_{ji}^k - \gamma) y_{ij}^k + \gamma \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n (\mu_i^k + \zeta_i^k \sigma_i^k) \leq \omega \quad \forall \zeta \in \mathcal{U} \quad (10b)$$

379 However, robust counterparts for equivalent formulations of the same problem are not
380 necessarily equivalent [47]. To reformulate the problem while maintaining an identical robust
381 counterpart, we make variables T_i^k *adaptive*, meaning that both variables are “wait-and-see”⁴
382 variables relating to uncertain parameters ζ , i.e., $T_i^k = T_i^k(\zeta)$. Introducing adaptive vari-
383 ables turns the initial RO problem into an Adaptive Robust Optimization (ARO) problem.
384 A commonly-used approximation method for solving ARO problems is the application of
385 Linear Decision Rules (LDRs), which has been shown to perform well in practice [34, 48].
386 Also, if the coefficients for the variables to be eliminated in the equality constraint do not in-
387 clude uncertain parameters and the constraint is linear in the uncertain parameters, making
388 such variables adaptive and applying LDRs is equivalent to directly eliminating them [47].

³Ratio of the standard deviation to the mean.

⁴The value of “wait-and-see” variables are determined only after the future demand is revealed.

389 Substitutions with equality constraint (7h) satisfies both conditions, therefore we eliminate
 390 variables T_i^k in the problem (P) to ensure no uncertain parameters appear in equality con-
 391 straints. The reformulation (P') is equivalent to an approximation for the original robust
 392 formulation (P) together with restriction (8) on the uncertainty set by applying LDRs:

$$(P') \quad \min \quad Z = \omega \quad (11a)$$

$$\text{s.t.} \quad \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n x_{ij}^k d_{ij}^k + \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n (\beta \cdot d_{ji}^k - \gamma) y_{ij}^k + \gamma \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n (\mu_i^k + \zeta_i^k \sigma_i^k) \leq \omega \quad \forall \zeta \in \mathcal{U} \quad (11b)$$

Constraints (7b) – (7g), (7i) – (7l), (8)

393 After the reformulation, uncertain parameters only appear in the constraints. The next
 394 step is to derive the robust counterpart for the robust MIVR model. Constraints (7g),
 395 (11b) and Equation (8) with uncertain parameter ζ can be written as the following generic
 396 formulation:

$$L(\cdot) + v^T \zeta \leq c \quad \forall \zeta \in \mathcal{U}, \quad (12)$$

397 where $L(\cdot)$ indicates a function of decision variables in problem (P'), v is a vector in
 398 dimension $n\kappa$ and c is a scalar. The robust counterpart for the generic constraint (12) is

$$\begin{cases} L(\cdot) + \rho \|\theta_0\|_1 + \Gamma \sum_{k=1}^{\kappa} (\eta_1^k + \eta_2^k) \leq c \\ (\eta_1^{k'} - \eta_2^{k'}) \sigma_i^{k'} = \theta_{k'}^{i,k} \quad \forall i \in N, \forall k = k' \in K \\ \theta_{k'}^{i,k} = 0 \quad \forall i \in N, \forall k \neq k' \in K \\ \eta_1^k, \eta_2^k \geq 0 \quad \forall k \in K \\ \sum_{k=0}^{\kappa} \theta_k = v \end{cases} \quad (13)$$

399 Where $\theta_k \in \mathbb{R}^{n\kappa}$ and $\theta_{k'}^{i,k}$ represents (ik) -th entry of vector $\theta_{k'}$, $\forall k' \in K$. The full
 400 derivation of the generic robust counterpart of (12) can be found in Appendix A. Then we
 401 derive the robust counterpart for problem (P'):

$$(RC) \quad \min \quad Z = \omega \quad (14a)$$

s.t. Constraints (7b) – (7f), (7i) – (7l)

$$\begin{aligned} & \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n x_{ij}^k d_{ij}^k + \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n (\beta \cdot d_{ji}^k - \gamma) y_{ij}^k + \gamma \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n \mu_i^k + \rho \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n \tilde{\theta}_0^{i,k} \\ & + \Gamma \cdot \sum_{k=1}^{\kappa} (\eta_1^k + \eta_2^k) \leq \omega \end{aligned} \quad (14b)$$

$$(\eta_1^{k'} - \eta_2^{k'}) \sigma_i^{k'} = \theta_{k'}^{i,k} \quad \forall i \in N, \forall k = k' \in K \quad (14c)$$

$$\theta_{k'}^{i,k} = 0 \quad \forall i \in N, \forall k \neq k' \in K \quad (14d)$$

$$\sum_{k'=0}^{\kappa} \theta_{k'}^{i,k} = \gamma \cdot \sigma_i^k \quad \forall i \in N, \forall k \in K \quad (14e)$$

$$-\tilde{\theta}_0^{i,k} \leq \theta_0^{i,k} \leq \tilde{\theta}_0^{i,k} \quad \forall i \in N, \forall k \in K \quad (14f)$$

$$\eta_1^k, \eta_2^k \geq 0 \quad \forall k \in K \quad (14g)$$

$$\sum_{j=1}^n y_{ij}^k + \rho \sum_{k'=1}^{\kappa} \sum_{i'=1}^n (\tau_{1,i,k}^{i',k'} + \tau_{2,i,k}^{i',k'}) + \Gamma \sum_{k'=1}^{\kappa} (\tau_{3,i,k}^{k'} + \tau_{4,i,k}^{k'}) \leq \mu_i^k \quad \forall i \in N, \forall k \in K \quad (14h)$$

$$\tau_{1,i,k}^{i',k'} - \tau_{2,i,k}^{i',k'} + \sigma_{i'}^{k'} (\tau_{3,i,k}^{k'} - \tau_{4,i,k}^{k'}) = 0 \quad \forall i', i \in N, \forall k', k \in K, (i', k') \neq (i, k) \quad (14i)$$

$$\tau_{1,i,k}^{i',k'} - \tau_{2,i,k}^{i',k'} + \sigma_{i'}^{k'} (\tau_{3,i,k}^{k'} - \tau_{4,i,k}^{k'}) = -\sigma_i^k \quad \forall i' = i \in N, \forall k' = k \in K \quad (14j)$$

$$\tau_{1,i,k}^{i',k'}, \tau_{2,i,k}^{i',k'} \geq 0 \quad \forall i', i \in N, \forall k', k \in K \quad (14k)$$

$$\tau_{3,i,k}^{k'}, \tau_{4,i,k}^{k'} \geq 0 \quad \forall i \in N, \forall k', k \in K \quad (14l)$$

$$\rho \sum_{k'=1}^{\kappa} \sum_{i'=1}^n (\nu_{1,i,k}^{i',k'} + \nu_{2,i,k}^{i',k'}) + \Gamma \sum_{k'=1}^{\kappa} (\nu_{3,i,k}^{k'} + \nu_{4,i,k}^{k'}) \leq \mu_i^k \quad \forall i \in N, \forall k \in K \quad (14m)$$

$$\nu_{1,i,k}^{i',k'} - \nu_{2,i,k}^{i',k'} + \sigma_{i'}^{k'} (\nu_{3,i,k}^{k'} - \nu_{4,i,k}^{k'}) = 0 \quad \forall i', i \in N, \forall k', k \in K, (i', k') \neq (i, k) \quad (14n)$$

$$\nu_{1,i,k}^{i',k'} - \nu_{2,i,k}^{i',k'} + \sigma_{i'}^{k'} (\nu_{3,i,k}^{k'} - \nu_{4,i,k}^{k'}) = -\sigma_i^k \quad \forall i' = i \in N, \forall k' = k \in K \quad (14o)$$

$$\nu_{1,i,k}^{i',k'}, \nu_{2,i,k}^{i',k'} \geq 0 \quad \forall i', i \in N, \forall k', k \in K \quad (14p)$$

$$\nu_{3,i,k}^{k'}, \nu_{4,i,k}^{k'} \geq 0 \quad \forall i \in N, \forall k', k \in K \quad (14q)$$

402 The constraints (14b) - (14g) represent the robust counterpart of constraints (11b). Con-
 403 straints (14h) - (14l) are the robust counterpart of constraints (7g) while constraints (14m)
 404 - (14q) are the robust counterpart of Equation (8). Compared to problem (P'), the robust
 405 counterpart (RC) introduces $(4n^2\kappa^2 + 5n\kappa^2 + 2n\kappa + 2\kappa)$ new auxiliary continuous variables.
 406 Although the number of decision variables increases considerably in the robust counterpart,
 407 this LP problem can be solved efficiently even for large-scale instances.

408 4. Empirical Study Design

409 In this section, we describe a real-time ride-hailing simulator used to compare the MIVR
 410 model with an independent VR model. To justify the benefit of introducing the robust op-
 411 timization technique into the vehicle rebalancing problem, a separate matching problem is
 412 solved over multiple demand scenarios to evaluate robust solutions and compare the nom-
 413 inal MIVR model with the robust MIVR model. We also describe the data used in the
 414 experiments.

415 4.1. Ride-hailing Simulator

416 The ride-hailing simulator is used to compare the nominal MIVR model with a benchmark
 417 VR model described in Appendix B. The results produced by this simulator allow us to
 418 evaluate the impact of the MIVR model independent of the robust optimization component.
 419 The simulation framework is shown in Figure 3.

420 *Data Input.* Data input for the ride-hailing simulator including the road network for
 421 the studied region with a shortest path distance matrix and a predecessor matrix, the set
 422 of n sub-regions N , a distance matrix d_{ij}^k and travel time matrix w_{ij}^k between centroids of

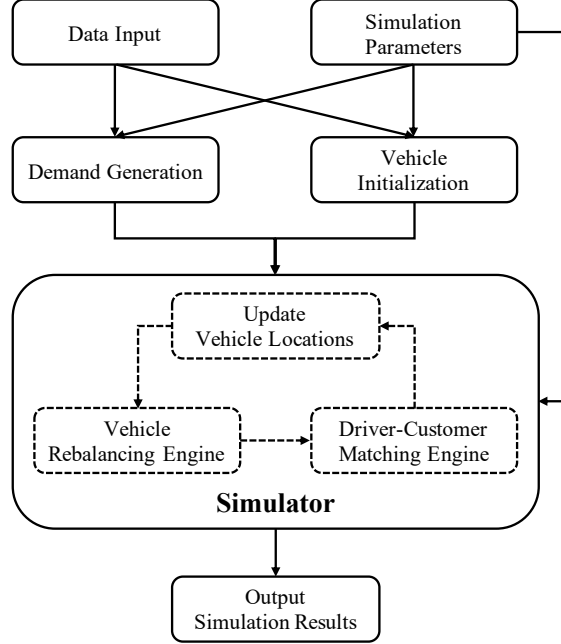


Figure 3: Ride-hailing simulation framework.

423 sub-regions, the set of Ω time intervals with length Δ , a mean μ_i^k of demand for each sub-
 424 region during each time interval, a full day of ride-hailing demand, and regional probability
 425 transition matrices for occupied vehicles P, Q and vacant vehicles P_v, Q_v . Details of the P, Q
 426 matrix estimation methods are provided in Appendix D. Data sources are described in detail
 427 in Section 4.3.

428 *Simulation Parameters.* Table 1 presents and explains the simulation parameters. Re-
 429 balancing decisions are solved with a model considering κ look-ahead time intervals.

Simulation Parameter	Explanation	Base Case Value
α	Cost parameter for regular rebalancing model	10^2
β	Weight parameter for pickup distance	1
γ	Cost parameter for unsatisfied requests	10^2
T_{start}	Start time of simulation	00:00
T_{end}	End time of simulation	24:00
Δ	Decision time interval length	300 (seconds)
δ	Matching batch size	30 (seconds)
κ	Number of time intervals considered in model	6
\bar{w}	Maximum pickup time	300 (seconds)
\tilde{w}	Maximum wait time	300 (seconds)
N_v	Number of vehicles	3000
\bar{v}	Average vehicle speed	20 (mph)

Table 1: Simulation parameters and base case value.

430 *Demand Generation.* Due to privacy concerns, historical TNC trip datasets typically

431 do not provide exact addresses or coordinates for trip origins and destinations. Given the
432 demand data at sub-regional level, we randomly assign road nodes within sub-regions as
433 origins and destinations.

434 *Vehicle Initialization.* At the start of the simulation period, the N_v vehicles are equally
435 likely to be in any sub-region i . The initial location for a vehicle within a sub-region i is
436 randomly assigned to a road node within i . All vehicles are considered to be available at the
437 beginning of the simulation.

438 *Simulator.* There are two main components contained in the simulator: the vehicle
439 rebalancing engine and driver-customer matching engine. Vehicle locations are updated at
440 the beginning of each simulation iteration. The simulator works as follows: at the beginning
441 of current simulation iteration, vacant and occupied vehicle locations are updated and used as
442 the input for vehicle rebalancing engine; vacant vehicles are rebalanced based on rebalancing
443 decision variables for the current iteration; within each simulation iteration, the driver-
444 customer matching engine can be run multiple times depend on the matching batch size
445 (e.g., 30 seconds); vehicles with assigned customers become occupied and start to pick up
446 customers and finish their trips.

447 *Driver-customer Matching Engine.* The optimal assignment problem for matching drivers
448 with customers in the simulator can be found in Appendix C. The objective of the optimal
449 assignment problem is minimizing the number of unsatisfied requests while minimizing the
450 pickup distance. The batch size of driver-customer matching engine is δ and customers will
451 leave the ride-hailing system if they wait longer than the maximum wait time \tilde{w} .

452 *Simulation Results.* We evaluated the simulation with the following vehicle-related indi-
453 cators: number of served customers, non-occupied VMT and number of rebalancing trips.
454 Customer wait time is used as the customer-related indicator to evaluate the simulation.
455 The customer wait time includes two components: the time for the vehicle to be assigned to
456 the customer, and the time for the assigned vehicle to travel to the pickup location.

457 4.2. Robust Solution Evaluation

458 Evaluating the solutions from the robust model requires multiple different demand sce-
459 narios due to the stochastic inputs. We compare the average performance of the model across
460 all demand scenarios in the study period for different uncertainty set sizes.

461 To evaluate the model performance under each demand scenario, we solve a separate
462 driver-customer matching problem after the demand is realized and the (nominal or ro-
463 bust) rebalancing decision x_{ij}^k (generated with estimated demand) is executed. The driver-
464 customer matching problem solved here is identical to the one solved in the simulator. The
465 overall pickup time and the number of unsatisfied customers are used as outputs to evaluate
466 robust solutions.

