



Joint infrastructure planning and fleet management for one-way electric car sharing under time-varying uncertain demand

Yikang Hua^a, Dongfang Zhao^b, Xin Wang^{c,*}, Xiaopeng Li^b

^a Department of Industrial and Systems Engineering, University of Wisconsin-Madison, United States

^b Department of Civil and Environmental Engineering, University of South Florida, United States

^c Department of Industrial and Systems Engineering & Grainger Institute for Engineering, University of Wisconsin-Madison, United States



ARTICLE INFO

Article history:

Received 20 August 2018

Revised 26 May 2019

Accepted 13 July 2019

Keywords:

Electric vehicle sharing

Infrastructure planning

Dynamic system,

ABSTRACT

We propose an innovative framework to deploy a one-way Electric Vehicle (EV) sharing system that serves an urban area. For the first time, long-term infrastructure planning (charging station location and fleet distribution) and real-time fleet operations (relocation and charging decisions) are jointly optimized under time-varying uncertain demand. This substantially advances EV sharing system efficiency and yields a practical management strategy. We propose a multistage stochastic model to address the critical challenge of time-varying uncertain demand. An accelerated solution algorithm is developed to conquer the curse of dimensionality in integer infrastructure planning decisions. Meaningful insights are delivered through hypothetical numerical experiments and a realistic case study with EV sharing service in the New York City.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

The overwhelming growth of private automobile ownership exacerbates social and environmental issues such as rush-hour traffic congestion, lacking of parking spaces, and greenhouse gas emissions. Many efforts have been made toward reducing the adoption of private vehicles. In particular, noticing that private vehicles are usually not utilized efficiently, people proposed the concept of vehicle sharing as an innovative trade-off between convenience and efficiency. Zipcar, for example, is a car sharing company that provides their members with accesses to a fleet of cars across a service area. Members can make one-way reservations (picking up a car at an origin and returning it at a destination), paying this service by time of usage. In such a car sharing service, due to the high utilization of each car, the travel cost is dramatically reduced. Furthermore, with the prosperous car sharing services nowadays, some travelers have already abandoned their private cars and complete their daily trips through car sharing services instead (Millard-Ball, 2005).

Motivated by the environmental benefit, electric vehicles (EV) based vehicle sharing services have been commercialized and attracted lots of attention. A life cycle analysis shows that, compared with the traditional internal combustion engine vehicles, EVs offer a great potential for reducing emissions and decreasing unit-distance travel cost (Wang et al., 2005). Due to environmental concerns, government agencies gradually strengthen the vehicle emission standards nowadays, which pushes automotive manufacturers to produce more and more EVs. However, private EV users face travel range and

* Corresponding author.

E-mail address: xin.wang@wisc.edu (X. Wang).

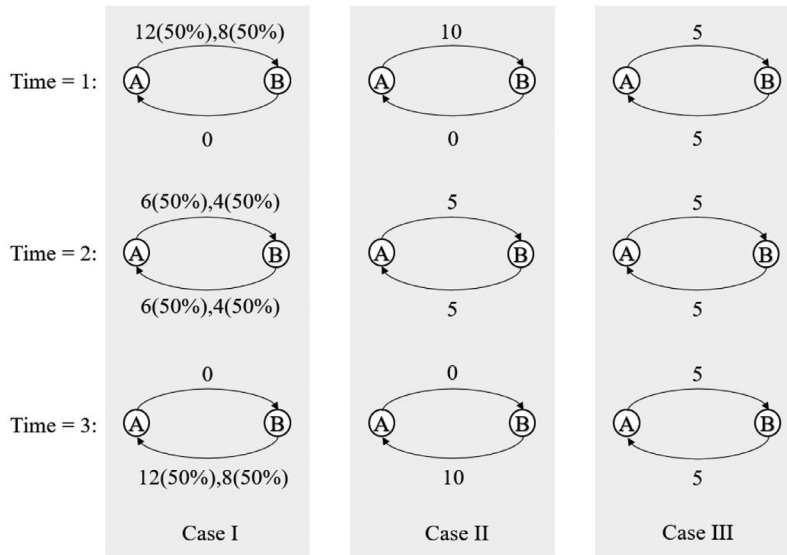


Fig. 1. Demand profiles of the two-nodes EV sharing problem, where the number of customer travel requests within each time period is indicated at the top/bottom the corresponding arrow. Case I: time-varying uncertain demand (e.g., profile 12 (50%), 8 (50%) indicates the demand is either 12 or 8 with equal chances); Case II: time-varying deterministic demand; Case III: time-invariant deterministic demand.

resell value anxieties (Lim et al., 2014), which significantly limits the EV adoption. On the contrary, EV sharing services can pass through such anxieties from private users upward to a company level, where mitigation strategies can be implemented through well-coordinated fleet management and economies of scale. One example is Car2Go, a subsidiary company of Daimler AG (a large automotive corporation owning Mercedes-Benz and Smart Automobile), which provides one-way free floating EV rentals. Another example is the BlueIndy services in Indianapolis, which applies exclusive electric vehicles to station-based point-to-point car sharing services (Voelcker, 2016). It is foreseen that, as the manufacturing cost decreases and infrastructure develops, EVs may replace conventional vehicles in the near future, and EV sharing services also become more attractive and affordable than traditional ones in the long term.

However, many challenges are raised by EV sharing services. The first one is the vehicle relocation, which is the common headache among all one-way vehicle sharing services. For example, people tend to drive from the suburb to the downtown in the morning and return in the afternoon, which indicates a huge demand imbalance. Without effective vehicle relocation operations, customers may fail to get or return a vehicle due to limited fleet size or capacity of stations. Furthermore, different from traditional vehicle sharing, EVs are with frequent charging needs and much slower charging speed, which brings battery charging management critical (Luè et al., 2012). In particular, at the early stage of EV adoption, lack of charging infrastructure are blamed to be the reason to block EV sharing services. For example, Car2Go claimed their EV sharing services in San Diego were heavily affected by lack of charging stations, hence they have to replace all EVs with gasoline-powered cars and further stopped the vehicle sharing service in San Diego in 2016 (Garrick, 2016). At present, Car2Go company insists on applying EVs in only two cities (Amsterdam in Netherlands and Stuttgart in Germany) among all 25 active service regions. Many evidences have shown that to maintain a good service quality, EV sharing system requires to coordinate fleet operations based on limited infrastructure availability.

Another critical challenge for EV sharing is the highly uncertain and time-varying travel demand. To obtain an intuition about the impact of time-varying uncertain demand on an EV system design, we consider a simple two-node example, where customers travel between locations A and B. The actual demand is uncertain over three time periods (e.g., morning, afternoon, evening), as shown in Case I of Fig. 1. If the uncertainty and time-varying features of demands are both ignored, we simply obtain a symmetric demand profile in Case III. With a certain parameter setting,¹ we should build charging stations in both A and B and equally allocate 10 EVs for each location. However, if we notice that the demand is actually time varying, which is given as Case II, there is no need to build charging station at A but should redistribute the EV fleet, 15 EVs at A and 5 EVs at B, to serve the peak demand at time 1. By doing this, the infrastructure cost is saved without sacrificing the service. Moreover, if both the uncertainty and time-varying features are captured (Case I), we can further save some EVs by wisely deploying them initially, say 16 EVs at A only. As more features are captured, the EV sharing business can become much more efficient (with the same demand profile as in Case I, the realized costs of the optimal solution at each case are Case I: 1040, Case II: 1260, and Case III: 1780).

¹ Assuming that each EV can just support one-way trip between the two locations with a full state of charge (SOC), which takes one unit time for travel. It also takes one unit time to charge the battery from empty to full SOC. We assume all customers are patient and will wait for the service. The unit holding cost for each EV is 50, the unit waiting cost for each customer is 20, and the penalty cost for each unsatisfied demand in the end is 100.

Although capturing time-varying uncertain demand is critical, it is extremely difficult due to the curse of dimensionality, despite some limited efforts (Li et al., 2016). To capture the EV fleet dynamics, the model needs to cover the dimension of location, time, and even the battery level. A direct way is to construct multiple stages to allow the stochastic parameters changing over multiple time periods. Then the tractability of the model becomes a well-known issue, especially when integer decisions are involved such as charging station location.

Noticing the above challenges, we establish a holistic framework to address infrastructure planning in a one-way EV sharing system with corresponding real-time fleet management policy. Given the probability distribution of time-varying demand profiles, we fill the gap to optimize EV sharing services through a multi-stage spatial-temporal network. The ultimate objective of the model is to find the best trade-off between the infrastructure fixed cost, fleet deployment and relocation cost, as well as a satisfactory service level. Our contribution is summarized as follows. (i) For the first time, we are able to address the joint infrastructure planning and daily operations decisions, which significantly improves the EV sharing efficiency; (ii) A multi-stage stochastic model is established, which captures the critical issue of time-varying uncertain demand, and it further allows a better utilization of historical customer demand information; (iii) We provide a heuristic algorithm with acceptable bounded errors to conquer the curse of dimensionality in multi-stage stochastic models with integer decisions, which can be further generalized to other related work involving time-varying uncertainties. (iv) A series of numerical experiments are conducted through both hypothetical and realistic case studies to show the performance of our algorithms as well as providing managerial insights.

The remainder of this paper is organized as follows. Section 2 provides a literature review for existing vehicle sharing work. In Section 3, we develop a full version of the scenario-based multistage stochastic model to address the joint EV fleet infrastructure deployment and vehicle relocation decision. Then in Section 4, we provide a solution based on Lagrangian relaxation and Stochastic Dual Dynamic Programming (SDDP) method. A series of numerical experiments and a real case optimal system design is presented in Section 5. Finally, Section 6 concludes our work.

2. Literature review

The core of our proposed research is infrastructure deployment problem in the field of EV sharing, which has been widely explored in other similar fields. Early studies attempted to determine the optimal locations to capture more customer flows with less facilities (Hodgson, 1990; Berman et al., 1992). Further studies extended the location design to consider the spatial correlations, such as the joint location-flow network design to optimally choose both the facility locations and the links between them (Melkote and Daskin, 2001) and the joint location-inventory problem where adjacent facilities can exhibit somewhat pooling effect to reduce the safety stock level (Shen et al., 2003; Yao et al., 2010). In addition to location design, determining the fleet size is another important factor in the infrastructure deployment problem. Wong et al. (2001) developed a two-level network equilibrium model of urban taxi services which explicitly offered some insights regarding the effects of fleet size on the system performance.

As the concept of car sharing becomes more and more popular, some mathematical models are developed to solve problems that have appeared in practical car sharing programs, such as vehicle relocation, location and fleet size problems. George and Xia (2011) formulated a closed queuing network model of vehicle rental systems to determine the optimal fleet size for a vehicle rental company. Nair and Miller-Hooks (2011) developed a mixed-integer program with joint chance constraints which generates least-cost relocation plans for the purpose of meeting a proportion of all near-term asymmetric demand scenarios. de Almeida Correia and Antunes (2012) presented an optimization approach based on mixed-integer programming to locate depot under certain vehicle allocation strategy to avoid one-way car sharing imbalance. Barrios and Godier (2014) built an agent-based model of flexible car-sharing system to explore the relation between fleet size and relocation. Correia et al. (2014) extended the model for trip selection and station location by considering more flexible choices of customers.