467 4.3. Data Description

468 The study area used in the experiments is the island of Manhattan in NYC. We used the
469 high-volume ride-hailing trip data collected by the NYC Taxi and Limousine Commission
470 [49] as the demand data. The sub-regions used in the experiments are “taxi zones” defined
471 within the high-volume ride-hailing trip dataset. There are 63 taxi zones on the island of
472 Manhattan ($N = 63$).

473 For benchmark comparisons of the nominal MIVR model, weekdays in June 2019 were
 474 chosen as the analysis period. Only trips that began and ended on the island of Manhattan
 475 were included. The mean and standard deviation of daily trip count by zone are shown in
 476 Figure 4 to illustrate the overall demand pattern. Demand is generally concentrated around
 477 dense residential areas on the eastern and western sides of Manhattan. There was an average
 478 of 294,422 high-volume ride-hailing trips per weekday during the sample period.

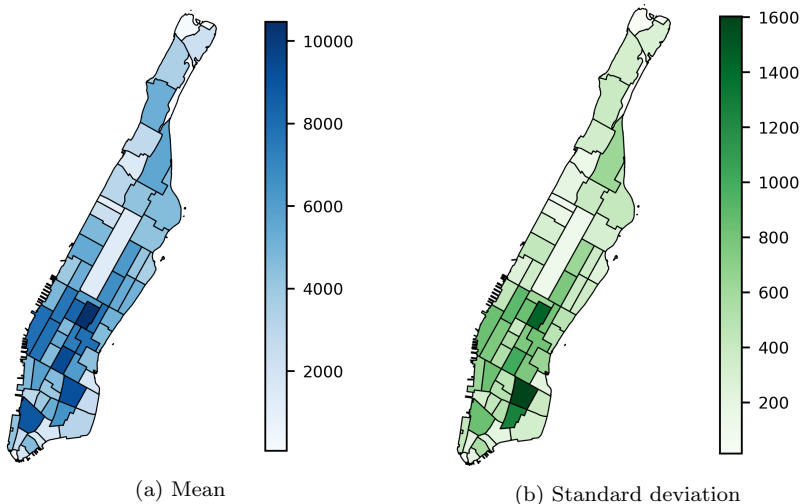


Figure 4: Average daily demand by zone (trips).

479 The full day of ride-hailing demand used in the simulation is from June 10, 2019. We chose
 480 a non-holiday Wednesday as it represents a typical day of demand from the study period.
 481 Figure 5 shows the comparison between the real demand and estimated demand⁵ aggregated
 482 into 5-minute time intervals. Based on the relationship between total real demand and total
 483 estimated demand, we identify four discrete demand scenarios over which the model can be
 484 tested:

- 485 I Low demand with accurate estimation (0 - 6): overall demand is relatively low and
 486 consistent with the historical average for this period.
- 487 II High demand with accurate estimation (6 - 10): overall demand is high and consistent
 488 with the historical average for this period.
- 489 III Demand underestimation (11 - 17): the total demand exceeds the historical average for
 490 this period.
- 491 IV Demand overestimation (20 - 24): the total demand is lower than the historical average
 492 for this period.

493 It is worth mentioning that an accurate prediction of the total demand does not lead
 494 to accurate sub-regional demand predictions. Demand uncertainties exist in every demand
 495 scenario and the overall level of uncertainty is higher in scenarios where demand is under-
 496 estimated or overestimated. Simulation results for each demand scenario are shown in the

⁵Mean demand μ_i^k is used as estimated demand.

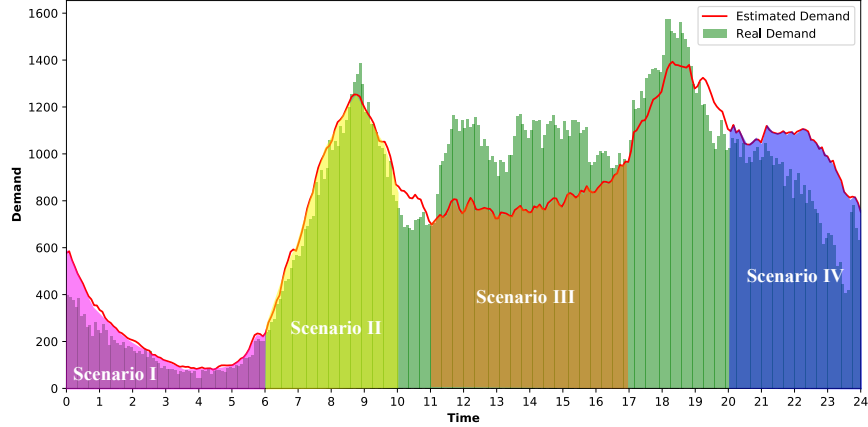


Figure 5: Estimated and real demand with four different types of demand scenarios. The demand is aggregated into 5-minute time intervals.

497 next section in order to illustrate the difference in model performance across two dimensions:
 498 demand level and prediction accuracy.

499 For evaluations of the robust MIVR model, we utilized the actual demand data for the
 500 65 week days from April to June 2019 to reflect the real demand uncertainty. Mean μ_i^k and
 501 standard deviation σ_i^k used in the robust MIVR model are generated from the same period.

502 The interzonal travel times for each time interval, w_{ij}^k , were collected from real travel
 503 speed data provided by the Uber Movement database for the study period of June 2019 [50].
 504 Hourly link-level travel speed is available for every link with at least five unique trips during
 505 the hour. First, the average hourly speed across all days in the study period was determined.
 506 The average hourly link travel speed was then used as an input to find the shortest path travel
 507 time between each zone pair for each hour in the day. Dijkstra’s algorithm [51] was used
 508 to determine the shortest path between zone centroids. The regional transition probability
 509 matrices for occupied and vacant vehicles, P , Q , P_v and Q_v , are generated based on the real
 510 travel time and demand data, and details are shown in Appendix D.

511 5. Results

512 All experiments in this paper are conducted on a 3.0 GHz AMD Threadripper 2970WX
 513 Processor with 128 GB Memory. The integer linear program and linear program in the
 514 experiments are solved with Gurobi 9.0 [52].

515 Presentation and discussion of the results is organized into three subsections. Section 5.1
 516 compares the MIVR model to two benchmark models: the VR model described earlier, and
 517 a recent state-of-the-art rebalancing model [25]. Section 5.2 explores the sensitivity of the
 518 MIVR results to variation in the model inputs. Finally, Section 5.4 provides the results for
 519 the robust MIVR model.

520 5.1. Benchmark Comparison

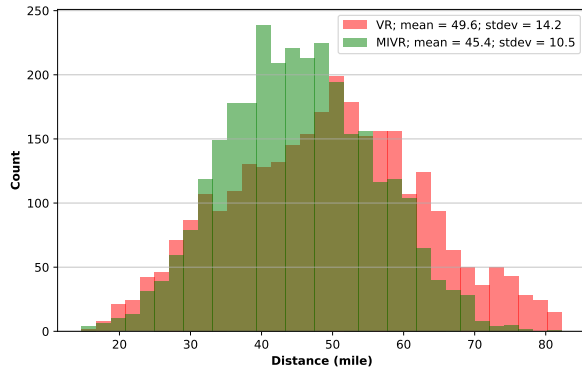
521 First, we compare the MIVR model with the benchmark VR model described in Appendix
 522 B and a fluid-based empty-car routing policy (FERP) proposed by Braverman et al. [25]. The
 523 performance of each model is assessed with the ride-hailing simulator described in Section 4.

524 To ensure a fair comparison, each vehicle rebalancing model uses the same demand profile
 525 and initial vehicle locations for each scenario.

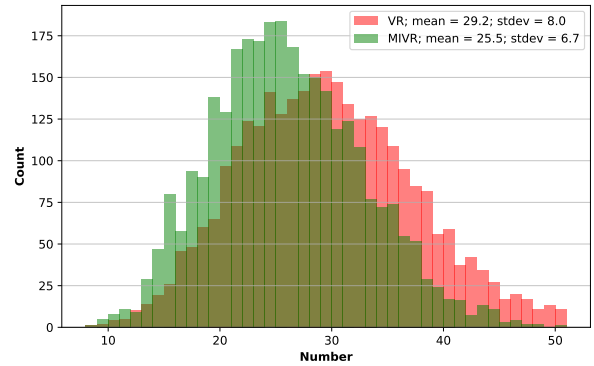
526 *5.1.1. Benchmark VR Model Comparison*

527 The base case scenario (full-day simulation) is tested with the simulation parameters
 528 shown in Table 1. The base case considers a scenario with 3000 vehicles, i.e., $N_v = 3000$,
 529 and 6 future time intervals in the vehicle rebalancing model, i.e., $\kappa = 6$. The base case
 530 scenario purely minimizes the number of unsatisfied requests and the total non-occupied
 531 VMT, i.e., $\beta = 1$. Both vehicle- and customer-related metrics are presented in Figure 6,
 532 where each figure shows the distributions for vehicles or customers for both MIVR and VR
 533 model results.

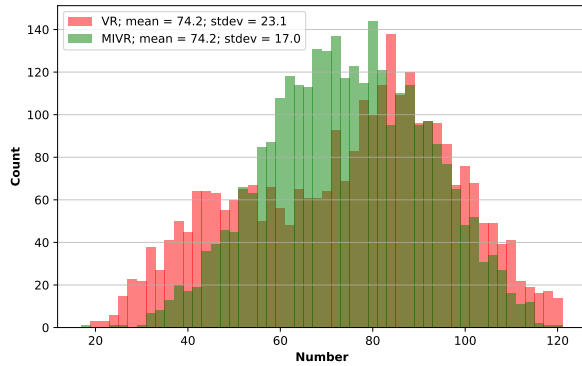
534 As shown in Figure 6a, the MIVR model reduces the non-occupied travel distance on
 535 average when compared to the VR model. Also, the number of vehicles with extremely
 536 long travel distance is reduced when utilizing the MIVR model. Figure 6b displays the
 537 rebalancing trip distributions, indicating that the MIVR dispatches fewer vacant vehicles
 538 for rebalancing purposes. The distribution of the number of served customers per vehicle
 539 is shown in Figure 6c. Although the average number of customers served by each vehicle
 540 is identical for two models, vehicles utilization is more evenly distributed under the MIVR
 541 model compared to the VR model. Figure 6d compares the wait time between the MIVR and



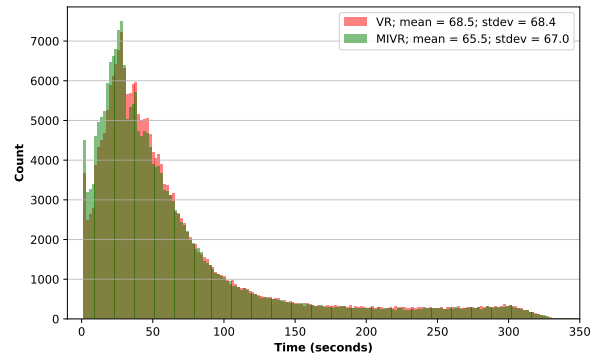
(a) Vehicle non-occupied travel distance distribution



(b) Vehicle rebalancing trip distribution



(c) Number of customers served (per vehicle) distribution



(d) Customer wait time distribution

Figure 6: Vehicle- and customer-related metrics in the simulation for the base case.

VR models. The average wait times are 65.5 and 68.5 seconds for each model, respectively. This occurs because the MIVR model reduces the number of customers with longer wait times. The fraction of unsatisfied requests for both models is less than 0.1%. Under the base case scenario, the MIVR model reduces customer wait time by 4.4% on average and total non-occupied VMT by 8.5%.

To better understand the model performance relative to the magnitude of demand and the level of prediction accuracy, we compared the MIVR model with the VR model over the four demand scenarios described in Section 4.3. Figure 7 displays the non-occupied vehicle travel distance distributions and Figure 8 shows the customer wait time distribution over the four demand scenarios. For the low demand with accurate estimation (I) and demand underestimation (III) scenarios, the MIVR model outperforms the VR model by significantly reducing customer wait time while also reducing the average vehicle non-occupied travel distance. In the high demand with accurate estimation scenario (II), the MIVR model reduces customer wait time by proactively rebalancing vehicles more frequently than the VR model. In the demand overestimation scenario (IV), the MIVR model is outperformed by the VR model as the VR model leads to lower average customer wait time and average vehicle non-occupied travel distance. The detailed simulation results for each demand scenario can be found in Appendix E.

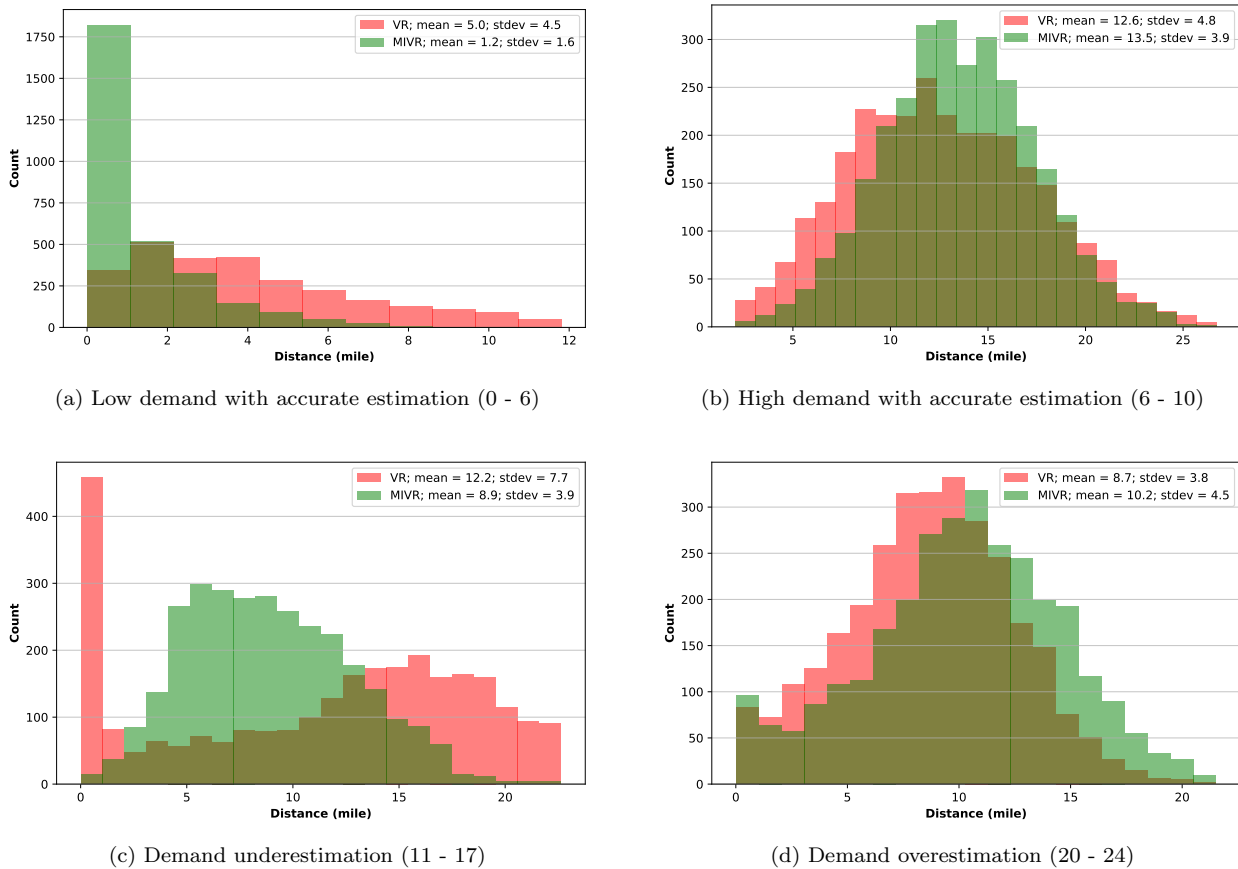


Figure 7: Vehicle non-occupied travel distance distributions for different demand scenarios.

To summarize, the MIVR model dispatches more vacant vehicles than the VR model

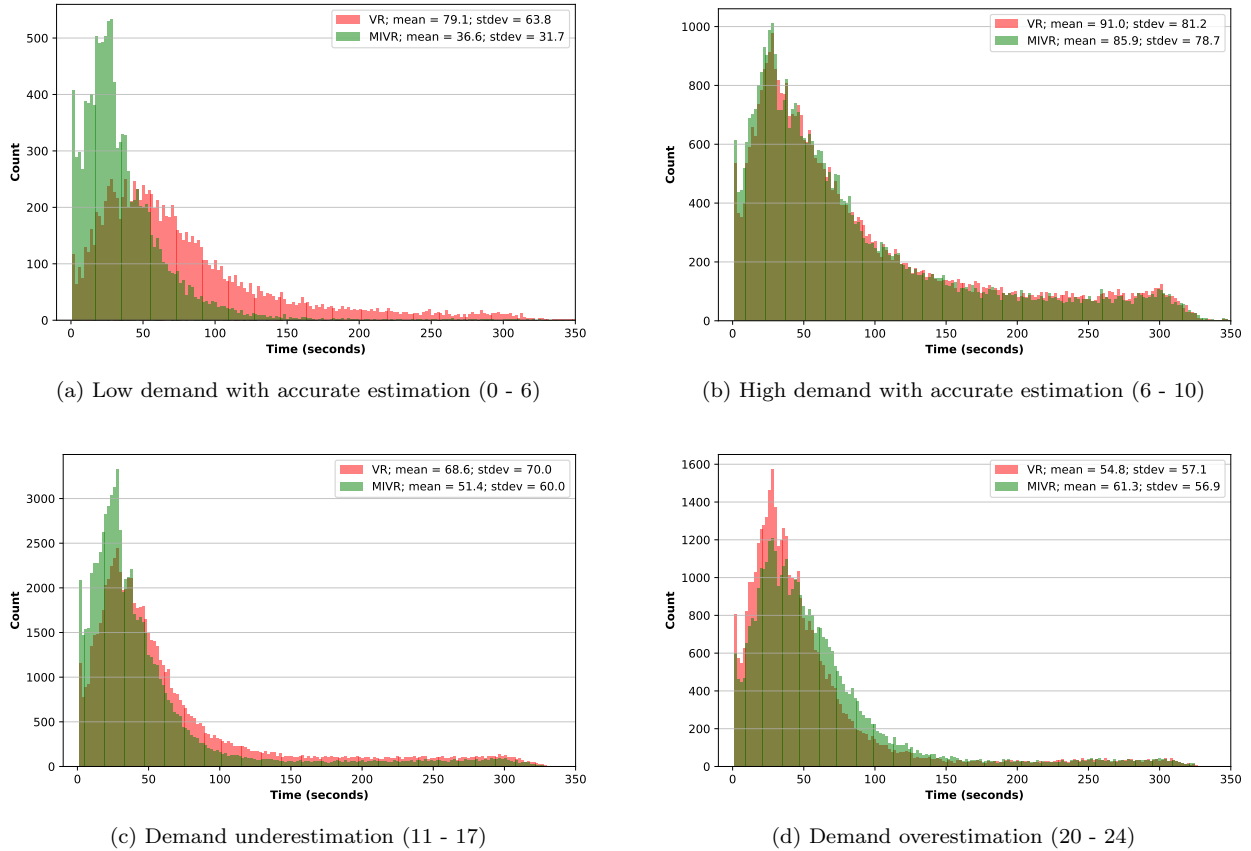


Figure 8: Customer wait time distributions for different demand scenarios.

561 when the level of estimated demand is high (given a specific fleet size N_v). On the other
 562 hand, fewer vehicles are dispatched by the MIVR model compared to the VR model when the
 563 level of estimated demand is low. This conclusion is further substantiated in Section 5.2.1,
 564 which discusses the results under different fleet sizes. We observe that the MIVR model is
 565 less proactive on dispatching vacant vehicles compared to the VR model when the fleet size
 566 is large relative to the level of demand.

567 In this section, we have shown that the performance of rebalancing models, as measured
 568 by the average customer wait time, depends on the accuracy of demand prediction and the
 569 level of demand. When the error in demand prediction is low, the MIVR model reduces
 570 the average customer wait time compared to the VR model. Model performance is penal-
 571 ized when the error in demand prediction is high (the total demand is underestimated or
 572 overestimated). Additionally, a rebalancing model which dispatches more vacant vehicles
 573 suffers higher penalties due to inaccurate demand estimation. In the demand scenario III,
 574 the level of predicted demand is low and the MIVR model dispatches fewer vacant vehicles
 575 than the VR model. Therefore, the MIVR model performs better than the VR model by
 576 reacting less often to inaccurate demand estimation. In the demand scenario IV, the level
 577 of predicted demand is high and the MIVR model dispatches more vacant vehicles than the
 578 VR model. The MIVR model experiences a higher penalty due to inaccurate demand esti-
 579 mations because of a proactive rebalancing strategy; hence, it performs worse than the VR

580 model under these conditions. The demand scenario IV implies that the demand prediction
 581 serves a critical role in the performance MIVR model. These results therefore demonstrate
 582 the value of a *robust* MIVR model that explicitly considers demand uncertainty.

583 *5.1.2. Benchmark FERP Comparison*

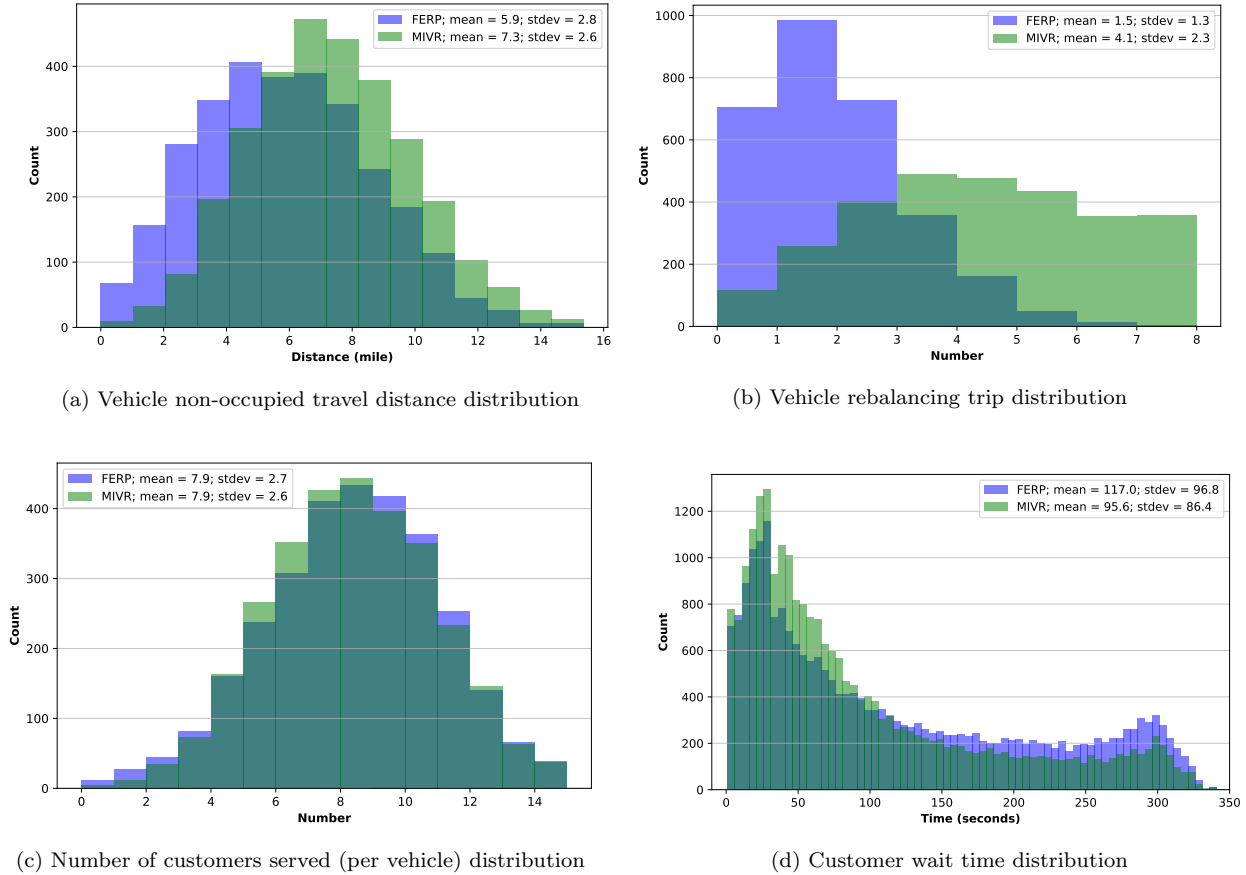


Figure 9: Benchmark comparison results between MIVR and FERP models.

584 To further evaluate the performance of proposed MIVR model, we compared our approach
 585 with a state-of-the-art method for solving the vehicle rebalancing problem [25]. Braverman
 586 et al. [25] formulated a fluid-based optimization problem to generate a static empty-car
 587 routing policy. To guarantee a fair comparison, we chose a two-hour time period (7AM -
 588 9AM) with historical demand and travel time data from June 2019 and 3000 vehicles to
 589 compute a static empty-car routing policy. We implemented the static routing policy in the
 590 simulator to dispatch vacant vehicles at each time interval instead of solving an optimization
 591 problem. Comparison results are shown in Figure 9.

592 Figure 9a displays the distributions of non-occupied vehicle travel distance and Figure
 593 9b shows the vehicle rebalancing trip distributions. The MIVR model dispatches vacant
 594 vehicles more proactively than the FERP. The distributions of number of customers served
 595 per vehicle are presented in Figure 9c, where vehicles are utilized slightly more evenly by the
 596 MIVR model than the FERP. Figure 9d displays the customer wait time distributions. The

597 MIVR model reduces the average customer wait time by 18% while increasing total non-
 598 occupied VMT by 24%. The proportion of unsatisfied requests for both approaches is less
 599 than 0.1%, which is a result of the adequate supply of vehicles. The MIVR model optimizes
 600 rebalancing decisions during each time interval and the FERP maintains the same vehicle
 601 rebalancing policy throughout the simulation period. In general, the MIVR model provides
 602 better service quality for customers by producing a more proactive rebalancing strategy, but
 603 it also results in a somewhat higher non-occupied VMT.

604 5.2. Scenario Testing

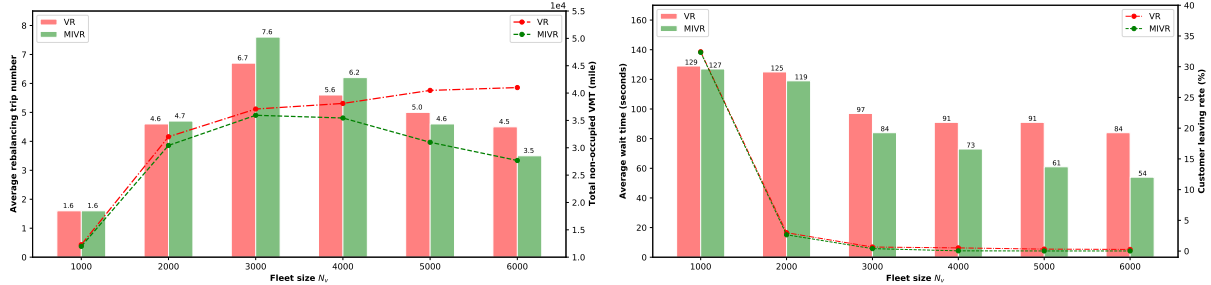
605 Second, we test the sensitivity of the results when changing input parameters of the
 606 MIVR model, including the fleet size, N_v , the length of decision time interval, κ , the weight
 607 parameter for pickup distance, β , and the size of the sub-regions. To avoid the effect of inac-
 608 curate demand estimation when testing different scenarios⁶, we tested different scenarios with
 609 N_v , κ and β over a four-hour time period (6 AM - 10 AM) assuming perfect future demand
 610 predictions. Alternative scenarios are generated by changing the simulation parameters for
 611 the base case.