Further, as EV receives more attention, the strategic optimization problems in electric mobility systems have been widely investigated to optimize decision variables including station location, fleet size, parking capacity and etc. Frade et al. (2011) presented a maximal covering model to define the number and capacity of the charging stations while ensuring an acceptable level of service, and showed the usefulness of this methodology in areas characterized with a strong concentration of population and employment. Worley et al. (2012) combined the problem of locating charging stations with that of designing EV routes based on the vehicle routing problem to overcome the challenge of long-distance travel demand. Mak et al. (2013) developed robust optimization models to aid the planning process for deploying battery-swapping infrastructure to support the transition from battery charging to battery swapping, which may be a potential solution to deal with the limited capacity of batteries. Bruglieri et al. (2014) exploited the relocation strategy in one-way EV sharing systems. In the paper, a mixed integer linear programming (MILP) model is formulated to forecast the unbalancing of a car-sharing system and generate the relocation requests of EVs with the objective of maximizing the total number of requests served. The model is then tested in the Milan area and it is shown in the result that 24% of the driving time has been reduced on average. Further, Boyacı et al. (2015) developed a multi-objective MILP model for planning one-way vehicle sharing systems considering both vehicle relocation and EV charging requirement based on a scenario-based stochastic demand and solved the model using an aggregation method. The objective function maximizes the net revenue for the operator as well as user's net benefits. Different from their simplified model setting where EVs are charged for a fixed time period after each dispatch, our work fully captures the dynamics of SOC levels to allow a better energy management policy, such as charging EVs to high

SOC levels as preparation before the rush hours come. Li et al. (2016) proposed a continuum approximation model for the design of a one-way EV sharing system which determines the optimal EV sharing station locations and the corresponding fleet sizes in order to minimize the comprehensive system cost. The model develops a comprehensive design framework for economically deploying a one-way EV sharing system and considers to provide a reliable service for fluctuating demands. However, in the model, an EV is not available before it is fully charged after it has been connected to a charger. An EV sharing system is able to serve more demands if EVs are allowed to provide services before they are fully charged. Besides, Brandstätter et al. (2017) proposed a two-stage stochastic time-dependent integer linear program to help with determining charging station locations, and provided a heuristic model to deal with large-scale instance. In the model, they also assume that only fully-charged vehicles can be assigned in order to prevent a car from running out of battery during a trip. Boyacı et al. (2017) further developed an integrated optimization-simulation framework for the operational planning problem of one-way EV sharing systems with reservations, in which an event-based simulator module is used to test the feasibility of electricity levels during the optimization procedure for operational decisions. Xu et al. (2018) proposed an optimization problem for one-way EV sharing system aiming to maximize its total profit by simultaneously determining the fleet size, pricing and relocation strategies.

The model we propose in this paper follows the framework of multistage mixed-integer stochastic programming where integers appear only at the first stage. Multistage stochastic programs are often used to model practical decision processes over time and under uncertainty, e.g., in energy planning (Pereira and Pinto, 1991; Fleten and Kristoffersen, 2008), insurance (Carino et al., 1994) and finance (Mulvey and Shetty, 2004). For more information on stochastic programming see (Kall et al., 1994; Shapiro et al., 2009). Compared with deterministic cases, stochastic programming can provide insight into a range of possible outcomes rather than just an expected outcome, so the solution of stochastic programming is more resilient in extreme cases. We can simplify a multistage stochastic programming formulation based on a scenario tree, but it could still be a really time-consuming task to solve due to the curse of dimensionality in the number of scenario combinations (Pereira and Pinto, 1991). Several exact solution methods have been studied and developed for scenario based multistage linear stochastic programming problems with relative complete recourse, such as SDDP method (Pereira and Pinto, 1991; Shapiro, 2011), the convergent cutting-plane algorithm (Chen and Powell, 1999) and the nested decomposition algorithm (Donohue and Birge, 2006).

3. Problem descriptions and modeling

We consider a one-way EV sharing system with spatial distributed stations I running through a finite planning horizon T . We assume an operation paradigm that repeats with an operational cycle of length $\bar{t}$ (e.g., a day) and only describes operations at some uniformly located discrete points within T , and therefore $T = \{0, 1, \dots, \bar{t}\}$ with a certain time point denoted by $t \in T$. Under this discrete framework, we aggregate all events happening within $[t, t+1)$ to time t . Each station can be either a physical station where idle EVs gather, or can be extended into a relatively larger zone where EVs are free floating. We assume the EV chargers are scarce and are installed in a selected subset of stations $I_c \subseteq I$. We denote the decision whether to install chargers in each station by binary variables $\{x_i\}_{i \in I}$, so that $I_c := \{i \in I : x_i = 1\}$. Since installing EV chargers are relatively long-term decisions comparing with the daily operations, we consider I_c is determined at the beginning and then remain fixed during the entire planning horizon. In the following paragraphs, we will just call the stations that have charging ability as *charging stations* for convenience. The capacity of each station i to hold EVs is given as $\bar{n}_i$, which is a predefined parameter reflecting the real scale of station i . Once a station is built as a charging station, the number of chargers in it is assumed to be equal to the capacity of the station, i.e., any EV at a charging station has a place to charge. We also need to decide the initial number of EVs deployed at each station, and we denote it by $\{n_i^0\}_{i \in I}$, and then the total number of EVs in the system can be represented by $\sum_{i \in I} n_i^0$.

All EVs in the system are assumed to be identical with finite states of charge (SOC) levels $E = \{0, 1, \dots, \bar{e}\}$, and a generic SOC level will be denoted by e . Notice that EVs in the same station may have different SOC levels, so in order to describe the EV profile at each station at any time, we denote the number of vehicles with an SOC level greater than or equal to e at station i at time t by n_{iet} . In this case, n_{i0t} also represents the total number of idle EVs at station i at time t . We require that at any time t , the number of EVs at station i cannot exceed the station's capacity $\bar{n}_i$,

$$n_{i0t} \leq \bar{n}_i, \quad \forall i \in I, t \in T. \quad (1)$$

Further, notice that the number of vehicles with an SOC level exactly equal to e at station i at time t can be represented as $n_{iet} - n_{i(e+1)t}$. Since this value cannot be negative, $\{n_{iet}\}$ should satisfy the following conditions,

$$n_{i(e-1)t} \geq n_{iet}, \quad \forall i \in I, e \in E \setminus \{0\}, t \in T. \quad (2)$$

The system aims to provide mobility service for customers from one station to another. We need to decide how to dispatch EVs over the system to meet the demand as customers appear. Moreover, it is necessary to relocate EVs when demands are spatially imbalanced over the system, i.e., the amount of EV picked up at a certain station is much smaller/larger than those returned. Let μ_{ijet} be the number of EVs with SOC levels greater than or equal to $e \in E$ to be dispatched at time t to serve customers who travel from station i to j . Therefore, we simply have the dispatched EVs satisfying $\mu_{ije't} \leq \mu_{ije''t}$ for any $e', e'' \in E$ such that $e'' < e'$. Similarly, let ν_{ijet} be the number of EVs relocated from station i to j with an SOC level greater

than or equal to e at time t . Following the condition of $\{n_{iet}\}$, we obtain the following conditions for $\{\mu_{ijet}\}$ and $\{v_{ijet}\}$,

$$\mu_{ij(e-1)t} \geq \mu_{ijet}, \quad \forall i, j \in I, e \in E \setminus \{0\}, t \in T, \quad (3)$$

$$v_{ij(e-1)t} \geq v_{ijet}, \quad \forall i, j \in I, e \in E \setminus \{0\}, t \in T. \quad (4)$$

In reality, relocation operations may need to consider the transition and availability of staff over the system. To capture this, we require that at any time t , the total number of EVs being relocated over the system cannot exceed a certain relocation capacity $\bar{v}$ (e.g., 10% of the total staff),

$$\sum_{i,j \in I} v_{ij0t} \leq \bar{v}, \quad \forall t \in T. \quad (5)$$

Suppose customer travel demands over any origin-destination pairs, e.g., from i to j , are subject to known time-varying distributions independently, which are denoted by $\{\lambda_{ijt}\}$. We further assume that travel from station i to j takes time τ_{ij} at an energy cost ε_{ij} . Similarly, we denote the relocation time and energy consumption by τ_{ij}^r and ε_{ij}^r , respectively. Notice that time (τ_{ij}, τ_{ij}^r) and energy cost $(\varepsilon_{ij}, \varepsilon_{ij}^r)$ should be integers under our discrete framework, and if they are not in reality, we need to round them up to fit in our formulation. To guarantee an EV can finish a trip, we exclude those EVs without sufficient SOC levels from services by the following constraints,

$$\mu_{ijet} = \mu_{ij\varepsilon_{ijt}} \quad \forall i, j \in I, e < \varepsilon_{ij}, t \in T, \quad (6)$$

$$v_{ijet} = v_{ij\varepsilon_{ij}^r t} \quad \forall i, j \in I, e < \varepsilon_{ij}^r, t \in T. \quad (7)$$

If an EV arrives at a charging station, we assume it starts to charge immediately at a constant charging speed regardless of its SOC. Without loss of generality, we assume the charging speed is exactly one SOC level per time interval for modeling convenience. To be more specific, if an EV with SOC level e ($e < \bar{e}$) is being charged at time t , its SOC level will increase to $e + 1$ at time $t + 1$. Under these settings, the EV profile n_{iet} change over time from $n_{ie(t-1)}$ due to the following two factors: (i) EVs with SOC level $e - 1$ at $t - 1$ are charged to level e at t , and (ii) EVs with SOC level equal or above e arrive/depart at station i from/to other stations due to dispatching or relocation at t . We assume that at the very beginning ($t = 0$), all EVs are at highest SOC level $\bar{e}$ (it is reasonable to make this assumption because we can get all EVs fully charged during the period when the service is closed, e.g., midnight), and hence the following joint vehicle flow-energy conservation law should be obeyed,

$$n_{ie0} = n_i^0, \quad \forall i \in I, e \in E, \quad (8)$$

$$n_{i0t} = n_{i0(t-1)} - \sum_{j \in I} \mu_{ij0(t-1)} - \sum_{j \in I} v_{ij0(t-1)} + \sum_{j \in I} \mu_{ji0(t-\tau_{ji})} + \sum_{j \in I} v_{ji0(t-\tau_{ji}^r)}, \quad \forall i \in I, t \in T \setminus \{0\}, \quad (9)$$

$$n_{iet} = \left(n_{i(e-1)(t-1)} - \sum_{j \in I} \mu_{ij(e-1)(t-1)} - \sum_{j \in I} v_{ij(e-1)(t-1)} \right) x_i + \left(n_{ie(t-1)} - \sum_{j \in I} \mu_{ije(t-1)} - \sum_{j \in I} v_{ije(t-1)} \right) (1 - x_i) \\ + \sum_{j \in I} \mu_{ji(e+\varepsilon_{ji})(t-\tau_{ji})} + \sum_{j \in I} v_{ji(e+\varepsilon_{ji}^r)(t-\tau_{ji}^r)}, \quad \forall i \in I, e \in E \setminus \{0\}, t \in T \setminus \{0\}. \quad (10)$$