612 5.2.1. Fleet size N_v

613 Results for scenarios with varying fleet sizes, represented by N_v in the simulation pa-
 614 rameters, are shown in Figure 10. When there is a limited number of vehicles ($N_v \leq 4000$)
 615 in the system, the MIVR model generates more rebalancing trips per vehicle compared to
 616 the VR model. When there are sufficient vehicles in the system ($N_v = 5000$ or 6000), the
 617 MIVR model dispatches fewer vacant vehicles and reduces the total non-occupied VMT com-
 618 pared to the VR model. This is intuitive; for the MIVR model, less rebalancing is needed
 619 when there is a higher concentration of idle vehicles since more passengers can be picked up
 620 (within the maximum wait time constraint) without significant rebalancing. Therefore, the
 621 MIVR model reduces the total non-occupied VMT. The MIVR model decreases the average
 622 customer wait time under all scenarios with different fleet sizes compared to the VR model.
 623 Customer wait time decreases significantly for the MIVR model when a larger fleet is avail-
 624 able. Even though rebalancing is not as critical for a large fleet, the MIVR model continues
 625 to minimize pickup distance and therefore customer wait time. The proportion of unsatisfied
 626 requests is marginally decreased for the MIVR model compared to the VR model, regardless
 627 of fleet size.

628 The scenario testing with different fleet sizes implies the existence of the *Pareto* improve-
 629 ment at aggregate level for the MIVR model compared to the VR model. When a sufficient
 630 number of vehicles is available, the MIVR model reduces the total non-occupied VMT, aver-
 631 age vehicle rebalancing trips and average customer wait time while satisfying more requests
 632 compared to the VR model. For instance, when there are 6000 vehicles in the system (with
 633 $\kappa = 6$), the MIVR model reduces the total non-occupied VMT by 33%, average vehicle
 634 rebalancing trips by 22% and average customer wait time by 36% when compared to the VR
 635 model. Under this scenario, the MIVR model clearly outperforms the VR model, indicating
 636 that the *Pareto* improvement exists.

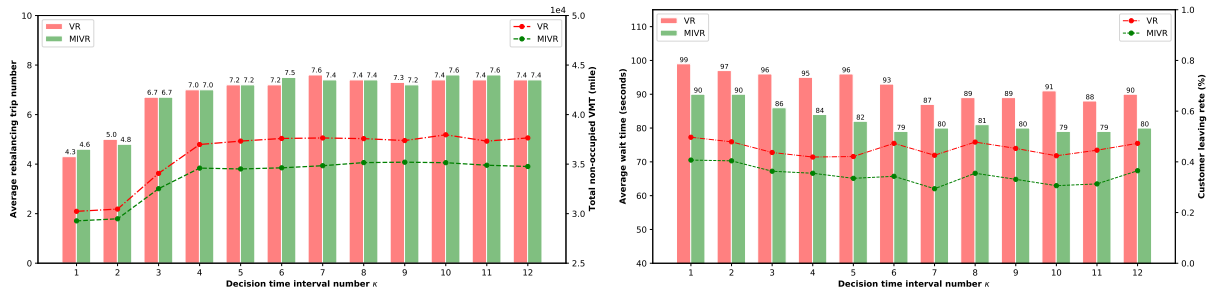
⁶The effect of input parameters on the simulation results can be overshadowed by the effect induced by inaccurate demand estimations when two have contradictory effects on certain performance metrics.



(a) Average number of rebalancing trips per vehicle and total non-occupied VMT. Bars indicate the average rebalancing trip number per vehicle and dashed lines show the total non-occupied VMT. (b) Average wait time per customer and proportion of unsatisfied requests. Bars indicate the average wait time per customer and dashed lines show the proportion of unsatisfied requests.

Figure 10: Scenario testing results for different fleet size N_v .

637 5.2.2. Decision time interval length κ



(a) Average number of rebalancing trips per vehicle and total non-occupied VMT. Bars indicate the average rebalancing trip number per vehicle and dashed lines show the total non-occupied VMT. (b) Average wait time per customer and proportion of unsatisfied requests. Bars indicate the average wait time per customer and dashed lines show the proportion of unsatisfied requests.

Figure 11: Scenario testing results for different decision time interval length κ .

638 Figure 11 shows the results under scenarios with varying decision time intervals κ . Both
 639 models dispatch more vehicles when considering additional future time intervals (i.e. κ
 640 becomes large), and similar amount of vacant vehicles are dispatched by both models. Also,
 641 the total non-occupied VMT increases when considering more future time intervals for both
 642 models, and the MIVR model leads to less non-occupied VMT compared to the VR model
 643 for all scenarios. With respect to customer wait time, considering additional time intervals
 644 benefits both models and the MIVR model reduces wait times for all scenarios compared
 645 to the VR model. The MIVR outperforms the VR model on the proportion of unsatisfied
 646 requests for all scenarios.

647 Note that selecting number of time intervals presents a trade-off between system perfor-
 648 mance and computation time. Increasing κ linearly increases the size of the problem, which
 649 may result in a solution time that is too long to use in practice. The average computation
 650 time for solving the MIVR model with $\kappa = 6$ is 3.8 seconds and the average computation
 651 time for the MIVR model with $\kappa = 12$ is 7.5 seconds. Platform operators must therefore
 652 choose a look-ahead window that is suited to their system size and computational capacity.

653 5.2.3. Weight parameter for pickup distance β

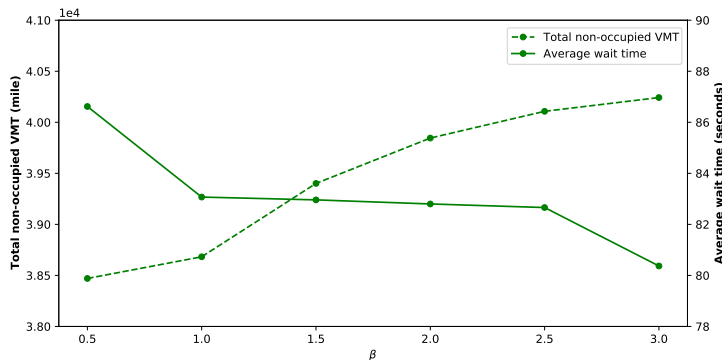


Figure 12: Sensitivity testing results for the weight parameter β in the MIVR model. Solid line indicates the average customer wait time and dashed line represents the total non-occupied VMT.

654 The weight parameter β in the MIVR model controls the trade-off between the total
 655 non-occupied VMT and the service quality. In previous experiments, $\beta = 1$ was used as a
 656 base case, leading to a MIVR model which purely minimized the total non-occupied VMT
 657 and the number of unsatisfied requests. In this section, different values of β are tested based
 658 on the base case simulation setting assuming perfect future demand predictions, and the
 659 total non-occupied VMT and the average customer wait time are shown in Figure 12.

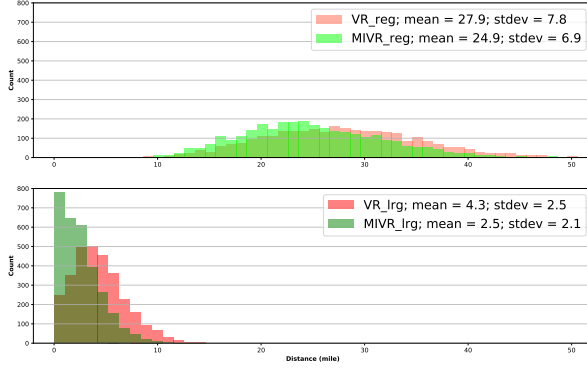
660 When β becomes larger, the MIVR model puts more weight on the service quality (cus-
 661 tomer wait times), and the total non-occupied VMT gets larger. The average customer wait
 662 time monotonically decreases when β increases. By increasing the value of β to 3, the average
 663 wait time is reduced by 3% while increasing the total non-occupied VMT by 4%. However,
 664 the MIVR model becomes more vulnerable to the demand uncertainty when the value of β
 665 is large. This is because more vacant vehicles are rebalanced when β is large, where a larger
 666 penalty is induced by the inaccurate demand estimations. Therefore, the service quality can
 667 be diminished if β is too large.

668 On the other hand, a negative weight is put on the service quality when $\beta < 1$, meaning
 669 that the service quality is sacrificed to reduce the total non-occupied VMT. For the scenario
 670 with $\beta = 0.5$, the total non-occupied VMT is reduced by 0.5% and the average wait time
 671 is increased by 4% compared to the base case. Since the vehicle rebalancing distance is
 672 highly correlated with customer wait time, reducing β does not significantly decrease the
 673 total non-occupied VMT.

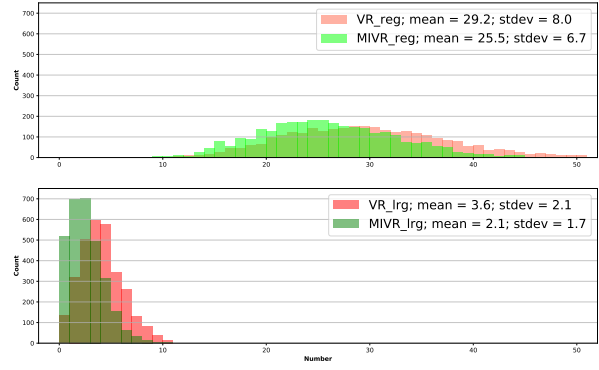
674 5.2.4. Sub-regional size

675 The MIVR model performance relies on the size of sub-regions. Smaller sub-regions leads
 676 to more rebalancing options (decision variables) and a better overall model performance.
 677 However, the model complexity increases when considering smaller sub-regions. To quantify
 678 the effect of changing the size of sub-regions, we combined 63 taxi zones into 13 larger zones
 679 and ran simulations for the 13 large sub-regions. Comparison results are shown in Figure
 680 13.

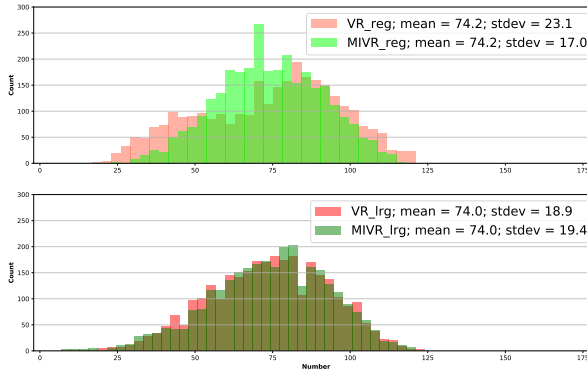
681 Figure 13a and 13b show the distributions of non-occupied vehicle travel distance and
 682 vehicle rebalancing trips. Fewer sub-regions with larger size reduces the opportunities for



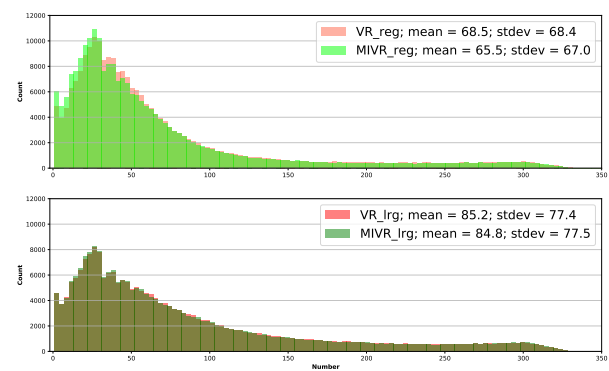
(a) Vehicle non-occupied travel distance distribution



(b) Vehicle rebalancing trip distribution



(c) Number of customers served (per vehicle) distribution



(d) Customer wait time distribution

Figure 13: Results comparison between simulations with 63 regular sub-regions and 13 large sub-regions.

683 rebalancing vacant vehicles between sub-regions. Therefore, both the average vehicle non-
 684 occupied VMT and rebalancing trips are significantly decreased. The distribution of number
 685 of customers served per vehicle is shown in Figure 13c, where vehicles are more evenly utilized
 686 by the MIVR model under a smaller sub-region size. Figure 13d displays the customer wait
 687 time distributions for both scenarios. Compared to the scenario with larger sub-regions, the
 688 scenario with 63 sub-regions leads to 20% and 23% reductions on the average customer wait
 689 time for the VR and the MIVR, respectively. Differences between the MIVR and VR models
 690 hold regardless of the size of the sub-regions.

691 As for the computation complexity, the average running time for producing rebalancing
 692 decisions during each iteration by the MIVR model under a regular sub-region size is 3.95
 693 seconds. The average running time for the MIVR model under a larger sub-region size is
 694 0.18 seconds. Reducing the number of sub-regions from 63 to 13 saves approximately 95% of
 695 the computation time on generating rebalancing decisions. In general, the size of sub-regions
 696 should be chosen to balance computation complexity and model performance.

697 5.3. Impact of Regional Transition Matrices

698 In the MIVR model, we utilized static regional transition matrices P and Q , which are
 699 estimated from the historical data, to reflect the movement of occupied vehicles. However,
 700 the true regional transition matrices depend spatio-temporal demand flows and operators'

701 dispatching and rebalancing strategies. In this section, we will quantify the impact of ap-
 702 proximating true regional transition matrices with the historical data.

703 To incorporate the true regional transition matrices in the model, we modified the sim-
 704 ulator by estimating regional transition matrices for occupied vehicles based on preceding
 705 matching decisions at the beginning of each simulation period. By using the previous match-
 706 ing decisions in the simulation, only regional transition matrices between the current time
 707 period k and the next time period $k + 1$ can be evaluated accurately. Therefore, we imple-
 708 mented a MIVR model with $\kappa = 2$ in the simulation, indicating that two time intervals were
 709 considered when making rebalancing decisions. Other simulation parameters are identical to
 710 the base case scenario. Such a modified simulator is able to produce rebalancing decisions
 711 based on the true regional transition matrices at each time interval.

712 To quantify the impact of approximating regional transition matrices with the historical
 713 data, we compared results from the modified simulator to results from a standard simulator
 714 described in section 4.1 with $\kappa = 2$, which guarantees identical look-ahead windows in the
 715 MIVR model. Results are compared within a four-hour time period (8AM - 12PM) and
 716 detailed comparison results are shown in Figure 14.