Since customer demands are random, there are chances when no idle EV or no EV with sufficient SOC is available for dispatching, especially when the demand surge happens in the rush hour. In such scenario, extra demands may either line up as a queue or abandon our EV service. In light of this, we suppose a ratio q of queuing customers will choose to stay in the queue to the next time interval, while the rest choose to leave. Let q_{ijt} be the customers waiting in queue at station i attempting to travel to station j at time t and l_{ijt} be the corresponding number of customers lost at time t . The following constraints capture the dynamic of queue length with an assumption that no customers waiting in queue at the very beginning,

$$q_{ij0} = 0, \quad \forall i, j \in I, \quad (11)$$

$$q_{ijt} = q(q_{ij(t-1)} + \lambda_{ij(t-1)} - \mu_{ij0(t-1)}), \quad \forall i, j \in I, t \in T \setminus \{0\}, \quad (12)$$

$$l_{ijt} = (1 - q)(q_{ij(t-1)} + \lambda_{ij(t-1)} - \mu_{ij0(t-1)}), \quad \forall i, j \in I, t \in T \setminus \{0\}. \quad (13)$$

To reduce the queue length and improve the service quality, real-time fleet charging and dispatching operations should be managed according to the demand uncertainties. That is, we seek a *real-time demand-adaptive fleet operation policy* for the EV sharing system. To evaluate the operation policy, we sample the random customer demands according to the known distribution to construct a scenario tree over T , where the total expected cost can be estimated through the average cost

over each scenario. To overcome the exploded number of scenarios required to get an unbiased estimation, we divide $T = \{0, 1, \dots, \bar{t}\}$ into K stages at equal time length, $T_k := \{t \in T : t_{k-1} \leq t < t_k\}$, where $t_0 = 0$ and $t_k = \frac{k\bar{t}}{K}$ for $k = 1, 2, \dots, K$. In addition, suppose at the beginning of stage k , demands over the whole stage is observed, i.e., the uncertainties are realized, while the demand over future stages remains uncertain. The length of each stage can be interpreted as a short in-advance reservation for customers. As such, fleet operation decisions for stage k are made at the beginning of the stage, t_{k-1} , with the realized demand information in T_k and the prediction of demand in future time horizon.

To minimize the total expected cost over T together with the initial infrastructure deployment cost, we refine the decision process into a multistage problem with the following $K + 1$ stages. First, at stage 0, the system decides the charging stations and the initial deployment of EVs at each station, i.e., $\{x_i\}$ and $\{n_i^0\}$, which together determine the infrastructure deployment decision. With a unit fixed infrastructure building cost c_i^f (including installment and maintenance fee per charger) and a unit EV holding cost c^h , we have the initial planning cost (stage 0 cost) for the system as

$$f_0(X_0) = \sum_{i \in I} (c_i^f \bar{n}_i x_i + c^h n_i^0),$$

where $X_0 := \{x_i, n_i^0 : i \in I\}$ indicates the corresponding initial stage decision.

Then based on the infrastructure deployment decision, we address the real-time EV fleet operation decisions in the following K stages. With the penalty for each unit length of queue c_t^w at each time t , the penalty for customer loss c_t^l at each time t , and the vehicle relocation cost c_{ijt}^r from station i to j at time t , we have the cost for each stage as

$$f_k(X_k, Y_k) = \sum_{i, j \in I, t \in T_k} (c_t^w q_{ijt} + c_t^l l_{ijt} + c_{ijt}^r v_{ijor}),$$

where $X_k := \{n_{iet}, q_{ijt}, l_{ijt} : i, j \in I, e \in E, t \in T_k\}$ and $Y_k := \{\mu_{ijet}, v_{ijet} : i, j \in I, e \in E, t \in T_k\}$ are state and stage decision variables, respectively. In particular, the stage decision Y_k captures the fleet operation decision in real time and the state decision X_k reflects the status of the system. Notice that X_k and Y_k are random decision variables relying on the observation of the demand till stage k . Therefore, the system has the following MultiStage Problem (MSP) to minimize the total expected cost over the entire planing horizon,

$$\zeta^{\text{MSP}} = \min_{X_0, \{X_k, Y_k\}} f_0(X_0) + \mathbb{E}[\min \{f_1(X_1, Y_1) + \mathbb{E}[\min \{f_2(X_2, Y_2) + \dots + \mathbb{E}[\min f_K(X_K, Y_K)]\}] \dots]$$

Constraints (1) to (13),

$$x_i \in \{0, 1\}, \forall i \in I, \quad (14)$$

$$n_i^0, n_{iet}, \mu_{ijet}, v_{ijet}, q_{ijt}, l_{ijt} \in \mathbb{Z}_{\geq 0}, \forall i, j \in I, e \in E, t \in T. \quad (15)$$

Here, Constraints (14) and (15) define the domains of all decision variables.

In fact, the MSP involves many implicit random decision variables as well as uncertain customer demands. Now we reformulate it into a scenario tree based model (Fleten and Kristoffersen, 2008) to facilitate the solution algorithm. A scenario tree is used to describe the uncertain data evolving over time in a multistage stochastic process. Within each stage k , suppose the observed customer demand at t_{k-1} is $\lambda_{ijt} = \lambda_{ijt}[\xi_k]$ for $t \in T_k$. When observing a particular demand profile, we determine the corresponding state and stage decisions, i.e., $X_k = X_k[\xi_k]$ and $Y_k = Y_k[\xi_k]$, respectively. Let $\xi_{[1,k]} := \{\xi_1, \xi_2, \dots, \xi_k\}$ be a sample path of demand realization till stage k . We assume that a specific realization of $\xi_{[1,K]} = \xi_{[1,K]}^s$ is taken from a discrete distribution indexed by a finite set S , with corresponding probabilities such that $\sum_{s \in S} p(s) = 1$. Each realization $\xi_{[1,K]}^s$ is called a scenario and captures a certain realization of demands over T represented by $\lambda_{ijt}(\xi_{[1,K]}^s) = \{\lambda_{ijt}[\xi_k^s] : k = 1, 2, \dots, K\}$, and we denote it by $\lambda(s)$ for demonstration convenience. Similarly, the corresponding state and stage decisions over a particular realized scenario are denoted as $X[\xi_{[1,K]}^s]$, $Y[\xi_{[1,K]}^s]$, or $X(s)$, $Y(s)$, respectively. We organize these realizations into a finite scenario tree with K levels. Each node at level k corresponds to a certain realization of $\xi_{[1,k]}$, and branches based on the demand realization at stage $k + 1$. Part of the scenario tree is shown in Fig. 2 as a specific example where each stage contains 15 time points. Let L_k be the index set of all level k nodes, and $\Omega_k^l (l \in L_k)$ be a specific level k node. In particular, each scenario s is a leaf of the tree and we say $s \in \Omega_k^l$ if node Ω_k^l is the ancestor of s . Therefore, if $s_1, s_2 \in \Omega_k^l$, we know s_1, s_2 have the same demand realization till stage k . In the following demonstration, for convenience, we let $\Psi(\lambda(s)) := X_0$,

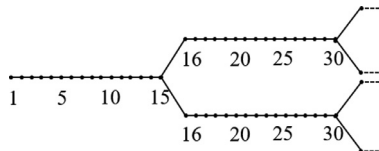


Fig. 2. Part of a scenario tree.

$\{X(s), Y(s)\}$, satisfying Constraints (1) to (15), be the feasible decision set given a realized demand $\lambda(s)$. Then we formulate the following scenario-based MultiStage Problem (SMSP) for an EV sharing system,

$$\begin{aligned} \zeta^{\text{SMSP}} = & \min_{X_0, \{X_k(s), Y_k(s)\}, \alpha, \beta} \left(\sum_{i \in I} (c_i^f \bar{n}_i x_i + c^h n_i^0) + \sum_{s \in S} p(s) \left(\sum_{i, j \in I, t \in T} (c_t^w q_{ijt}(s) + c_t^l l_{ijt}(s) + c_{ijt}^r v_{ij0t}(s)) \right) \right) \\ \text{s.t. } & v_{ijet}(s) = \alpha_{ijet}^{kl}, \quad \forall i, j \in I, e \in E, k \in K, \forall t \in T_k, l \in L_k, s \in \Omega_k^l, \end{aligned} \quad (16)$$

$$\mu_{ijet}(s) = \beta_{ijet}^{kl}, \quad \forall i, j \in I, e \in E, k \in K, \forall t \in T_k, l \in L_k, s \in \Omega_k^l, \quad (17)$$

$$X_0, \{X(s), Y(s)\} \in \Psi(\lambda(s)), \quad \forall s \in S. \quad (18)$$

In SMSP, $\alpha := \{\alpha_{ijet}^{kl} : i, j \in I, e \in E, k \in \{1, 2, \dots, K-1\}, t \in T_k, l \in L_k\}$ and $\beta := \{\beta_{ijet}^{kl} : i, j \in I, e \in E, k \in \{1, 2, \dots, K-1\}, t \in T_k, l \in L_k\}$ are newly introduced auxiliary variables. Constraints (16) and (17) are linear nonanticipativity constraints which guarantee such scenario based decisions are realistic. In particular, at stage k , $\{s : s \in \Omega_k^l\}$ is a set of scenarios which share a common sample path till this stage. So their decisions $\{X_k, Y_k\}$ should be consistent with each other at this stage. Furthermore, since $\{X_k\}$ are forced to be the same if $\{Y_k\}$ are the same due to Constraints (9) and (10), we only need to consider the nonanticipativity constraints on $\{Y_k\}$. Constraints (18) indicate the feasibility of decisions at scenario s .

4. Solution algorithm

SMSP is a multistage nonlinear integer stochastic program, which is extremely time-consuming and requires unacceptably high computational resources to solve directly, due to the curse of dimensionality. This is the well known difficulty in such multistage stochastic problems involving integer decisions and nonlinearity. In this section, we will develop a solution algorithm for the LP relaxation of SMSP which can provide a good infrastructure decision within much shorter time, and also a way to check the optimality gap of the solution. First, we apply Lagrangian relaxation method by relaxing nonanticipativity constraints. A projected subgradient method is developed to solve the corresponding Lagrangian dual problem, which naturally leads to a feasible solution to the initial stage decision as well as a lower bound of the optimal objective value. Further, notice that the integer decisions are only shown in the initial stage, while the following K stages, as a subproblem given the initial stage decision, are only with continuous decision variables. We can further linearize it into a multistage linear stochastic problem, and utilize an algorithm based on SDDP to provide an optimality gap under the assumption of stage-wise independent distribution of demands.

4.1. Linearization reformulation

First, we address the nonlinearity. Notice that in SMSP, Constraints (10) are bi-linear due to the products of binary variables x_i and non-negative variables n_{iet} , μ_{ijet} and v_{ijet} . This can be linearized by the so-called big M reformulation. By introducing new variables N_{iet} , U_{ijet} , V_{ijet} such that $N_{iet} = n_{iet}x_i$, $U_{ijet} = \mu_{ijet}x_i$, $V_{ijet} = v_{ijet}x_i$. Constraints (10) can be linearized by replacing all bi-linear terms with N_{iet} , U_{ijet} , V_{ijet} , and all we need is to add corresponding linear constraints to impose the definition of N_{iet} , U_{ijet} , V_{ijet} . Take $N_{iet} = n_{iet}x_i$ as an example, it is equivalent to the following constraints:

$$\begin{aligned} 0 & \leq N_{iet} \leq n_{iet}, \quad \forall i \in I, e \in E, t \in T, \\ N_{iet} & \leq \bar{n}_i x_i, \quad \forall i \in I, e \in E, t \in T, \\ N_{iet} & \geq \bar{n}_i (x_i - 1) + n_{iet}, \quad \forall i \in I, e \in E, t \in T. \end{aligned}$$

Notice that we set the big M as $\bar{n}_i$ in the above reformulation since $\bar{n}_i$ is an upper bound to n_{iet} based on Constraints (1) and (2).