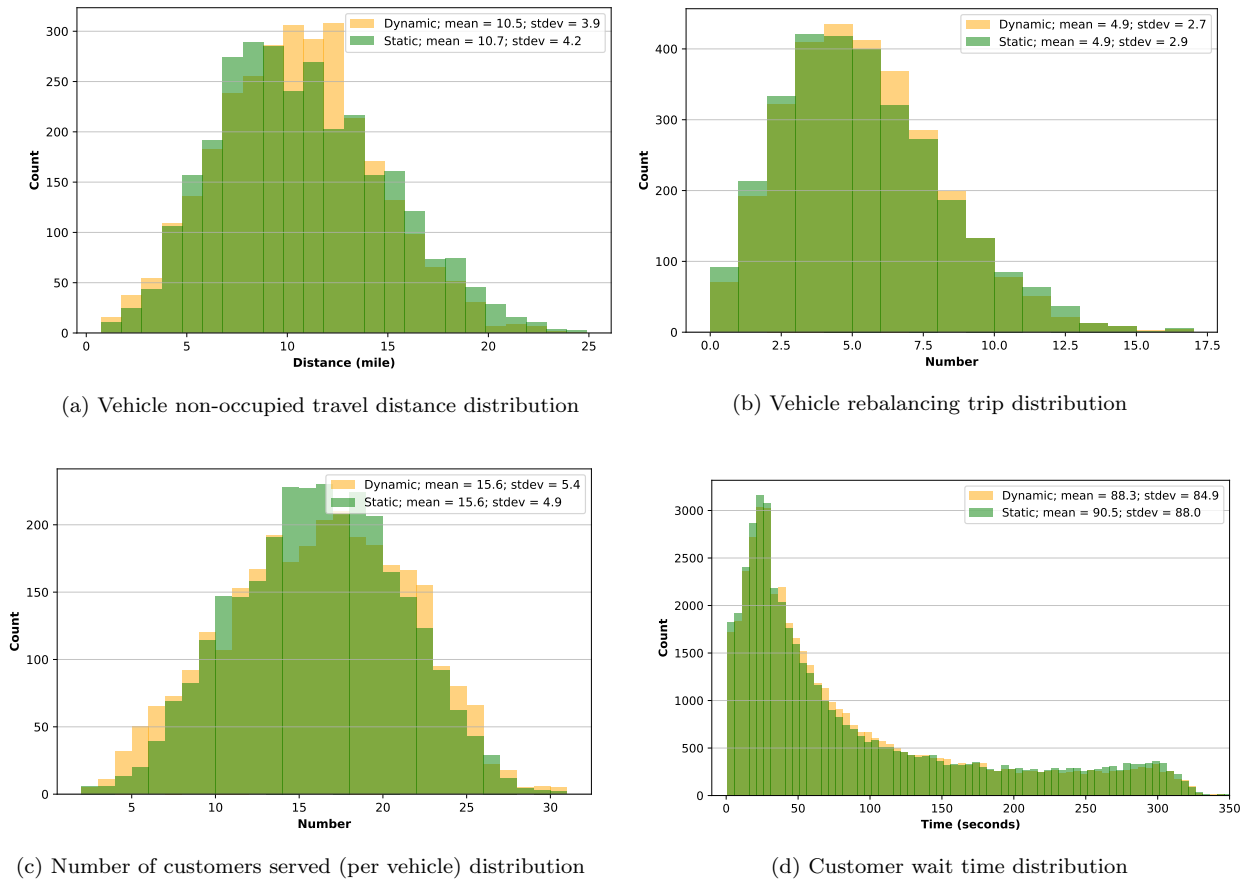


Figure 14: Comparison results between simulators with dynamic and static regional transition matrices. *Dynamic* indicates that regional transition matrices are estimated at the beginning of every simulation time interval. *Static* implies that regional transition matrices estimated by the historical data are utilized.

717 Figure 14a shows the distributions of vehicle non-occupied travel distance. Utilizing true

718 regional demand matrices reduces the total non-occupied VMT by 1.9%. Distributions of
719 vehicle rebalancing trips and number of customers served are displayed in Figure 14b and
720 14c, where two simulators have identical performance on average. Figure 14d presents the
721 distributions of customer wait time. Using true regional transition matrices reduces the
722 average customer wait time by 2.4%.

723 The comparison results imply that approximating the true regional transition matrices
724 with static matrices estimated from the historical data has a marginal impact on model
725 performance. This is intuitive; the regional transition matrices are used for constructing a
726 forward-looking vehicle rebalancing model. In the simulation, only the rebalancing decisions
727 for the first time interval will be implemented, although rebalancing decisions for κ time
728 periods are generated. When moving to the next time period, real-time information (e.g.,
729 vehicle locations) is updated and a separate MIVR model considering κ time intervals is
730 solved. Therefore, regional transition matrices have a limited impact on rebalancing decisions
731 at the first time interval, which subsequently has a marginal impact on model performance.

732

733 5.4. Robust Model Results

734 To evaluate the robust optimization model, we tested multiple scenarios with different
735 levels of uncertainty as defined by the uncertainty set size parameters ρ and Γ . Each robust
736 solution was generated for the robust MIVR model considering 6 future time intervals, i.e.,
737 $\kappa = 6$. The model parameters were set as $\beta = 1$, $\gamma = 10^2$ and $\bar{w} = \tilde{w} = 300$. For the
738 number of vehicles N_v , we considered the scenario with 3000 vehicles, indicating a sufficient
739 supply (almost all customers can be served) given the demand profile, and 2000 vehicles,
740 representing an insufficient supply. The initial vehicle distributions V_1 and O_1 are generated
741 using the following process: each vehicle in the fleet with size N_v is either vacant or occupied
742 with equal probability and is randomly assigned to a sub-region. To test the performance
743 for different solutions, we utilized the real demand data from 9 AM - 9:30 AM for 65 work
744 days from April to June 2019 and solved a driver-customer matching problem with realized
745 demand and vehicle distributions after rebalancing. The performance of each solution was
746 evaluated based on the average values of the total pickup time and the number of unsatisfied
747 requests over the 65 demand scenarios. The solution generated by the nominal MIVR model
748 was used as the benchmark for evaluating robust solutions. The performance of each robust
749 solution is displayed as the percentage reduction in performance measurements compared to
750 the nominal solution.

751 For the scenario with insufficient supply ($N_v = 2000$ and a proportion of customers can
752 not be served), Table 2 shows the results about the total pickup time and and Table 3
753 displays the percentage reduction for the number of unsatisfied requests over the nominal
754 MIVR model. Introducing uncertainty into the model generates solutions that outperform
755 the nominal solution for all values of ρ and Γ . The uncertain parameter ρ significantly affects
756 the total pickup time and the number of unsatisfied requests while the uncertain parameter
757 Γ has limited impact on them. When a high level of uncertainty is considered in development
758 of the robust MIVR model, more customers can be served with less total pickup time.

759 For the scenario with sufficient number of vehicles ($N_v = 3000$ and almost all customers
760 can be served), the percentage reduction of the total pickup time is shown in Table 4. The
761 robust MIVR model benefits more when having a large fleet of vehicles in the system. The

$\rho \backslash \Gamma$	0	1	2	3	4	5	6	7	8	9	10
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
0.2	0.51	1.38	1.18	0.66	1.18	0.66	1.18	1.18	0.66	0.79	1.38
0.3	2.45	2.45	4.52	2.45	2.45	2.45	2.45	4.52	4.52	2.45	4.52
0.4	3.47	4.15	4.15	4.15	4.15	4.15	5.67	4.15	4.15	5.67	5.6
0.5	5.62	5.62	5.62	5.62	5.62	5.62	5.62	5.62	7.68	7.68	5.62
0.6	7.89	7.89	7.89	7.89	7.89	7.89	7.89	7.89	7.89	7.89	7.89
0.7	8.78	10.32	10.32	10.32	10.32	10.32	10.32	10.32	10.32	10.32	10.32
0.8	13.24	13.24	13.24	13.24	13.24	13.24	13.24	13.24	13.24	13.24	13.24
0.9	17.17	17.17	17.17	18.59	17.17	17.17	17.17	19.78	18.59	18.59	17.17
1.0	21.19	19.92	21.23	21.23	21.23	21.23	21.23	21.23	21.23	21.23	21.23

Table 2: Percentage reduction in the total pickup time compared to the nominal MIVR solution with insufficient supply ($N_v = 2000$), for different values of ρ and Γ . Gray cells indicate uncertain scenarios with the largest reduction in pickup time.

$\rho \backslash \Gamma$	0	1	2	3	4	5	6	7	8	9	10
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.2	0.2	0.41	0.21	0.2	0.21	0.2	0.21	0.21	0.2	0.2	0.41
0.3	0.17	0.17	0.61	0.17	0.17	0.17	0.17	0.61	0.61	0.17	0.61
0.4	0.14	0.14	0.14	0.14	0.14	0.14	0.22	0.14	0.14	0.22	0.3
0.5	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.3	0.3	0.08
0.6	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
0.7	0.2	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
0.8	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
0.9	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
1.0	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54

Table 3: Percentage reduction in the number of unsatisfied requests compared to the nominal MIVR solution with insufficient supply ($N_v = 2000$). Gray cells indicate uncertain scenarios with the largest reduction in unsatisfied requests.

762 largest total pickup time reduction for the robust MIVR model with sufficient supply is
763 41.03% compared to 21.23% for the scenario with insufficient supply. Under the scenario
764 with sufficient supply, all customers can be served and introducing uncertainty into the model
765 generates solutions that outperform the nominal solution for all values of ρ and Γ .

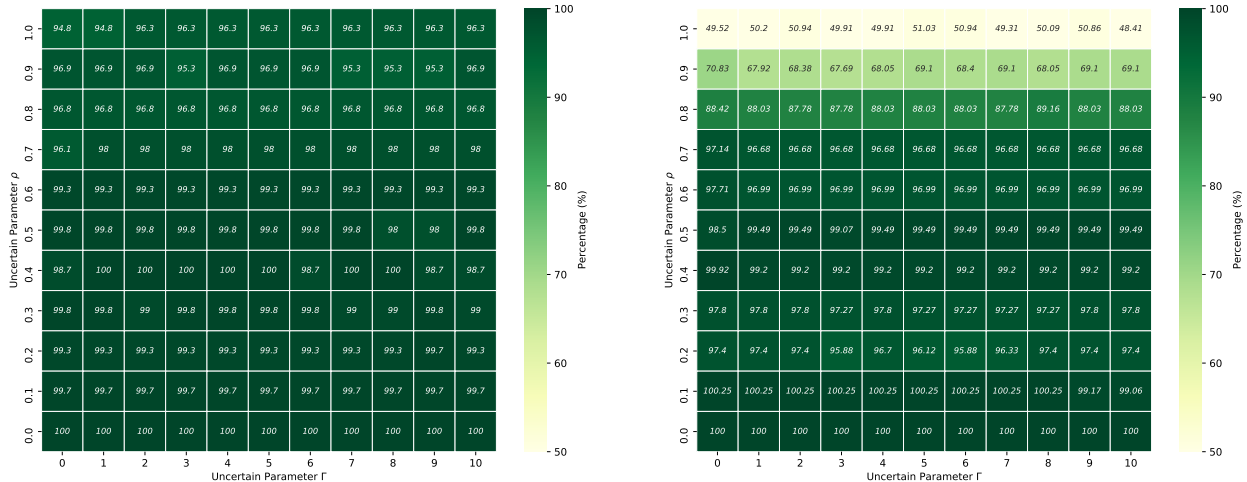
766 The robust MIVR model protects the rebalancing decisions against demand uncertainty
767 by restricting the number of rebalancing trips compared to the nominal MIVR model, which
768 is shown in Figure 15. When dispatching fewer vacant vehicles compared to the nominal
769 case, the penalty incurred due to inaccurate demand estimations is decreased and the system
770 becomes more robust against the demand uncertainty, hence has less total pickup time.

$\rho \backslash \Gamma$	0	1	2	3	4	5	6	7	8	9	10
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	5.23	4.19
0.2	6.4	6.4	6.4	7.44	7.01	7.27	6.67	6.45	6.4	6.4	6.4
0.3	12.51	12.51	12.51	12.0	12.51	12.0	12.0	12.0	12.0	12.51	12.51
0.4	16.23	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33
0.5	18.19	18.32	18.32	17.57	18.32	18.32	18.32	18.32	18.32	18.32	18.32
0.6	24.14	22.96	22.96	22.98	22.98	22.96	22.98	22.96	22.96	22.96	22.98
0.7	25.62	25.18	25.18	25.18	25.18	25.18	25.18	25.18	25.18	25.18	25.18
0.8	30.89	29.39	29.22	29.13	31.44	29.39	30.98	29.82	29.4	29.73	29.39
0.9	39.82	36.55	38.02	38.92	36.6	36.44	37.16	36.44	38.1	36.44	36.44
1.0	38.22	39.1	39.49	39.01	39.01	40.41	39.49	39.41	41.03	40.47	40.93

Table 4: Percentage reduction in the total pickup time compared to the nominal MIVR solution with sufficient supply ($N_v = 3000$), for different values of ρ and Γ .

771 The number of rebalancing trips is significantly restricted (less than 50% compared to the
772 nominal MIVR model) when introducing a high level of uncertainty into the robust MIVR
773 model under the sufficient supply scenario.

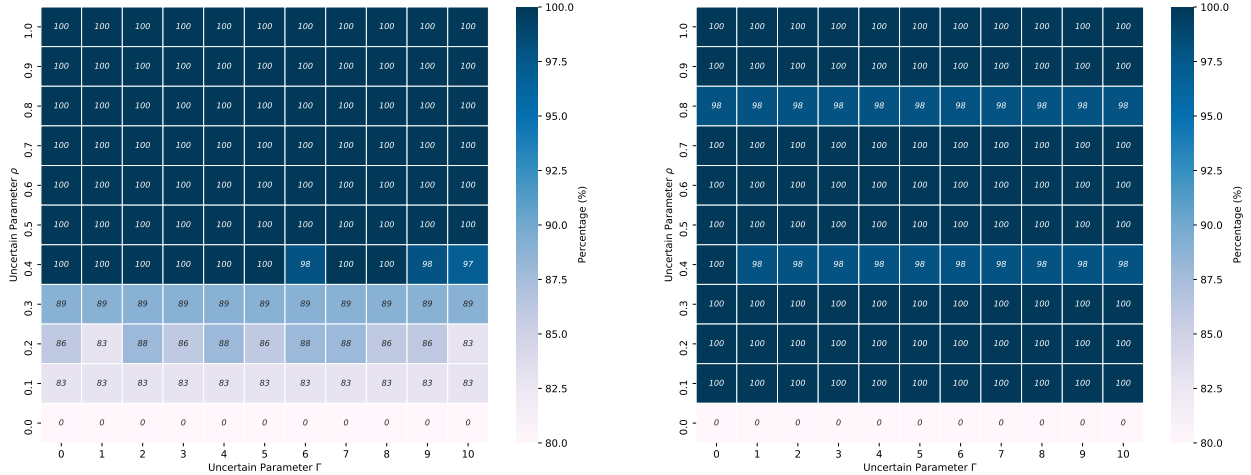
774 Figure 16 shows the daily performance of the robust MIVR model compared to the
775 nominal MIVR model. Under the insufficient supply scenario, even considering a low level
776 of uncertainty ($\rho = 0.1$) can significantly improve the performance of the robust MIVR
777 model (better performance than the nominal MIVR model for 83% of the 65 days tested).