In the following demonstration, we will treat Constraints (10) as “linear” constraints set for demonstration convenience.

4.2. Linear programming relaxation

As shown in Constraints (15), the fleet size and daily operation decisions should be nonnegative integer variables in reality, especially when the volume is close to zero. However, when the problem scale becomes large, involving these integer decisions makes our problem easily intractable. To this end, we relax Constraints (15) to its continuous version,

$$n_i^0, n_{iet}, \mu_{ijet}, v_{ijet}, q_{ijt}, l_{ijt} \geq 0, \quad \forall i, j \in I, e \in E, t \in T, \quad (19)$$

so that SMSP only have 0–1 integers and nonnegative continuous variables. We can foresee such relaxation approximates the optimal solution well when the value of optimal decision variables are large. However, such approximation need to be validated when demand is small, which will be conducted in Section 5.1. We will show that it is effective to relax the integer constraints to linear constraints even when the demand is low.

In the following demonstration, we will consider linear constraints (19) instead of (15) by default without further illustration.

4.3. Lagrangian relaxation

Due to the curse of dimensionality, directly solving the deterministic equivalent SMSP requires unacceptably high computational resources. Luckily, SMSP has a potential decomposable structure that can be algorithmically exploited. We introduce scenario-dependent copies $X_0(s)$ of initial stage decision X_0 for each scenario $s \in S$ to create the following reformulation of SMSP:

$$\zeta^{\text{SMSP}} = \min_{X_0(s), \{X_k(s), Y_k(s)\}, \alpha, \beta, \gamma, \delta} \sum_{s \in S} p(s) \sum_{i \in I} (c_i^f \bar{n}_i x_i(s) + c_i^h n_i^0(s)) + \sum_{s \in S} p(s) \sum_{i, j \in I, t \in T} (c_t^w q_{ijt}(s) + c_t^l l_{ijt}(s) + c_{ijt}^r v_{ij0t}(s)) \quad (20)$$

$$\text{s.t. } x_i(s) = \gamma_i, \quad \forall i \in I, s \in S, \quad (21)$$

$$n_i^0(s) = \delta_i, \quad \forall i \in I, s \in S, \{X_0(s), \{X_k(s), Y_k(s)\}, \alpha, \beta\} \in \Phi(\lambda(s)), \quad \forall s \in S, \quad (22)$$

where $\gamma := \{\gamma_i : i \in I\}$ and $\delta := \{\delta_i : i \in I\}$ are newly introduced auxiliary variables. $\Phi(\lambda(s))$ consists of Constraints (16) to (18). Constraints (21) and (22) enforce nonanticipativity for $X_0(s)$; the initial infrastructure design $x_i(s)$ and $n_i^0(s)$ must be the same (γ_i and δ_i) for each scenario $s \in S$ respectively. Considering the difficulty of dealing with all the nonanticipativity constraints (16), (17), (21), and (22), we apply Lagrangian relaxation to these constraints and yield the following nonanticipativity Lagrangian dual function

$$\phi(\mathcal{L}) := \min_{X_0(s), \{X_k(s), Y_k(s)\}, \alpha, \beta, \gamma, \delta} \left\{ \begin{aligned} & \sum_{s \in S} p(s) \sum_{i \in I} (c_i^f \bar{n}_i x_i(s) + c_i^h n_i^0(s)) + \\ & \sum_{s \in S} p(s) \sum_{i, j \in I, t \in T} (c_t^w q_{ijt}(s) + c_t^l l_{ijt}(s) + c_{ijt}^r v_{ij0t}(s)) \\ & + \sum_{s \in S, i \in I} \mathcal{X}_i(s) (x_i(s) - \gamma_i) + \sum_{s \in S, i \in I} \mathcal{N}_i(s) (n_i^0(s) - \delta_i) \\ & + \sum_{k \in K} \left(\sum_{i, j \in I, e \in E, t \in T_k, l \in L_k} \left(\sum_{s \in \Omega_k^l} \mathcal{V}_{ijet}^{kl}(s) (v_{ijet}(s) - \alpha_{ijet}^{kl}) \right) \right) \\ & + \sum_{k \in K} \left(\sum_{i, j \in I, e \in E, t \in T_k, l \in L_k} \left(\sum_{s \in \Omega_k^l} \mathcal{U}_{ijet}^{kl}(s) (\mu_{ijet}(s) - \beta_{ijet}^{kl}) \right) \right) \end{aligned} \right\}, \quad (23)$$

s.t. $\{X_0(s), X_k(s), Y_k(s)\} \in \Psi(\lambda(s)), \quad \forall s \in S$

where $\mathcal{X}_i(s), \mathcal{N}_i(s), \mathcal{V}_{ijet}^{kl}(s), \mathcal{U}_{ijet}^{kl}(s)$ are corresponding Lagrangian multipliers associated with the relaxed constraints (21), (22), (16), and (17) respectively, and $\mathcal{L} := \{\mathcal{X}_i(s), \mathcal{N}_i(s), \mathcal{V}_{ijet}^{kl}(s), \mathcal{U}_{ijet}^{kl}(s) : s \in S\}$ is the set of all Lagrangian multipliers. Since $\alpha, \beta, \gamma, \delta$ are unconstrained in the optimization problem in (23), to guarantee that $\phi(\mathcal{L})$ is bounded from below, we require the following constraints for $\mathcal{L}$ to be satisfied as a condition of dual feasibility:

$$\sum_{s \in S} \mathcal{X}_i(s) = 0, \quad \forall i \in I, \quad (24)$$

$$\sum_{s \in S} \mathcal{N}_i(s) = 0, \quad \forall i \in I, \quad (25)$$

$$\sum_{s \in \Omega_k^l} \mathcal{V}_{ijet}^{kl}(s) = 0, \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k, \quad (26)$$

$$\sum_{s \in \Omega_k^l} \mathcal{U}_{ijet}^{kl}(s) = 0, \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k. \quad (27)$$

Under this condition, the terms in (23) for $\alpha, \beta, \gamma, \delta$ vanish. Let $X := \{X_0(s), \{X_k(s), Y_k(s)\} : s \in S\}$ denote the overall decision variable set, and $\Psi := \{\Psi(\lambda(s)) : s \in S\}$ be the overall constraint set, the Lagrangian dual function (23) can be reformulated as

$$\phi(\mathcal{L}) := \min_X \{ \mathcal{F}(X, \mathcal{L}) : X \in \Psi \}, \quad (28)$$

where

$$\begin{aligned} \mathcal{F}(X, \mathcal{L}) = & \sum_{s \in S} p(s) \sum_{i \in I} (c_i^f \bar{n}_i x_i(s) + c_i^h n_i^0(s)) \\ & + \sum_{s \in S} p(s) \sum_{i, j \in I, t \in T} (c_t^w q_{ijt}(s) + c_t^l l_{ijt}(s) + c_{ijt}^r v_{ij0t}(s)) \end{aligned}$$

$$\begin{aligned}
& + \sum_{s \in S, i \in I} \mathcal{X}_i(s) \mathcal{X}_i(s) + \sum_{s \in S, i \in I} \mathcal{N}_i(s) n_i^0(s) \\
& + \sum_{k \in K} \left(\sum_{i, j \in I, e \in E, t \in T_k, l \in L_k} \left(\sum_{s \in \Omega_k^l} \mathcal{V}_{ijet}^{kl}(s) \mathcal{V}_{ijet}(s) \right) \right) \\
& + \sum_{k \in K} \left(\sum_{i, j \in I, e \in E, t \in T_k, l \in L_k} \left(\sum_{s \in \Omega_k^l} \mathcal{U}_{ijet}^{kl}(s) \mathcal{U}_{ijet}(s) \right) \right).
\end{aligned}$$

Notice that the overall constraints set Ψ is scenario-decomposable. Let $X(s) := \{X_0(s), \{X_k(s), Y_k(s)\}\}$ and $\mathcal{L}(s) := \{\mathcal{X}_i(s), \mathcal{N}_i(s), \mathcal{V}_{ijet}^{kl}(s), \mathcal{U}_{ijet}^{kl}(s)\}$ denote the decision variables set and Lagrangian multipliers set under scenario s respectively, and $\mathcal{F}_s(X(s), \mathcal{L}(s))$ be a function of terms related to $X(s)$ in $\mathcal{F}(X, \mathcal{L})$, then problem (28) can be decomposed as:

$$\phi(\mathcal{L}) = \sum_{s \in S} \phi_s(\mathcal{L}(s)),$$

where

$$\phi_s(\mathcal{L}(s)) = \min_{X(s)} \{\mathcal{F}_s(X(s), \mathcal{L}(s)) : X(s) \in \Psi(\lambda(s))\}.$$

For any choice of $\mathcal{L}$ that satisfies the dual feasibility condition, i.e., $\mathcal{L} \in \Upsilon := \{\mathcal{L} \text{ satisfying Constraints (24) to (27)}\}$, the value of the Lagrangian dual function provides a lower bound to the optimal objective value of SMSP: $\phi(\mathcal{L}) \leq \zeta^{\text{SMSP}}$. We would like to find the best lower bound, i.e., to solve the Lagrangian Dual (LD) problem:

$$\zeta^{\text{LD}} := \max_{\mathcal{L}} \{\phi(\mathcal{L}) : \mathcal{L} \in \Upsilon\}. \quad (29)$$

4.4. Projected subgradient method

The subgradient method can be used to minimize a nondifferentiable convex function (Boyd et al., 2003). To solve problem (29), first we need to find a subgradient $g(\mathcal{L})$ for the negative dual function $-\phi(\mathcal{L})$, which satisfies the follows. Given any $\mathcal{L}_1 \in \Upsilon$, the inequality

$$-\phi(\mathcal{L}_2) \geq -\phi(\mathcal{L}_1) + g(\mathcal{L}_1) \cdot (\mathcal{L}_2 - \mathcal{L}_1) \quad (30)$$

holds for any $\mathcal{L}_2 \in \Upsilon$. Let $X^*(\mathcal{L})$ be one of the minimizers for problem (28), i.e., $\phi(\mathcal{L}) = \mathcal{F}(X^*(\mathcal{L}), \mathcal{L})$. Since $\phi(\mathcal{L}_2) = \mathcal{F}(X^*(\mathcal{L}_2), \mathcal{L}_2) \leq \mathcal{F}(X^*(\mathcal{L}_1), \mathcal{L}_2)$, if the inequality

$$-\mathcal{F}(X^*(\mathcal{L}_1), \mathcal{L}_2) \geq -\mathcal{F}(X^*(\mathcal{L}_1), \mathcal{L}_1) + g(\mathcal{L}_1) \cdot (\mathcal{L}_2 - \mathcal{L}_1) \quad (31)$$

holds, the equality (30) holds as well. So all we need to do is to find a subgradient $g(\mathcal{L})$ for $-\mathcal{F}(X^*(\mathcal{L}), \mathcal{L})$. Since $\mathcal{F}(X^*(\mathcal{L}), \mathcal{L})$ is linear over $\mathcal{L}$, we can easily obtain a subgradient for $-\mathcal{F}(X^*(\mathcal{L}), \mathcal{L})$ as

$$g(\mathcal{L}) = -\frac{\partial \mathcal{F}(X^*(\mathcal{L}), \mathcal{L})}{\partial \mathcal{L}}.$$

Written out in components, we have

$$\begin{aligned}
g(\mathcal{X}_i(s)) &= -\mathcal{X}_i^*(s), \quad \forall i \in I, s \in S, \\
g(\mathcal{N}_i(s)) &= -n_i^{0*}(s), \quad \forall i \in I, s \in S, \\
g(\mathcal{V}_{ijet}^{kl}(s)) &= -v_{ijet}^{*kl}(s), \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k, s \in \Omega_k^l, \\
g(\mathcal{U}_{ijet}^{kl}(s)) &= -\mu_{ijet}^{*kl}(s), \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k, s \in \Omega_k^l,
\end{aligned}$$

where $\{x_i^*(s), n_i^{0*}(s), v_{ijet}^{*kl}(s), \mu_{ijet}^{*kl}(s)\}$ can be obtained by solving problem (28).