(a) Insufficient supply scenario with fleet size $N_v = 2000$ (the nominal MIVR model conducts 864 vehicle rebalancing trips)

(b) Sufficient supply scenario with fleet size $N_v = 3000$ (the nominal MIVR model conducts 1228 vehicle rebalancing trips)

Figure 15: Rebalancing trips for the robust MIVR model under multiple uncertain scenarios. Each cell represents the percentage of rebalancing trips under a specific level of uncertainty compared to the number rebalancing trips in the nominal MIVR model.



(a) Insufficient supply scenario with fleet size $N_v = 2000$

(b) Sufficient supply scenario with fleet size $N_v = 3000$

Figure 16: Daily robust MIVR model performance under multiple uncertain scenarios. Each cell represents the percentage of the 65 input days that the robust MIVR model performs strictly better than the nominal MIVR model under a given level of uncertainty.

778 When incorporating a moderate level of uncertainty ($\rho \geq 0.5$) into the model, the robust
 779 MIVR model outperforms the nominal MIVR model for every day of demand tested. When
 780 a sufficient supply of vehicles is available, the robust MIVR model performs better than the
 781 nominal MIVR model for every weekday tested over most uncertain scenarios.

782 Overall, the robust MIVR model generates rebalancing decisions based on out-of-sample
 783 demand uncertainty defined by parameters ρ and Γ , and solutions are evaluated with real
 784 demand data reflecting in-sample demand uncertainty. The parameters ρ and Γ for uncer-
 785 tainty sets indicate the level of demand uncertainty that ride-hailing operators are willing to
 786 protect rebalancing decisions against. Based on experiment results, introducing robustness
 787 into the MIVR model and protecting rebalancing decisions against demand uncertainty im-
 788 prove the system performance effectively under insufficient and sufficient supply cases. The
 789 robust MIVR model performs even better when having sufficient number of vehicles in the
 790 system.

791 6. Conclusions and future work

792 In this paper, we formulate the MIVR model, which incorporates the driver-customer
 793 matching component into the consideration of vehicle rebalancing decisions made by ride-
 794 hailing operators, to protect rebalancing decisions against future demand uncertainty induced
 795 by inaccurate demand estimates. We evaluate the performance of our model by comparing
 796 against a benchmark VR model and a state-of-the-art model, named fluid-based empty-car
 797 routing policy (FERP), using actual ride-hailing trip data. Comparing to the VR model, the
 798 MIVR model reduces the average customer wait time and the total non-occupied VMT under
 799 most scenarios. When a large fleet is available, a *Pareto* improvement can be found regarding
 800 the overall non-occupied VMT, the average vehicle rebalancing trips, the average customer
 801 wait time and the number of unsatisfied requests. Comparing to the FERP, the MIVR model
 802 reduces the average customer wait time by generating a more proactive rebalancing strategy.

803 To further immunize solutions against demand uncertainty, we propose the robust MIVR
804 model by introducing RO techniques. The robust MIVR is especially effective when the
805 supply of ride-hailing vehicles is sufficient and most requests can be satisfied. Under both
806 sufficient-supply and insufficient-supply cases, the robust MIVR model prevents rebalancing
807 decisions from inaccurate demand estimation by rebalancing fewer vehicles. Additionally,
808 introducing robustness into the MIVR model generates rebalancing decisions that performs
809 better than decisions produced by the nominal MIVR model under most demand scenarios.

810 The main limitations of this study are a result of approximations embedded in the MIVR
811 model. First, we are only able to model trips aggregated to the zonal level given the data
812 availability. While we simulate actual pickups and drop-off locations within those zones,
813 future work could incorporate disaggregate data to test rebalancing and matching at the in-
814 dividual address level. The model could be improved if these data were made available. We
815 also assume static regional transition matrices estimated from the historical data. [Though](#)
816 [having limited impacts on model performances](#), matching and rebalancing decisions-based re-
817 gional transition matrices can be considered in the model to better reflect vehicle trajectories
818 across multiple time periods.

819 This paper shows how internalization of matching costs can be used to protect rebalancing
820 decisions against demand uncertainty and improve the efficiency of ride-hailing operations
821 regarding customers (satisfy more customers with shorter wait times), and under what condi-
822 tions the proposed method is beneficial. Furthermore, it illustrates how robust optimization
823 complements the MIVR model by further limiting the risk of increased cost due to incor-
824 rect demand estimations. Ride-hailing service operators should consider adopting the robust
825 MIVR model for improved customer outcomes, such as wait time and unsatisfied requests,
826 and reduced costs for operators.

827 There are several future research directions we identified in this paper. [First, the un-](#)
828 [certainty set \$\bar{U}^k\(\Gamma\)\$ has a limited impact on system performance. More effective and in-](#)
829 [terpretable uncertainty sets could be designed to model the uncertainty in the ride-hailing](#)
830 [system.](#) Secondly, additional uncertainty variables could be considered besides the demand
831 uncertainty, such as travel time. Thirdly, we used the historical average as the future de-
832 mand estimates in this paper. Advanced demand prediction algorithms can be incorporated
833 within the robust MIVR model to further improve operational performances. Lastly, the
834 MIVR model could be extended to solve the vehicle rebalancing problem in the shared MoD
835 system.

836 7. Acknowledgements

837 The authors would like to thank Qing Yi Wang and Nate Bailey for their comments on
838 the paper.

839 References

- 840 [1] S. Shaheen, A. Cohen, B. Yelchuru, S. Sarkhili, Mobility on demand operational concept
841 report (2017).
- 842 [2] D. Kerr, Lyft grows gangbusters in 2017, bringing competition to Uber, 2018.

- 843 [3] H. Wang, H. Yang, Ridesourcing systems: A framework and review, *Transportation*
844 *Research Part B: Methodological* 129 (2019) 122–155.
- 845 [4] N. Agatz, A. Erera, M. Savelsbergh, X. Wang, Optimization for dynamic ride-sharing:
846 A review, *European Journal of Operational Research* 223 (2012) 295–303.
- 847 [5] A. Wallar, M. Van Der Zee, J. Alonso-Mora, D. Rus, Vehicle rebalancing for mobility-
848 on-demand systems with ride-sharing, in: *2018 IEEE/RSJ International Conference on*
849 *Intelligent Robots and Systems (IROS)*, IEEE, pp. 4539–4546.
- 850 [6] J. Wen, J. Zhao, P. Jaillet, Rebalancing shared mobility-on-demand systems: A rein-
851 forcement learning approach, in: *2017 IEEE 20th International Conference on Intelligent*
852 *Transportation Systems (ITSC)*, IEEE, pp. 220–225.
- 853 [7] F. Miao, S. Han, A. M. Hendawi, M. E. Khalefa, J. A. Stankovic, G. J. Pappas, Data-
854 driven distributionally robust vehicle balancing using dynamic region partitions, in:
855 *Proceedings of the 8th International Conference on Cyber-Physical Systems - ICCPS*
856 *'17*, ACM Press, 2017, p. 261–271.
- 857 [8] K. Spieser, S. Samaranayake, W. Gruel, E. Frazzoli, Shared-vehicle mobility-on-demand
858 systems: a fleet operator’s guide to rebalancing empty vehicles, in: *Transportation*
859 *Research Board 95th Annual Meeting*, 16-5987, Transportation Research Board.
- 860 [9] R. Baldacci, V. Maniezzo, A. Mingozzi, An exact method for the car pooling problem
861 based on lagrangean column generation, *Operations Research* 52 (2004) 422–439.
- 862 [10] J. Alonso-Mora, S. Samaranayake, A. Wallar, E. Frazzoli, D. Rus, On-demand high-
863 capacity ride-sharing via dynamic trip-vehicle assignment, *Proceedings of the National*
864 *Academy of Sciences* 114 (2017) 462.
- 865 [11] Z. Xu, Z. Li, Q. Guan, D. Zhang, Q. Li, J. Nan, C. Liu, W. Bian, J. Ye, Large-scale
866 order dispatch in on-demand ride-hailing platforms: A learning and planning approach,
867 in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge*
868 *Discovery & Data Mining*, pp. 905–913.
- 869 [12] A. Mourad, J. Puchinger, C. Chu, A survey of models and algorithms for optimizing
870 shared mobility, *Transportation Research Part B: Methodological* 123 (2019) 323–346.
- 871 [13] S. C. Ho, W. Y. Szeto, Y.-H. Kuo, J. M. Leung, M. Petering, T. W. Tou, A survey
872 of dial-a-ride problems: Literature review and recent developments, *Transportation*
873 *Research Part B: Methodological* 111 (2018) 395–421.
- 874 [14] T. A. Taylor, On-demand service platforms, *Manufacturing & Service Operations*
875 *Management* 20 (2018) 704–720.
- 876 [15] J. Bai, K. C. So, C. S. Tang, X. M. Chen, H. Wang, Coordinating supply and demand
877 on an on-demand service platform with impatient customers, *Manufacturing & Service*
878 *Operations Management* 21 (2018) 556–570.

- 879 [16] G. P. Cachon, K. M. Daniels, R. Lobel, The role of surge pricing on a service platform
880 with self-scheduling capacity, *Manufacturing & Service Operations Management* 19
881 (2017) 368–384.
- 882 [17] J. Ke, H. Yang, X. Li, H. Wang, J. Ye, Pricing and equilibrium in on-demand ride-
883 pooling markets, *Transportation Research Part B: Methodological* 139 (2020) 411–431.
- 884 [18] A. A. Syed, I. Gaponova, K. Bogenberger, Neural network-based metaheuristic param-
885 eterization with application to the vehicle matching problem in ride-hailing services,
886 *Transportation Research Record* 2673 (2019) 311–320.
- 887 [19] M. Erdmann, F. Dandl, K. Bogenberger, Dynamic car-passenger matching based on
888 tabu search using global optimization with time windows, in: 2019 8th International
889 Conference on Modeling Simulation and Applied Optimization (ICMSAO), IEEE, pp.
890 1–5.
- 891 [20] H. Yang, X. Qin, J. Ke, J. Ye, Optimizing matching time interval and matching radius
892 in on-demand ride-sourcing markets, *Transportation Research Part B: Methodological*
893 131 (2020) 84–105.
- 894 [21] G. Lyu, W. C. Cheung, C.-P. Teo, H. Wang, Multi-objective online ride-matching,
895 Available at SSRN 3356823 (2019).
- 896 [22] R. Chen, M. W. Levin, Dynamic user equilibrium of mobility-on-demand system with
897 linear programming rebalancing strategy, *Transportation Research Record* 2673 (2019)
898 447–459.
- 899 [23] R. Zhang, F. Rossi, M. Pavone, Model predictive control of autonomous mobility-on-
900 demand systems, in: 2016 IEEE International Conference on Robotics and Automation
901 (ICRA), IEEE, pp. 1382–1389.
- 902 [24] R. Iglesias, F. Rossi, K. Wang, D. Hallac, J. Leskovec, M. Pavone, Data-driven model
903 predictive control of autonomous mobility-on-demand systems, arXiv:1709.07032 [cs,
904 stat] (2017). ArXiv: 1709.07032.
- 905 [25] A. Braverman, J. G. Dai, X. Liu, L. Ying, Empty-car routing in ridesharing systems,
906 *Operations Research* 67 (2019) 1437–1452.
- 907 [26] L. Al-Kanj, J. Nascimento, W. B. Powell, Approximate dynamic programming for
908 planning a ride-hailing system using autonomous fleets of electric vehicles, *European*
909 *Journal of Operational Research* 284 (2020) 1088–1106.
- 910 [27] F. Dandl, M. Hyland, K. Bogenberger, H. S. Mahmassani, Evaluating the impact of
911 spatio-temporal demand forecast aggregation on the operational performance of shared
912 autonomous mobility fleets, *Transportation* 46 (2019) 1975–1996.
- 913 [28] M. Guériau, I. Dusparic, SAMoD: Shared autonomous mobility-on-demand using de-
914 centralized reinforcement learning, in: 2018 21st International Conference on Intelligent
915 Transportation Systems (ITSC), IEEE, pp. 1558–1563.

- 916 [29] C. Yan, H. Zhu, N. Korolko, D. Woodard, Dynamic pricing and matching in ride-hailing
917 platforms, *Naval Research Logistics (NRL)* 67 (2020) 705–724.
- 918 [30] D. Z. Leon Yang Chu, Zhixi Wan, Harnessing the double-edged sword via routing:
919 Information provision on ride-hailing platforms, Available at SSRN 3266250 (2018).
- 920 [31] H. Yang, C. Shao, H. Wang, J. Ye, Integrated reward scheme and surge pricing in
921 a ridesourcing market, *Transportation Research Part B: Methodological* 134 (2020)
922 126–142.
- 923 [32] L. Zha, Y. Yin, H. Yang, Economic analysis of ride-sourcing markets, *Transportation*
924 *Research Part C: Emerging Technologies* 71 (2016) 249–266.
- 925 [33] I. Jindal, Z. T. Qin, X. Chen, M. Nokleby, J. Ye, Optimizing taxi carpool policies
926 via reinforcement learning and spatio-temporal mining, in: *2018 IEEE International*
927 *Conference on Big Data (Big Data)*, IEEE, pp. 1417–1426.
- 928 [34] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, *Robust Optimization*, Princeton Series in
929 *Applied Mathematics*, Princeton University Press, 2009.
- 930 [35] D. Bertsimas, D. B. Brown, C. Caramanis, Theory and applications of robust optimiza-
931 tion, *SIAM Review* 53 (2011) 464–501.
- 932 [36] A. Ben-Tal, A. Nemirovski, Robust convex optimization, *Mathematics of Operations*
933 *Research* 23 (1998) 769–805.
- 934 [37] A. Ben-Tal, A. Nemirovski, Robust solutions of uncertain linear programs, *Operations*
935 *Research Letters* 25 (1999) 1–13.
- 936 [38] D. Bertsimas, M. Sim, The price of robustness, *Operations Research* 52 (2004) 35–53.
- 937 [39] P. Xiong, P. Jirutitijaroen, C. Singh, A distributionally robust optimization model for
938 unit commitment considering uncertain wind power generation, *IEEE Transactions on*
939 *Power Systems* 32 (2016) 39–49.
- 940 [40] C. Ma, W. Hao, R. He, X. Jia, F. Pan, J. Fan, R. Xiong, Distribution path robust
941 optimization of electric vehicle with multiple distribution centers, *PloS One* 13 (2018).
- 942 [41] Y. Wang, Y. Zhang, J. Tang, A distributionally robust optimization approach for
943 surgery block allocation, *European Journal of Operational Research* 273 (2019) 740–
944 753.
- 945 [42] Y. Liu, W. Skinner, C. Xiang, Globally-optimized realtime supply-demand matching in
946 on-demand ridesharing, in: *The World Wide Web Conference*, pp. 3034–3040.
- 947 [43] F. Miao, S. Han, S. Lin, Q. Wang, J. A. Stankovic, A. Hendawi, D. Zhang, T. He,
948 G. J. Pappas, Data-driven robust taxi dispatch under demand uncertainties, *IEEE*
949 *Transactions on Control Systems Technology* 27 (2017) 175–191.