The subgradient method applied to the dual problem (29) can be expressed as

$$\begin{aligned}
X &:= X^*(\mathcal{L}), \\
g(\mathcal{L}) &:= -\frac{\partial \mathcal{F}(X, \mathcal{L})}{\partial \mathcal{L}}, \\
\mathcal{L} &:= P(\mathcal{L} - \sigma g(\mathcal{L})),
\end{aligned}$$

where σ is the step length which may vary over iterations, and P is a (Euclidean) projection on Υ . Specifically, for $\mathcal{L} \notin \Upsilon$, we calculate the average of left-hand side in Υ as

$$\frac{\sum_{s \in S} \mathcal{X}_i(s)}{|S|} = \bar{\mathcal{X}}_i, \quad \forall i \in I,$$

$$\frac{\sum_{s \in S} \mathcal{N}_i(s)}{|S|} = \bar{\mathcal{N}}_i, \quad \forall i \in I,$$

$$\frac{\sum_{s \in \Omega_k^l} \mathcal{V}_{ijet}^{kl}(s)}{|\{s : s \in \Omega_k^l\}|} = \bar{\mathcal{V}}_{ijet}^{kl}, \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k,$$

$$\frac{\sum_{s \in \Omega_k^l} \mathcal{U}_{ijet}^{kl}(s)}{|\{s : s \in \Omega_k^l\}|} = \bar{\mathcal{U}}_{ijet}^{kl}, \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k,$$

where $\bar{\mathcal{Z}} := \{\bar{\mathcal{X}}_i, \bar{\mathcal{N}}_i, \bar{\mathcal{V}}_{ijet}^{kl}, \bar{\mathcal{U}}_{ijet}^{kl}\}$ represents the distance to Υ . Then the projection P can be constructed in components as

$$P(\mathcal{X}_i(s)) = \mathcal{X}_i(s) - \bar{\mathcal{X}}_i, \quad \forall i \in I, s \in S,$$

$$P(\mathcal{N}_i(s)) = \mathcal{N}_i(s) - \bar{\mathcal{N}}_i, \quad \forall i \in I, s \in S,$$

$$P(\mathcal{V}_{ijet}^{kl}(s)) = \mathcal{V}_{ijet}^{kl}(s) - \bar{\mathcal{V}}_{ijet}^{kl}, \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k, s \in \Omega_k^l,$$

$$P(\mathcal{U}_{ijet}^{kl}(s)) = \mathcal{U}_{ijet}^{kl}(s) - \bar{\mathcal{U}}_{ijet}^{kl}, \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k, s \in \Omega_k^l,$$

which satisfies $P(\mathcal{Z}) \in \Upsilon$.

The algorithm converges when the step size σ is summable, and we let it terminate if the violation of $X^*(\mathcal{Z})$ for the relaxed nonanticipativity constraints (16), (17), (18), and (22) is within a tolerance ϵ . For a given X^* , we calculated the expected decision as

$$\bar{x}_i = \sum_{s \in S} p(s) x_i(s), \quad \forall i \in I,$$

$$\bar{n}_i^0 = \sum_{s \in S} p(s) n_i^0(s), \quad \forall i \in I,$$

$$\bar{\alpha}_{ijet}^{kl} = \sum_{s \in \Omega_k^l} p(s) v_{ijet}(s), \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k,$$

$$\bar{\beta}_{ijet}^{kl} = \sum_{s \in \Omega_k^l} p(s) \mu_{ijet}(s), \quad \forall k \in K, \forall i, j \in I, e \in E, t \in T_k, l \in L_k.$$

The motivated gap can be calculated as

$$\begin{aligned} \varepsilon^2(X^*) = & \sum_{s \in S} \|x_i(s) - \bar{x}_i\|_2^2 + \sum_{s \in S} \|n_i^0(s) - \bar{n}_i^0\|_2^2 \\ & + \sum_{k \in K} \sum_{i, j \in I, e \in E, t \in T_k, l \in L_k} \|v_{ijet}(s) - \bar{\alpha}_{ijet}^{kl}\|_2^2 \\ & + \sum_{k \in K} \sum_{i, j \in I, e \in E, t \in T_k, l \in L_k} \|\mu_{ijet}(s) - \bar{\beta}_{ijet}^{kl}\|_2^2, \end{aligned}$$

which is the distance between scenario-decomposed decisions and its average, measured by its square norm. If $\varepsilon^2 = 0$, the nonanticipativity constraints (16), (17), (18), and (22) are satisfied.

We can obtain the initial stage decision $\bar{X}_0$ by properly rounding the expected decision based on the final result of X^* as well as a lower bound given by the final result of ζ^{LD} . We denote this solution as $\{\bar{x}_i, \bar{n}_i^0\}$ and the lower bound as z^{LR} for future demonstration convenience. We summarize the projected subgradient algorithm in Algorithm 1.

4.5. SDDP based algorithm

Once we get the initial stage decision $\bar{X}_0$, we can reformulate the rest of the problem into a pure MultiStage Linear Program (MSLP) whose optimal value gives an upper bound for SMSP. We assume the maximum travel time and relocation time is smaller than the length of each stage so that there is no chance that an EV is on the road through an entire stage (e.g., starting from Stage 1 but ends at Stage 3), and thus $\{X_{k-1}, Y_{k-1}\}$ and demand realization λ_k together provide complete information of the system state for determining $\{X_k, Y_k\}$. In particular, trips starting and ending in consecutive stages (e.g., starting from Stage $k-1$ but ending at Stage k) are easily captured, since $\{Y_{k-1}\}$ contain all departure information of such cross-stage trips and arrival information is provided through the flow-energy conservation constraints (9) and (10). We also assume that the probability distribution of demand at stage k is independent of the specific realization of demand at the previous stage $k-1$, with probability $p_k(\xi_k)$ such that $\sum_{\xi_k \in \Xi_k} p_k(\xi_k) = 1$ for all k where Ξ_k is the set of all possible

Algorithm 1 Projected Subgradient Algorithm applied to problem (29).

```

function PSA( $\sigma, \epsilon, r_{\max}$ )
  Initialize:  $\mathcal{L}^{(0)} := 0$ ;
  for  $r = 0, \dots, r_{\max}$  do
    for  $s \in S$  do
      Solve  $X^{(r)}(s) = \arg \min_{X(s)} \{ \mathcal{F}_s(X(s), \mathcal{L}^{(r)}(s)), X(s) \in \Psi(\lambda(s)) \}$ ;
    end for
    Set  $\zeta^{(r)} := \sum_{s \in S} p(s) \left( \sum_{i \in I} \left( c_i^f \bar{n}_i \chi_i^{(r)}(s) + c_i^h \bar{n}_i^{0(r)}(s) \right) + \sum_{i,j \in I, t \in T} \left( c_t^w q_{ijt}^{(r)}(s) + c_t^l l_{ijt}^{(r)}(s) + c_{ijt}^r v_{ij0t}^{(r)}(s) \right) \right)$ ;
    if  $r > 0$  and  $\epsilon^2(X^{(r)}) < \epsilon$ 
      Set  $\{\bar{x}_i, \bar{n}_i^0\} := \text{int}(\sum_{s \in S} p(s) X^{(r)}(0))$ ;
      Set  $z^{\text{LR}} := \zeta^{(r)}$ ;
      return  $(\bar{x}_i, \bar{n}_i^0, z^{\text{LR}})$ ;
    end if
    Set  $g^{(r)} := -\frac{\partial \mathcal{F}(X^{(r)}, \mathcal{L})}{\partial \mathcal{L}}$ ;
    Set  $\mathcal{L}^{(r+1)} := P(\mathcal{L}^{(r)} - \sigma^{(r)} g^{(r)})$ ;
  end for
  Set  $\{\bar{x}_i, \bar{n}_i^0\} := \text{int}(\sum_{s \in S} p(s) X^{(r_{\max})}(0))$ ;
  Set  $z^{\text{LR}} := \zeta^{(r_{\max})}$ ;
  return  $(\bar{x}_i, \bar{n}_i^0, z^{\text{LR}})$ ;
end function

```

realizations of demand at stage k . Under such independence assumption, we have $p(s) = \prod_k p_k(\xi_k^s)$ for each $s \in S$. Let

$$\bar{\Psi}_k(X_{k-1}, Y_{k-1}, \xi_k) = \{X_k, Y_k \text{ satisfying } \bar{\Psi}(\lambda[\xi_k]) \text{ for } t \in T_k \text{ given } \{X_{k-1}, Y_{k-1}\}\}$$

be the corresponding constraint set given the decision $\{X_{k-1}, Y_{k-1}\}$ at previous stage and the realization of demand $\lambda[\xi_k]$ at current stage. So the condition that the multistage problem achieves optimum is given by the Bellman equation

$$\alpha_k(X_{k-1}, Y_{k-1}) = \sum_{\xi_k \in \Xi_k} p_k(\xi_k) \left\{ \min_{\{X_k, Y_k\} \in \bar{\Psi}_k(X_{k-1}, Y_{k-1}, \xi_k)} \{f_k(X_k, Y_k) + \alpha_{k+1}(X_k, Y_k)\} \right\}, \quad (32)$$

where $\alpha_k(X_{k-1}, Y_{k-1})$ is the expected cost-to-go function at stage k given $\{X_{k-1}, Y_{k-1}\}$. Notice that here the cost-to-go function $\alpha_k(X_{k-1}, Y_{k-1})$ does not depend on the data process due to the stage-wise independence assumption.

In particular, given the initial stage decision $\{\bar{x}_i, \bar{n}_i^0\}$, the optimal value of MSLP can be represented as

$$\begin{aligned} \zeta^{\text{MSLP}} &= \sum_{i \in I} (c_i^f \bar{n}_i \bar{x}_i + c_i^h \bar{n}_i^0) + \alpha_1(\{\bar{x}_i, \bar{n}_i^0\}), \\ \text{s.t. } \alpha_k(X_{k-1}, Y_{k-1}) &= \sum_{\xi_k \in \Xi_k} p_k(\xi_k) \\ &\left\{ \min_{\{X_k, Y_k\} \in \bar{\Psi}_k(X_{k-1}, Y_{k-1}, \xi_k)} \sum_{i,j \in I, t \in T_k} (c_t^w q_{ijt}(s) + c_t^l l_{ijt}(s) + c_{ijt}^r v_{ij0t}(s)) + \alpha_{k+1}(X_k, Y_k) \right\}, \quad \forall k \in K, \end{aligned}$$

where $\{X_0, Y_0\} = \{\bar{x}_i, \bar{n}_i^0\}$ and $\alpha_{K+1}(X_K, Y_K) = 0$. Also notice that for any reasonable $\{X_{k-1}, Y_{k-1}\}$ and demand $\lambda[\xi_k]$, the constraint set $\bar{\Psi}_k(X_{k-1}, Y_{k-1}, \xi_k)$ is not empty (e.g., at least the system can do nothing and leave all customers waiting in queue or being lost, which will not break any constraints), so the optimal solution exists.

When $K = 1$, MSLP can be easily solved by dealing with each scenario separately. When $K = 2$, MSLP is a set of two-stage stochastic problems featured by first stage scenarios, and each can be solved by the famous L-shaped method (Van Slyke and Wets, 1969). The basic principle behind the L-shaped method is to approximate the recourse term $\alpha_3(X_2, Y_2)$ in the objective function by adding Bender's cuts iteratively. Following a similar procedure called SDDP which was introduced by Pereira and Pinto (1991), we extend the method to the case when $K > 2$. For demonstration convenience, we represent linear constraint set $\bar{\Psi}_k(X_{k-1}, Y_{k-1}, \xi_k)$ as $C_{\xi_k, k} X_k + D_{\xi_k, k} Y_k \geq b_{\xi_k, k} - A_{\xi_k, k} X_{k-1} - B_{\xi_k, k} Y_{k-1}$ where $A_{\xi_k, k}$, $B_{\xi_k, k}$, $C_{\xi_k, k}$, $D_{\xi_k, k}$, and $b_{\xi_k, k}$ are the corresponding constant coefficients matrices (vectors) captured in $\bar{\Psi}_k(X_{k-1}, Y_{k-1}, \xi_k)$, and introduce a new variable θ to convert each minimization problem in (32) to the following formulation:

$$\begin{aligned} &(\text{LP}_{\xi_k, k}) \min_{X_k, Y_k, \theta \geq 0} f_k(X_k, Y_k) + \theta \\ \text{s.t. } &C_{\xi_k, k} X_k + D_{\xi_k, k} Y_k \geq b_{\xi_k, k} - A_{\xi_k, k} X_{k-1} - B_{\xi_k, k} Y_{k-1}, \end{aligned} \quad (33)$$

$$\theta \geq \alpha_{k+1}(X_k, Y_k). \quad (34)$$

Similar to the two-stage case, Constraint (34) is not explicit, so $LP_{\xi_k, k}$ cannot be solved directly. Following the principle of Bender's decomposition, we ignore the recursive constraint (34) on the cost-to-go function but add Benders' cut to $LP_{\xi_k, k}$ iteratively later to improve the relaxation. Specifically, at each iteration, given the current solution of $\{\bar{X}_k, \bar{Y}_k\}$, we solve all $LP_{\xi_{k+1}, k+1}$ for each $\xi_{k+1} \in \Xi_{k+1}$ and denote the optimal dual variable (vector) to Constraint (33) be $\gamma_{\xi_{k+1}, k+1}$, then the single Benders' cut that may be added to $LP_{\xi_k, k}$ is given by

$$(\text{Benders'-cut}) \quad \theta \geq \sum_{\xi_{k+1} \in \Xi_{k+1}} p_{k+1}(\xi_{k+1}) \left[(\gamma_{\xi_{k+1}, k+1})^T (b_{\xi_{k+1}, k+1} - A_{\xi_{k+1}, k+1} X_k - B_{\xi_{k+1}, k+1} Y_k) \right].$$

Given a first stage scenario ξ_1 , we sample a random path $\xi_{[2, K]}$ with replacement and solve $LP_{\xi_k, k}$ following the order of $k = 2, \dots, K$, which we call a forward pass. Notice that for $LP_{\xi_K, K}$ we enforce $\theta = 0$ as we know $\alpha_{K+1}(X_K, Y_K) = 0$. Then we solve the set of problems $\{LP_{\xi_k, k} : \xi_k \in \Xi_k\}$ and add Bender's cut to the corresponding previous stage $\{LP_{\xi_{k-1}, k-1} : \xi_{k-1} \in \Xi_{k-1}\}$, following the order of $k = K, K-1, \dots, 2$, which we call a backward pass. It can be seen that the optimal value of first stage problem $LP_{\xi_1, 1}$ provides a deterministic lower bound $\underline{z}$ due to relaxation. The upper bound $\bar{z}$ in turn, can be estimated through the results of forward passes since each forward pass can be regarded as one Monte Carlo simulation. The uncertainty around this upper bound estimate can be measured in the following way: After N iterations, an approximated 95% confidential interval for the upper bound is given by $[\bar{u} - 1.96 \frac{\bar{s}}{\sqrt{N}}, \bar{u} + 1.96 \frac{\bar{s}}{\sqrt{N}}]$, where $\bar{u}$ and $\bar{s}$ are the mean and sample standard deviation of N sample objective values $\bar{z}$. Following the same procedure of Pereira and Pinto (1991), we use the uncertainty bound around the upper bound estimate as a convergence criterion and stop the algorithm while the lower bound $\underline{z}$ is within the confidential interval of the upper bound. The above procedure is repeated for each first stage scenario ξ_1 , with each provides a statistic upper bound $\bar{z}(\xi_1)$. The final upper bound estimate is obtained by

$$z^{\text{SDDP}} := \sum_{i \in I} (c_i^f \bar{n}_i \bar{x}_i + c_i^h \bar{n}_i^0) + \sum_{\xi_1 \in \Xi_1} p_1(\xi_1) \bar{z}(\xi_1).$$

We summarize the above algorithm in Algorithm.

In summary, we obtain the initial stage decision $\{\bar{x}_i, \bar{n}_i^0\}$ by solving Lagrangian dual problem via Algorithm 1, which at the same time provides a lower bound z^{LR} for the decision $\{\bar{x}_i, \bar{n}_i^0\}$. To further validate the decision, we calculate a stochastic upper bound z^{SDDP} via Algorithm. The optimality gap of our initial stage decision $\{\bar{x}_i, \bar{n}_i^0\}$ can be estimated as $\rho = \frac{(z^{\text{SDDP}} - z^{\text{LR}})}{z^{\text{LR}}}$. If the optimality gap ρ is too large (e.g., larger than 10%), we may continue the iteration procedure in Algorithm 1 with a larger maximum iteration number $r_{\max}$ or a smaller convergence criterion ϵ so that we can get another potentially better decision with relatively less effort by making the most of previous results. We repeat the above procedures until we get a decision $\{\bar{x}_i, \bar{n}_i^0\}$ with satisfactory optimality gap. We summarize the comprehensive solution algorithm in Algorithm.

5. Numerical experiments

In this section, we first show the performance of Algorithm 1 by comparing with a time-varying deterministic solution and the true optimal solution, based on a hypothetical data set. Then we validate the proposed model by analyzing a set of numerical cases with some parameters changed. To further reveal some managerial insights, we first apply the model to three typical demand patterns and then apply it to a real case study in New York City. We solve all models by Gurobi using a desktop computer with i5-6500 CPU @3.20 GHz and 16.0GB RAM.

5.1. Algorithm performance

We consider I to be five locations randomly generated from a 10-by-10 grid. To focus on the hourly demand fluctuation, we aggregate the planning horizon into one day in the following analysis; i.e., the time-varying distribution of demands are assumed to be identical over days. In particular, we set the planning horizon to be 18 h from 6:00 to 24:00 at a 30 min time step, i.e., $T = \{0, 1, 2, \dots, 36\}$. The service adjourns from 0:00 am to 6:00 am for necessary adjustment (e.g., finishing necessary relocation operations) and preparation (e.g., charging all EV to full for service next day). We further assume that each EV has five SOC levels, $E = \{0, 1, 2, 3, 4\}$, indicating a full charging cycle of two hours. Then we refine the operation decision process into $K = 3$ stages where each stage lasts for six hours. This refinement considers the trade-off of tractability and representability. We may use more stages if the computational time permits. The demand distribution is assumed to be stage-wise independent. We adopted the default value of some parameters based on the work of Li et al. (2016). The unit fixed cost (aggregated to one day) c^f is set to be \$25. The daily holding cost c^h is set to \$54.5, considering a life cycle cost of an EV to be \$83,000 and a service life of 5 years. We consider the case that customers are too impatient to wait for 30 min for the service and will transfer to other means of transportation immediately (i.e., $q = 0$), and therefore there will be no waiting queue over the system. The penalty cost for customer loss c^l is assumed to remain \$20 across the whole planning horizon. We also assume the relocation cost c^r for each EV between each two stations is fixed to be \$10 regardless of the distance. The default capacity of each stations $\bar{n}$ is assumed to be 20 vehicles. Moving capacity $\bar{v}$ is 10 vehicles. The time (energy cost) for customer traveling is normalized and rounded into one to three time intervals (levels of SOC) based on the range of Euclidean distances between the origin and the destination. The time (energy cost) for

Algorithm 2 SDDP Algorithm.

```

function SDDP( $\bar{x}_i, \bar{n}_i^0$ )
  Initialize:  $z^{\text{SDDP}} := 0$ 
  for  $\xi_1 \in \Xi_1$  do
    Step 0: Initial Set
    Set upper bound set  $\bar{z} := \emptyset$ , lower bound  $\underline{z} := 0$ , iteration  $N := 1$ ,
    Set estimated upper bound  $\bar{z}(\xi_1) := 0$ .
    Step 1: Forward Pass
    Randomly sample a scenario  $s = \xi_{[2,k]}^s$  with replacement;
    Solve  $\text{LP}_{\xi_{1,1}}$  and let  $X_1^*, Y_1^*, \theta_1^*$  be the optimal solution;
    for  $k = 2, \dots, K$  do
      Update  $\{X_{k-1}, Y_{k-1}\} = \{X_{k-1}^*, Y_{k-1}^*\}$  and solve  $\text{LP}_{\xi_{k,k}^s}$  to obtain  $\{X_k^*, Y_k^*\}$ ;
    end for
    Update  $\bar{z} := \bar{z} \cup \left\{ \sum_{k=1}^K f_k(X_k^*, Y_k^*) \right\}$ ;
    Update  $\underline{z} := f_1(X_1^*, Y_1^*) + \theta_1^*$ ;
    Step 2: Convergence Check
    Calculate mean  $\bar{u}$  and sample deviation  $\bar{s}$  of set  $\bar{z}$ 
    if  $\bar{u} - 1.96 \frac{\bar{s}}{\sqrt{N}} \leq \underline{z}$ 
      Set  $\bar{z}(\xi_1) := \sum_{k=1}^K f_k(X_k^*, Y_k^*)$ ;
    else
      go to Step 3.
    end if
    Step 3: Backward Pass
    for  $k = K, K-1, \dots, 2$  do
      for all scenario  $\xi_k$  in  $\Xi_k$  do
        Solve  $\text{LP}_{\xi_{k,k}}$  to obtain  $\gamma_{\xi_{k+1},k+1}$ ;
      end for
      Add corresponding Benders'-cut to  $\{\text{LP}_{\xi_{k-1,k-1}} : \xi_{k-1} \in \Xi_{k-1}\}$ ;
    end for
    Set  $N := N + 1$ ;
    Go to Step 1.
  end for
  Set  $z^{\text{SDDP}} := \sum_{i \in I} (c_i^f \bar{n}_i \bar{x}_i + c_i^h \bar{n}_i^0) + \sum_{\xi_1 \in \Xi_1} p_1(\xi_1) \bar{z}(\xi_1)$ ;
  return  $z^{\text{SDDP}}$ ;
end function