- 950 [44] D. Bertsimas, V. Gupta, N. Kallus, Data-driven robust optimization, *Mathematical*
951 *Programming* 167 (2018) 235–292.
- 952 [45] L. He, Z. Hu, M. Zhang, Robust repositioning for vehicle sharing, *Manufacturing &*
953 *Service Operations Management* (2019).
- 954 [46] D. Bertsimas, M. Sim, M. Zhang, Adaptive distributionally robust optimization, *Man-*
955 *agement Science* 65 (2019) 604–618.
- 956 [47] B. L. Gorissen, I. Yanıkoğlu, D. d. Hertog, A practical guide to robust optimization,
957 *Omega* 53 (2015) 124–137. ArXiv: 1501.02634.
- 958 [48] D. Bertsimas, D. den Hertog, Robust and adaptive optimization, Dynamic Ideas LLC,
959 Belmont, Massachusetts, 2020.
- 960 [49] NYC Taxi and Limousine Commission, TLC Trip Record Data, 2019. "[Online; accessed
961 15-June-2020]".
- 962 [50] Uber Technologies, Inc., Uber Movement, New York City travel speeds, 2019. "[Online;
963 accessed 15-June-2020]".
- 964 [51] E. W. Dijkstra, et al., A note on two problems in connexion with graphs, *Numerische*
965 *Mathematik* 1 (1959) 269–271.
- 966 [52] Gurobi Optimization, LLC, Gurobi optimizer reference manual, 2020.

967 **Appendix A. Derivation of The Robust Counterpart**

968 Given the following generic constraint

$$L(\cdot) + v^T \zeta \leq c \quad \forall \zeta \in \mathcal{U}, \quad (\text{A.1})$$

969 where $L(\cdot)$ indicates a function of decision variables in problem (P') , v is a vector in
970 dimension $n\kappa$ and c is a scalar, it is equivalent to

$$L(\cdot) + \max_{\zeta \in \mathcal{U}} v^T \zeta \leq c. \quad (\text{A.2})$$

971 By taking the convex conjugate of constraint (A.2) we derive the following equivalent
972 constraint

$$L(\cdot) + \delta^*(v \mid \mathcal{U}) \leq c, \quad (\text{A.3})$$

973 where $\delta(v \mid \mathcal{U})$ is an indicator function such that $\delta(v \mid \mathcal{U}) = 0$ if $v \in \mathcal{U}$, otherwise
974 $\delta(v \mid \mathcal{U}) = \infty$. $\delta^*(v \mid \mathcal{U})$ is the convex conjugate of $\delta(v \mid \mathcal{U})$. Then we introduce Lemma 1 to
975 help with deriving the robust counterpart [48].

Lemma 1. For a constraint $\bar{a}^T x + \delta^*(P^T x \mid Z) \leq b$, let Z_1, \dots, Z_k be closed convex sets, such that $\bigcap_i \text{ri}(Z_i) \neq \emptyset^7$, and let $Z = \bigcap_{i=1}^k Z_i$. Then,

$$\delta^*(y \mid Z) = \min_{y^1, \dots, y^k} \left\{ \sum_{i=1}^k \delta^*(y^i \mid Z_i) \mid \sum_{i=1}^k y^i = y \right\},$$

and the constraint becomes

$$\begin{cases} \bar{a}^T x + \sum_{i=1}^k \delta^*(y^i \mid Z_i) \leq b \\ \sum_{i=1}^k y^i = P^T x \end{cases}$$

976 Let $\mathcal{U}_0 = \{\zeta : \|\zeta\|_\infty \leq \rho\}$ and $\mathcal{U}_k = \{\zeta : |e^T(\zeta \circ \Sigma^k)| \leq \Gamma\}, \forall k \in K$, where $\Sigma^k \in \mathbb{R}^{n\kappa}$
977 denotes a vector with (ik) -th entry equals to $\sigma_i^k, \forall i \in N$, and other entries equal to zero.
978 The uncertainty set \mathcal{U} can be written as: $\mathcal{U} = \bigcap_{k=0}^\kappa \mathcal{U}_k$. By applying Lemma 1 to constraint
979 (A.3), we develop the following robust counterpart for constraint (A.1):

$$\begin{cases} L(\cdot) + \sum_{k=0}^\kappa \delta^*(\theta_k \mid \mathcal{U}_k) \leq c \\ \sum_{k=0}^\kappa \theta_k = v \end{cases} \quad (\text{A.4})$$

⁷ $\text{ri}(Z_i)$ indicates the relative interior of the set Z_i .

980

Which is equivalent to

$$\begin{cases} L(\cdot) + \rho \|\theta_0\|_1 + \Gamma \sum_{k=1}^{\kappa} (\eta_1^k + \eta_2^k) \leq c \\ (\eta_1^{k'} - \eta_2^{k'}) \sigma_i^{k'} = \theta_{k'}^{i,k} \quad \forall i \in N, \forall k = k' \in K \\ \theta_{k'}^{i,k} = 0 \quad \forall i \in N, \forall k \neq k' \in K \\ \eta_1^k, \eta_2^k \geq 0 \quad \forall k \in K \\ \sum_{k=0}^{\kappa} \theta_k = v \end{cases} \quad (\text{A.5})$$

981

Where $\theta_k \in \mathbb{R}^{n\kappa}$ and $\theta_{k'}^{i,k}$ represents (ik) -th entry of vector $\theta_{k'}$, $\forall k' \in K$.

982

Appendix B. Benchmark Vehicle Rebalancing (VR) Model

983

984

985

986

987

988

989

990

In this section, we formulate a benchmark vehicle rebalancing (VR) model to test the performance of our MIVR model. With similar notations to the MIVR model, we introduce several additional parameters. Let P_v^k, Q_v^k be regional transition matrices regarding vacant vehicles in time period k , which are learned from the historical data. $P_{v,ij}^k$ stands for the probability for a vacant vehicle in sub-region i at time k to be in sub-region j at time $k+1$ and becomes occupied. Similarly, $Q_{v,ij}^k$ denotes the probability for a vacant vehicle in sub-region i at time k to be in sub-region j at time $k+1$ and remains vacant. Two regional transition matrices satisfy the following condition:

$$\sum_{j=1}^n (P_{v,ij}^k + Q_{v,ij}^k) = 1, \quad \forall i \in N, \forall k \in K.$$

991

992

Then the benchmark VR model is:

$$(VR) \quad \min_{x_{ij}^k} \sum_{k=1}^{\kappa} \sum_{i=1}^n \sum_{j=1}^n x_{ij}^k d_{ij}^k + \alpha \cdot \sum_{k=1}^{\kappa} \sum_{i=1}^n |S_i^k - r_i^k| \quad (\text{B.1a})$$

$$\text{s.t.} \quad S_i^k = \sum_{j=1}^n x_{ji}^k - \sum_{j=1}^n x_{ij}^k + V_i^k \quad \forall i \in N, \forall k \in K \quad (\text{B.1b})$$

$$V_i^{k+1} = \sum_{j=1}^n Q_{v,ji}^k S_j^k + \sum_{j=1}^n Q_{ji}^k O_j^k \quad \forall i \in N, \forall k \in K \setminus \{\kappa\} \quad (\text{B.1c})$$

$$O_i^{k+1} = \sum_{j=1}^n P_{v,ji}^k S_j^k + \sum_{j=1}^n P_{ji}^k O_j^k \quad \forall i \in N, \forall k \in K \setminus \{\kappa\} \quad (\text{B.1d})$$

$$\sum_{j=1}^n x_{ij}^k \leq V_i^k \quad \forall i \in N, \forall k \in K \quad (\text{B.1e})$$

$$a_{ij}^k \cdot x_{ij}^k = 0 \quad \forall i \in N, \forall k \in K \quad (\text{B.1f})$$

$$x_{ij}^k \in \mathbb{R}^+ \quad \forall i, j \in N, \forall k \in K \quad (\text{B.1g})$$

$$S_i^k, V_i^k, O_i^k \in \mathbb{R}^+ \quad \forall i \in N, \forall k \in K \quad (\text{B.1h})$$

993 Where the objective function (B.1a) consists of vehicle rebalancing cost and a service
 994 availability function with a weight parameter α to minimize the difference between available
 995 vehicles and estimated demand in each sub-region. Constraints (B.1b) to (B.1d) define the
 996 relationship between available vehicles S_i^k , vacant vehicles V_i^k and occupied vehicles O_i^k . The
 997 maximum number of available vehicles that can be rebalanced is restricted by constraints
 998 (B.1e). Constraints (B.1f) impose the feasibility restrictions for rebalancing decisions, and
 999 the non-negativity of integer decision variables are guaranteed by constraints (B.1g) and
 1000 (B.1h). To increase the computational efficiency while maintaining a satisfying solution, we
 1001 further relax integer decision variables x_{ij}^k, S_i^k, V_i^k and O_i^k to positive real numbers \mathbb{R}^+ .

1002 The VR model proposed in this section is sufficient to show the benefit of integrating
 1003 matching into the VR problem. When having different VR models with the area partitioning
 1004 assumption, a matching-integrated version can always be constructed.

1005 Appendix C. Optimal Assignment of Drivers to Customers

1006 In this section, the driver-customer assignment problem implemented in the matching
 1007 engine of the simulator is described. Within each matching decision time interval δ , let
 1008 $\mathcal{R} = \{r_1, \dots, r_n\}$ denote a set of waiting customers and $\mathcal{V} = \{v_1, \dots, v_m\}$ represent a set of
 1009 vacant vehicles in the system. Between a customer r_i and a vehicle v_j , let $\tau(r_i, v_j)$ indicate
 1010 the minimum travel time for the vehicle to pick up the customer. The maximum pickup
 1011 time for customers is denoted by \bar{w} . First, we construct a bipartite graph $G = (V, E)$,
 1012 where $V = \mathcal{R} \cup \mathcal{V}$ and $E = \{e(r_i, v_j) : \forall r_i \in \mathcal{R}, \forall v_j \in \mathcal{V}, \tau(r_i, v_j) \leq \bar{w}\}$, meaning that
 1013 an edge exists between a vehicle and a customer if the customer can be picked up by the
 1014 vehicle within the maximum pickup time. The cost of each edge $e(r_i, v_j)$ equals to the pickup
 1015 time, i.e., $c_{e(r_i, v_j)} = \tau(r_i, v_j)$. The decision variables for the optimal assignment problem are
 1016 $x_{e(r_i, v_j)} \in \{0, 1\}$ for each edge $e(r_i, v_j) \in E$ in the bipartite graph G , and $y_{r_i} \in \{0, 1\}$ for each
 1017 customer $r_i \in \mathcal{R}$. $x_{e(r_i, v_j)} = 1$ indicates that the customer r_i will be picked up by the vehicle
 1018 v_j in the optimal assignment. $y_{r_i} = 1$ implies that the customer r_i will not be assigned
 1019 to any vehicles during the current decision time interval δ . Let $\mathcal{I}(r_i)$ represent the set of
 1020 edges connected to a customer vertex r_i in G . Similarly, let $\mathcal{I}(v_j)$ indicate the set of edges
 1021 connected to a driver vertex v_j in G . The optimal driver-customer assignment problem is:

$$\min \sum_{e(r_i, v_j) \in E} c_{e(r_i, v_j)} x_{e(r_i, v_j)} + \gamma \cdot \sum_{r_i \in \mathcal{R}} y_{r_i} \quad (\text{C.1a})$$

$$\text{s.t.} \quad \sum_{e(r_i, v_j) \in \mathcal{I}(v_j)} x_{e(r_i, v_j)} \leq 1 \quad \forall v_j \in \mathcal{V} \quad (\text{C.1b})$$

$$\sum_{e(r_i, v_j) \in \mathcal{I}(r_i)} x_{e(r_i, v_j)} + y_{r_i} = 1 \quad \forall r_i \in \mathcal{R} \quad (\text{C.1c})$$

$$x_{e(r_i, v_j)} \in \{0, 1\} \quad \forall e(r_i, v_j) \in E \quad (\text{C.1d})$$

$$y_{r_i} \in \{0, 1\} \quad \forall r_i \in \mathcal{R} \quad (\text{C.1e})$$

1022 The objective function (C.1a) minimizes the summation of the total pickup time and
 1023 penalties for unsatisfied requests, where γ stands for the penalty VMT for each unsatisfied
 1024 customer. Constraints (C.1b) ensure that each vehicle can only be assigned to at most one

1025 customer. Constraints (C.1c) guarantee that each customer is either served by a vehicle or
1026 remained waiting during the current matching period. Constraints (C.1d) and (C.1e) make
1027 sure that the decision variables are binary. The optimal driver-customer assignment problem
1028 can be solved efficiently by the off-the-shelf ILP solvers (e.g., Gurobi) in the simulation.