```

Algorithm 3 Comprehensive Solution Algorithm.

```

function main()
  Set step size:  $\sigma$ 
  Set maximum iteration number:  $r_{\max}$ ; convergence criterion:  $\epsilon$ 
   $(\bar{x}_i, \bar{n}_i^0, z^{\text{LR}}) = \text{PSA}(\sigma, \epsilon, r_{\max})$ 
   $z^{\text{SDDP}} = \text{SDDP}(\bar{x}_i, \bar{n}_i^0)$ 
  Calculate  $\rho := \frac{(z^{\text{SDDP}} - z^{\text{LR}})}{z^{\text{LR}}}$ 
  while  $\rho > 10\%$  do
    Update  $r_{\max}, \epsilon$ ;
     $(\bar{x}_i, \bar{n}_i^0, z^{\text{LR}}) = \text{PSA}(\sigma, \epsilon, r_{\max})$  (Continue with previous result)
     $z^{\text{SDDP}} = \text{SDDP}(\bar{x}_i, \bar{n}_i^0)$ 
    Update  $\rho := \frac{(z^{\text{SDDP}} - z^{\text{LR}})}{z^{\text{LR}}}$ 
  end while
  return  $\{\bar{x}_i, \bar{n}_i^0\}$ ;
end function

```

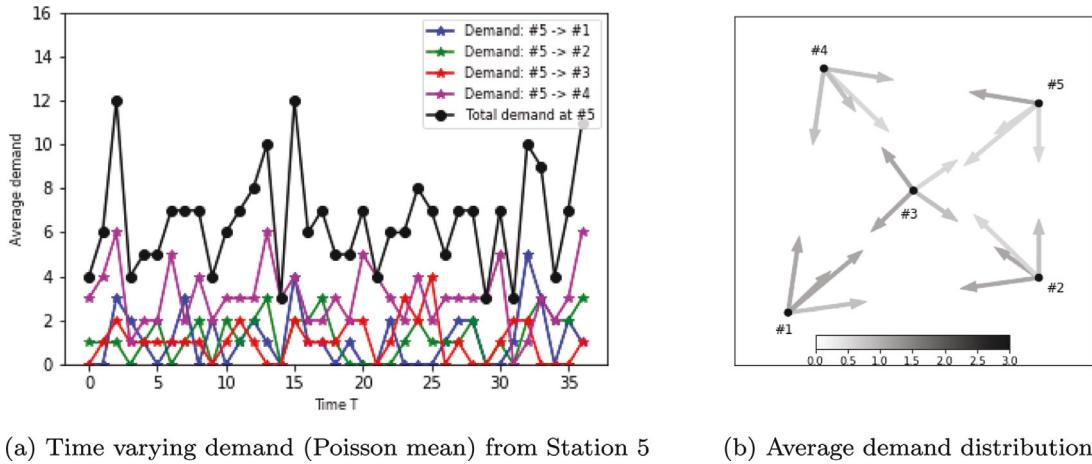


Fig. 3. Illustration for demand distribution.

Table 1
Results of the Algorithm 1.

Iteration No.	Lower Bound (ζ)	Gap (ε^2)	Accumulated CPU time (s)
1	2504	37,231	13
2	4829	12,031	31
3	6276	4,680	53
4	6562	1,648	73
5	6631	324	97
6	6645	89	120

Table 2
Comparison of different solution methods.

	LR	Mean	Mean(IP)	Opt
z	6703	6944	6952	6602
CPU time (s)	120	<1	306	12350
Gap	1.53%	5.2%	5.3%	0

relocation is assumed to be one time interval (level of SOC) for all origin-destination pairs. The hourly demand λ for each origin-destination pair follows time-varying Poisson distributions, with a predetermined time varying mean value within [0,3]. Fig. 3(a) shows the generated time-varying average demand pattern from Station 5 to all other stations as an example, and Fig. 3(b) shows the daily average demand across each origin-destination pair, where arrow directions indicate customer traveling directions, and the gray scale of arrow color indicates the average demand volume. We use Monte-Carlo sampling to generate 5 random scenarios for each stage, which can be expanded to 125 scenarios over the entire planning horizon, under stage-wise independence.

We solve the initial stage decision $X_0^{LR} := \{\tilde{x}_i, \tilde{n}_i^0\}$ based on Algorithm 1. After some warming trials, we set the step size as $\sigma = \frac{1}{10^6 N^2}$, where N is the number of current iteration, and the stopping criteria $\epsilon = 100$. Convergence criteria is met after 6 iterations with each iteration takes around 20 s. More details are shown in Table 1.

As a comparison, we solve the initial stage decision X_0^{Mean} for the time-varying deterministic case, in which only one scenario (obtained by calculating the average of 125 sampled scenarios) is considered. Second, we solve the deterministic case again but consider integer decisions (i.e., use Constraints (15) instead of (19)) and we can get another initial stage decision which we denote by $X_0^{Mean(IP)}$. Third, we directly solve the expanded SMSP to get the optimal initial stage decision X_0^{Opt} as well as the optimal objective value z^{Opt} . Further we directly solve the expanded SMSP fixing initial stage variables to X_0^{LR} , X_0^{Mean} and $X_0^{Mean(IP)}$ respectively, to get corresponding objective value z^{LR} , z^{Mean} and $z^{Mean(IP)}$. We compare the objective value, CPU running time for solving initial stage decision and optimality gap of each solution as shown in Table 2. First we analyze the effect of LP relaxation (Recall that in Section 4.2 we relax the integer variables related to fleet size and daily operation decisions and allow them to take any positive values). Under this relatively small demand (hourly OD pair demand is within [0, 3]), solving the deterministic integer case takes a much longer time (306 s) than solving the linear deterministic case (<1 s), but provides initial stage decision without too much difference (<1.53%). Our subgradient method still works under the integer setting, but we can only complete 2 iterations with a final gap (ε^2) of 30,695 within a 24-h running time. On the other hand, it only takes 120 s if we apply the LP relaxation. So we believe that it is reasonable to adopt the LP

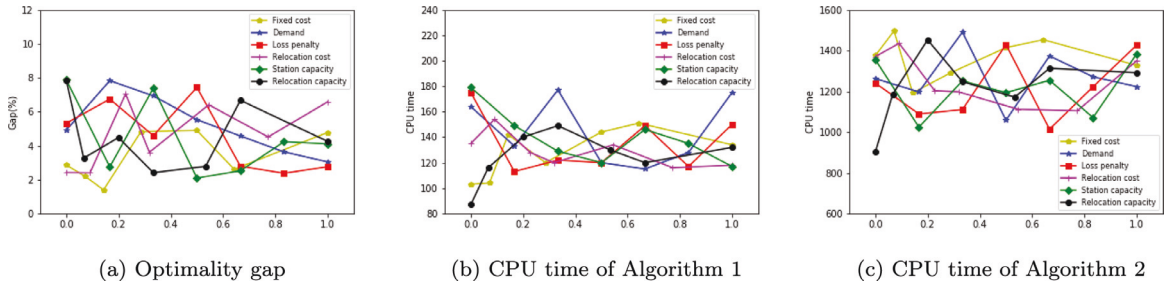


Fig. 4. Algorithm performance analysis.

relaxation for our dispatching and relocation problem. Next we show the performance of our solution. Comparing with the deterministic solution X_0^{Mean} (with an optimality gap of 5.2%), our stochastic model provides a better initial stage decision X_0^{LR} with an optimality gap of 1.53%. Our solution method (Algorithm 1) is also very efficient since it only takes 1% of time needed for obtaining the true optimum X_0^{Opt} . Our solution can be even more efficient when the scale of problem increases or more scenarios are considered, since its complexity grows linearly with the number of scenarios, and has a potential to take advantage of parallel computing.

5.2. Sensitivity analysis

We study several cases with each of the key parameters $\{c^f, \lambda, c^l, c^r, \bar{n}, \bar{v}\}$ changes from the default setting, while maintaining the rest. The change of each parameter is determined by $\eta \in [0, 1]$, with the following parameter settings $c^f = 5 + 70\eta$, $\lambda = (0.25 + 1.5\eta)\lambda_0$, $c^l = 5 + 30\eta$, $c^r = 3 + 22\eta$, $\bar{n} = 5 + 30\eta$, $\bar{v} = 0 + 15\eta$, where λ_0 indicates the generated benchmark demand. For each parameter we test 7 cases, and in all 42 cases are tested.

We obtain the initial planning decision by Algorithm 1, and then applying Algorithm to get corresponding optimality gap ρ as derived in Section 4.5. Changes of the optimality gap with each parameter are shown in Fig. 4(a). Overall, the gap ρ is around 4.5%. The optimality gap tends to decrease as the demand increases. CPU running times for each case with respect to η are shown in Fig. 4(b) and (c). The average running time to obtain initial planning decision using Algorithm 1 is 134 s and further to check optimality gap using Algorithm 2 is 1257 s.

Second, we investigate how the expected total cost as well as each cost component vary with the change of parameters. In particular, we consider the following cost terms based on the solution of final iteration in Algorithm 1: $f^b = \sum_{i \in I} c_i^f \bar{n}_i \bar{x}_i$ as the fixed cost for chargers installation, $f^h = \sum_{i \in I} c_i^h \bar{n}_i^0$ as the fleet holding cost, $f^l = \sum_{s \in S} p(s) \sum_{i, j \in I, t \in T} c_{ijt}^l \bar{v}_{ijt}(s)$ as the expected demand loss penalty cost, $f^r = \sum_{s \in S} p(s) \sum_{i, j \in I, t \in T} c_{ijt}^r \bar{v}_{ijt}(s)$ as the expected relocation cost, and the expected total cost $f := f^b + f^h + f^l + f^r$. The results are shown in Fig. 5(a)–(f).

From Fig. 5(a), when unit fixed cost c^f increases, most cost terms remain unchanged except the total fixed cost f^b , leading to an increase in the total cost f . Overall the total number of charging stations will reduce from 5 to 4 when unit fixed cost c^f exceeds 40.

From Fig. 5(b), when demand increases, all cost components increase except loss penalty cost f^l . In other words, to serve an area with higher demand, we need more chargers and more EVs, while at the same time we need to schedule more relocation operations. However, when there is too little demand, it may not be so worth to run such a business due to high infrastructure cost, and as a result, many customers cannot be served as reflected by relative high loss penalty f^l . Further, when the demand becomes too high, the loss penalty f^l increases immediately since the service capability of the system is limited by some other factors like station and relocation capacity.