1029 **Appendix D. Estimation of Regional Transition Matrix**

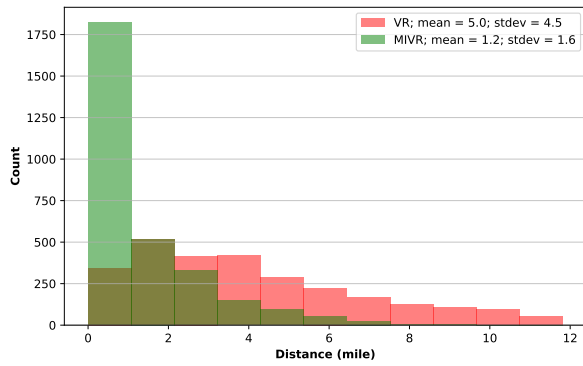
1030 In this section, the process for estimating the regional transition probability matrices for
1031 occupied and vacant vehicles, P , Q , P_v and Q_v , with the real travel time and demand data
1032 are described. There are several assumptions we made to generate these matrices:

- 1033 • Given a travel time and distance between the origin and the destination of a request,
1034 the vehicle travels with a constant speed.
- 1035 • Given the origin and the destination of a request, the vehicle travels along the shortest
1036 path with regards to travel time.
- 1037 • For vacant vehicles within sub-regions, 100% of vehicles remain in the same sub-region.

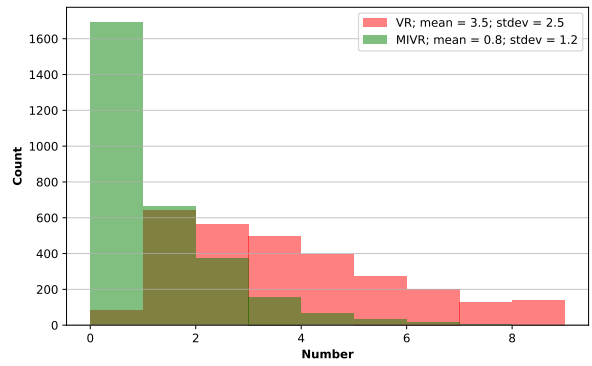
1038 The detailed procedure is described as follows. First, the list of sub-regions crossed by
1039 the shortest path between each origin and destination pair was determined. The time spent
1040 within each sub-region for each origin-destination pair was weighted by the total demand
1041 to get the average time spent in each sub-region across all trips. For a given starting sub-
1042 region, the interzonal shortest paths, sub-region durations and origin-destination demand
1043 patterns were used to determine the likelihood of a given vehicle remaining in the starting
1044 sub-region, transitioning to a nearby sub-region or making a dropoff within a time interval.
1045 These probabilities were then used to populate P and Q . Because the taxi dataset only
1046 contains information about occupied vehicles, assumptions were made for the vacant vehicle
1047 zone transition probability matrices P_v and Q_v .

1048 **Appendix E. Benchmark VR Comparison Results for Different Demand Sce-** 1049 **narios**

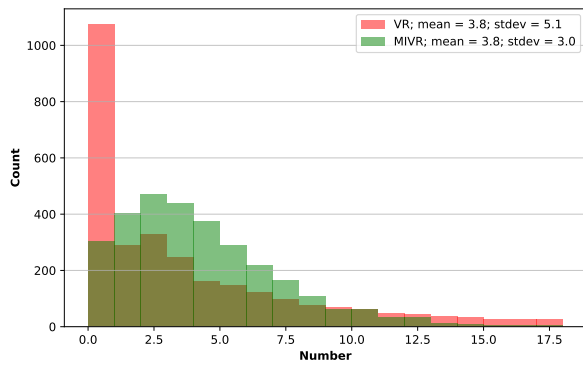
1050 In this section, we provide the base case simulation results for four different demand
1051 scenarios: low demand with accurate estimation in Figure E.17, high demand with accurate
1052 estimation in Figure E.18, demand underestimation in Figure E.19 and demand overestima-
1053 tion in Figure E.20.



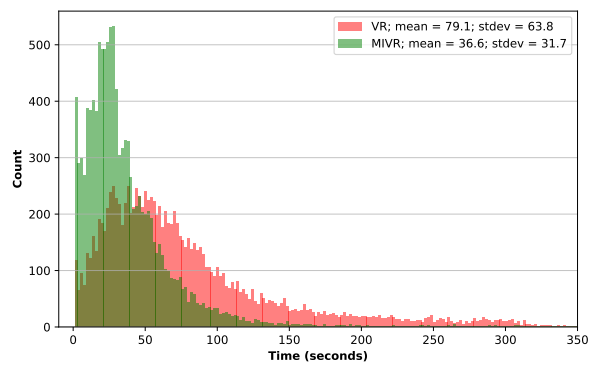
(a) Vehicle non-occupied travel distance distribution



(b) Vehicle rebalancing trip distribution

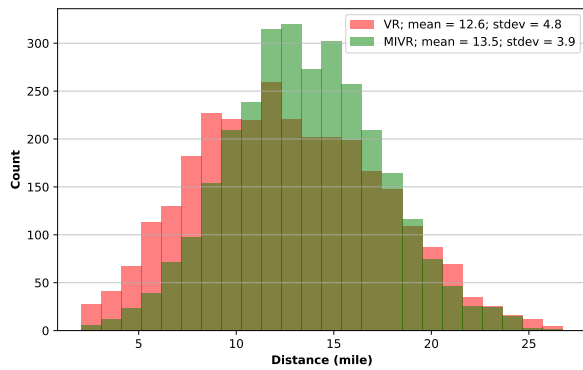


(c) Number of customers served (per vehicle) distribution

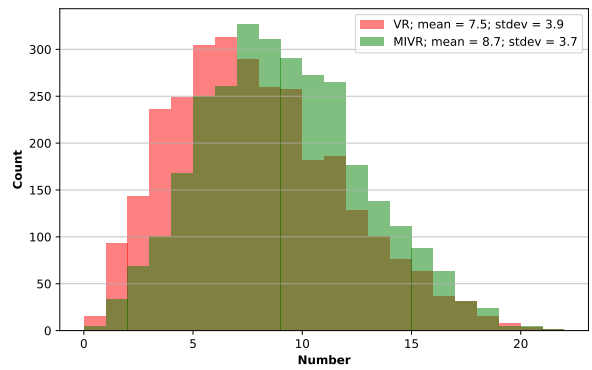


(d) Customer wait time distribution

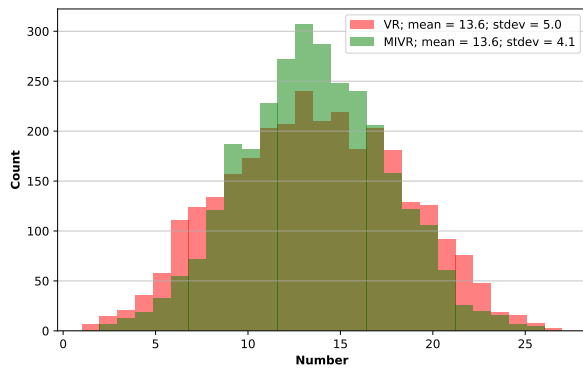
Figure E.17: Vehicle- and customer-related metrics in the simulation for the base case under the low demand with accurate estimation scenario (0 - 6).



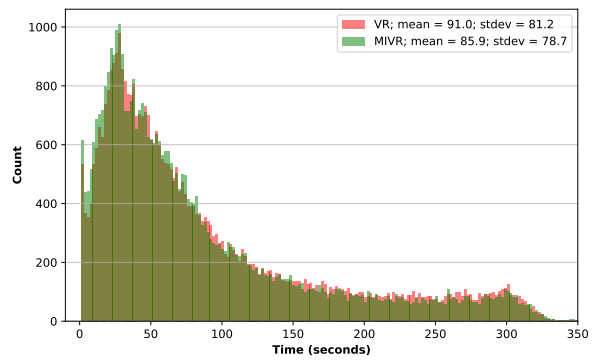
(a) Vehicle non-occupied travel distance distribution



(b) Vehicle rebalancing trip distribution

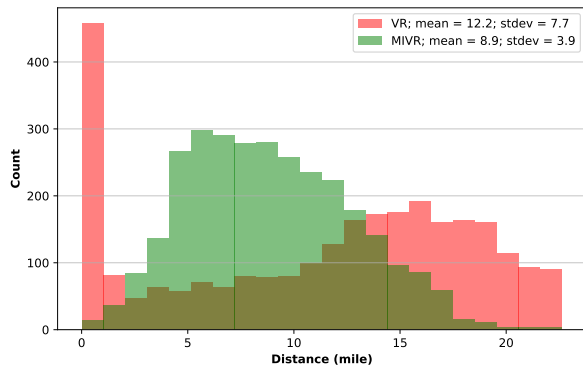


(c) Number of customers served (per vehicle) distribution

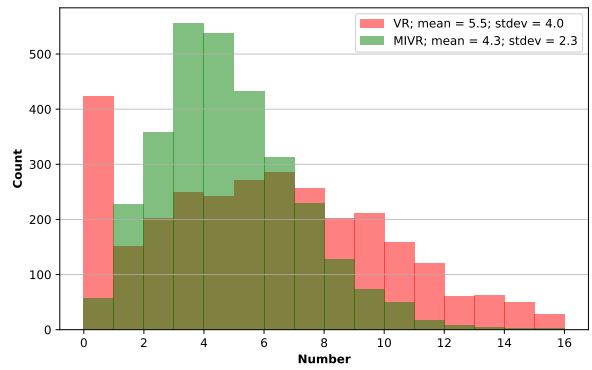


(d) Customer wait time distribution

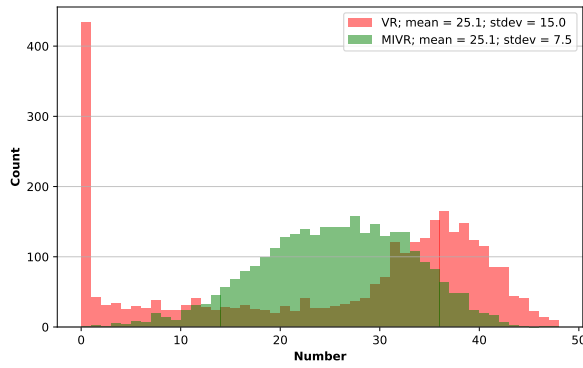
Figure E.18: Vehicle- and customer-related metrics in the simulation for the base case under the high demand with accurate estimation scenario (6 - 10).



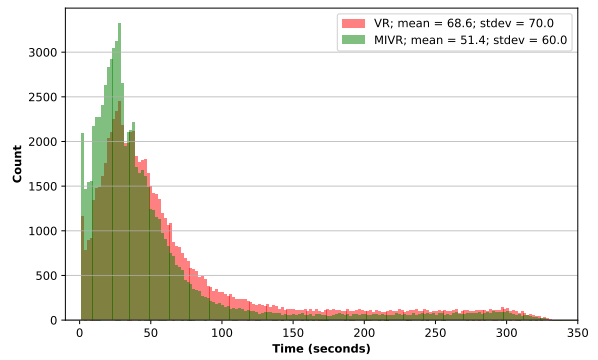
(a) Vehicle non-occupied travel distance distribution



(b) Vehicle rebalancing trip distribution

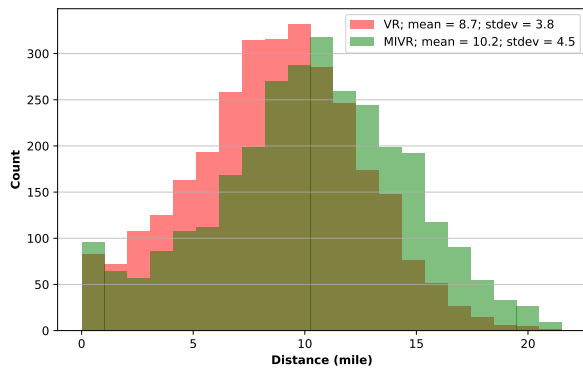


(c) Number of customers served (per vehicle) distribution

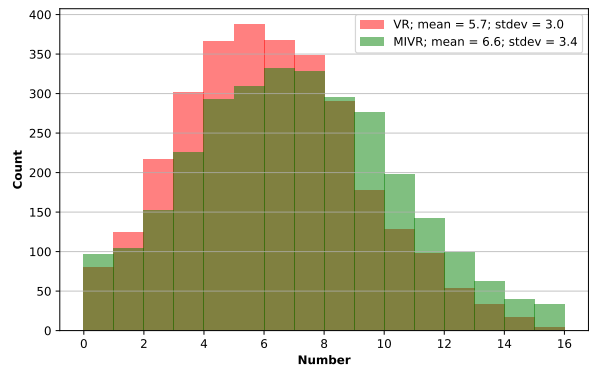


(d) Customer wait time distribution

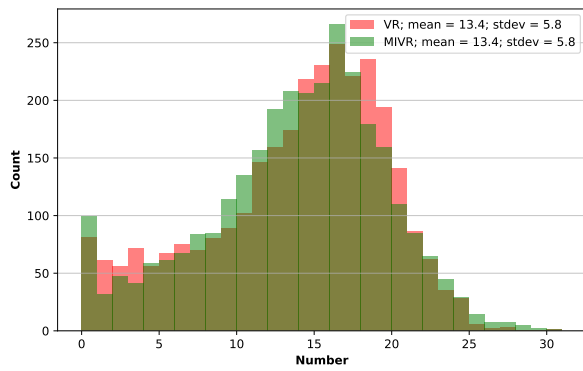
Figure E.19: Vehicle- and customer-related metrics in the simulation for the base case under demand under-estimation scenario (11 - 17).



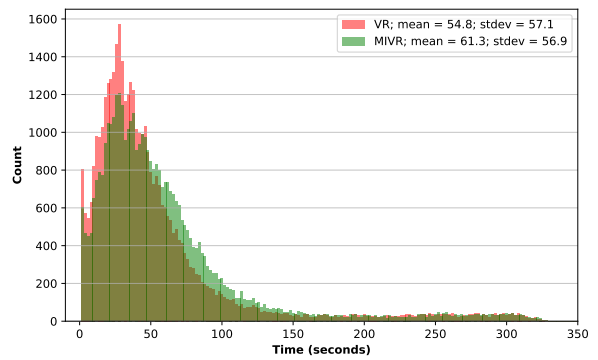
(a) Vehicle non-occupied travel distance distribution



(b) Vehicle rebalancing trip distribution



(c) Number of customers served (per vehicle) distribution



(d) Customer wait time distribution

Figure E.20: Vehicle- and customer-related metrics in the simulation for the base case under demand over-estimation scenario (20 - 24).