From Fig. 5(c), when loss penalty cost c^l increases, most cost terms including the total cost increase, leading to an increase in total cost f . An interesting observation is that the total penalty for demand loss f^l decreases even though the unit penalty cost c^l increases. When c^l is low, performing relocation operations to increase the service level is not economic (f^r is almost 0), and when c^l is higher than 30, there will be no customer loss as more relocation is performed to satisfy demand as much as possible.

From Fig. 5(d), when relocation cost c^r increases, holding cost f^h decreases while other terms increase, leading to an overall increase in the total cost f . We notice that more charging stations should be built, as reflected by an increase in fixed cost f^b . However, when relocation cost c^r becomes higher than 20, relocation cost f^r immediately decreases to 0, indicating that at this point, relocation is not economic and just letting demand lose is a better choice (increasing f^l). Hence, reducing the relocation cost is very critical to improving the service level of the system.

From Fig. 5(e), when station capacity $\bar{n}$ increases, the total cost f decreases first mainly due to the decrease of loss penalty cost f^l , but then increases because the grow in the total fixed cost f^b exceeds the benefit from the improvement in the service level. When station capacity $\bar{n}$ is small ($\bar{n} \leq 15$), the increment in station capacity will lead to the increase in holding cost f^h and relocation cost f^r as more EVs can be deployed and thus relocated over the system. However, when

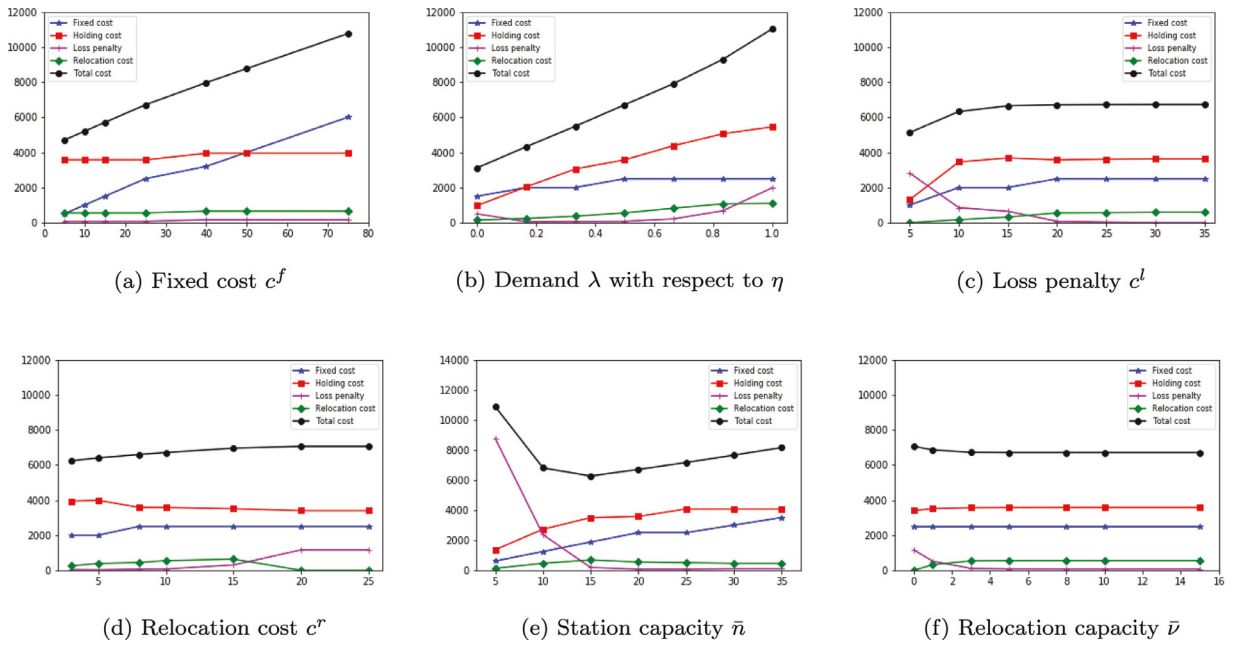


Fig. 5. Sensitivity analysis.

station capacity $\bar{n}$ exceeds 15, further increasing it will lead to a decrease in relocation cost f^r . The trend becomes stable when $\bar{n} \approx 25$, indicating the capacity of the system becomes redundant.

From Fig. 5(f) when relocation capacity $\bar{v}$ increases, while the total cost f almost remains unchanged, more relocations (increasing f^r) are preferred to provide a better service level (decreasing f^l). Further, it indicates that $\bar{v}$ greater than 8 is sufficient for the best service level.

5.3. Impact of demand pattern

In the previous section, although assumed time-varying, the daily variation of the mean customer flows, or the demand patterns, are relatively smooth. In this section, our purpose is to analyze the impact of different time-varying demand patterns on the specific system design. In particular, we consider two patterns: (i) daily commute pattern, where most customer trips are from residential areas to working areas in the morning and vice-versa in the afternoon; and (ii) event pattern, where most customers all go to a certain area at a certain time and then evacuate after a while. The daily commute pattern can be seen for most urban areas with obvious regional split of working and living functions, where the EV sharing system mainly provides commute transit service. This is expected to be seen when the system is either widely adopted or the distance of commute trips are short enough. The event pattern can be found in many scenarios, e.g., when the EV sharing system serves as a complimentary role in the existing transportation, which will be highly utilized when there is a demand surge due to special events. We will analyze the corresponding infrastructure design, including charger locations and initial deployment of EVs at each station, and real-time operation policies. The parameter settings follow the same as Section 5.1 except the demand pattern.

Although different cities have different commute patterns, we consider two representatives for illustration purposes: single-center city and twin-center city. In this setting, the residential areas are scattered around while the working areas are concentrated in one or two centers in the city. Based on the 5-station system as above, we consider 1 or 2 of them located at work places, indicated by “center” and “twin” respectively.

We first consider the single-center city commute pattern, where Station 3 is set to be the working area, while the rests are residential areas. To capture the commute feature, we divide the entire day into three time horizons: morning rush hours, afternoon rush hours, and the rest time. The morning rush hours are set from 6:00 am to 9:00 am, where the demands are mainly from residential to working areas, as shown in Fig. 6(a). The major commute trips are reflected in the plot with dark black arrows pointing from each of Stations 1, 2, 4, 5 to Station 3. The afternoon rush hours are set from 5:00 pm to 8:00 pm, where the demands are in the opposite directions, as shown in Fig. 6(b). The demand distribution at other hours are relatively smooth, as shown in Fig. 6(c). Under such a demand pattern, the corresponding optimal decision based on our algorithm is obtained as shown in Fig. 6(d)–(i). In these figures, the red circles indicate the locations of chargers, while arrows indicate the EV flows from a station to another. The width and color of arrows indicate the average flow and the SOC level (green: high SOC, yellow: medium SOC, red: low SOC) within the particular time horizon, respectively. Fig. 6(d)–(f) provide the EV transit flow for each time horizon. In addition, the bars with numbers near each station indicate

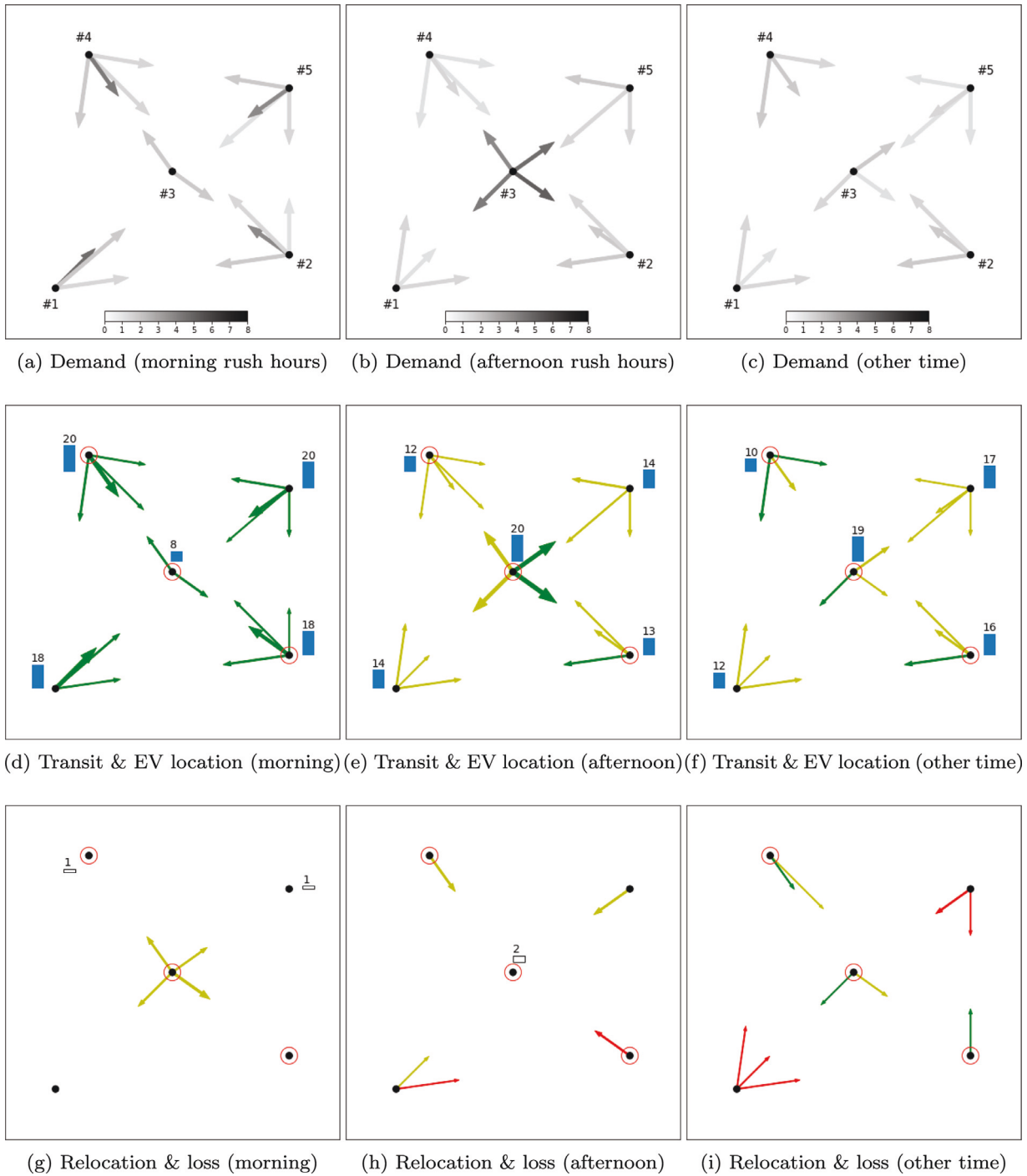


Fig. 6. Illustrations of results for daily commute pattern (single-center city).

the corresponding average EV distribution at the beginning of each time horizon. On the other hand, Fig. 6(g)–(i) show the EV relocation flow for each time horizon, while the hollow bars with numbers capture the average demand loss (rounded to integer) at each station.

The solution shows that chargers should be installed in the center, Station 3, as well as Stations 2 and 4. Initially, stations around in residential areas are fully deployed, as shown in Fig. 6(d), to satisfy the morning rush hour demand. Similarly, Fig. 6(e) shows Station 3 is fully deployed for the afternoon rush hour demand. Fig. 6(f) captures the EV concentration in Station 3 due to the huge EV returns at the end of morning rush hours. It can also be observed that the transit flow is

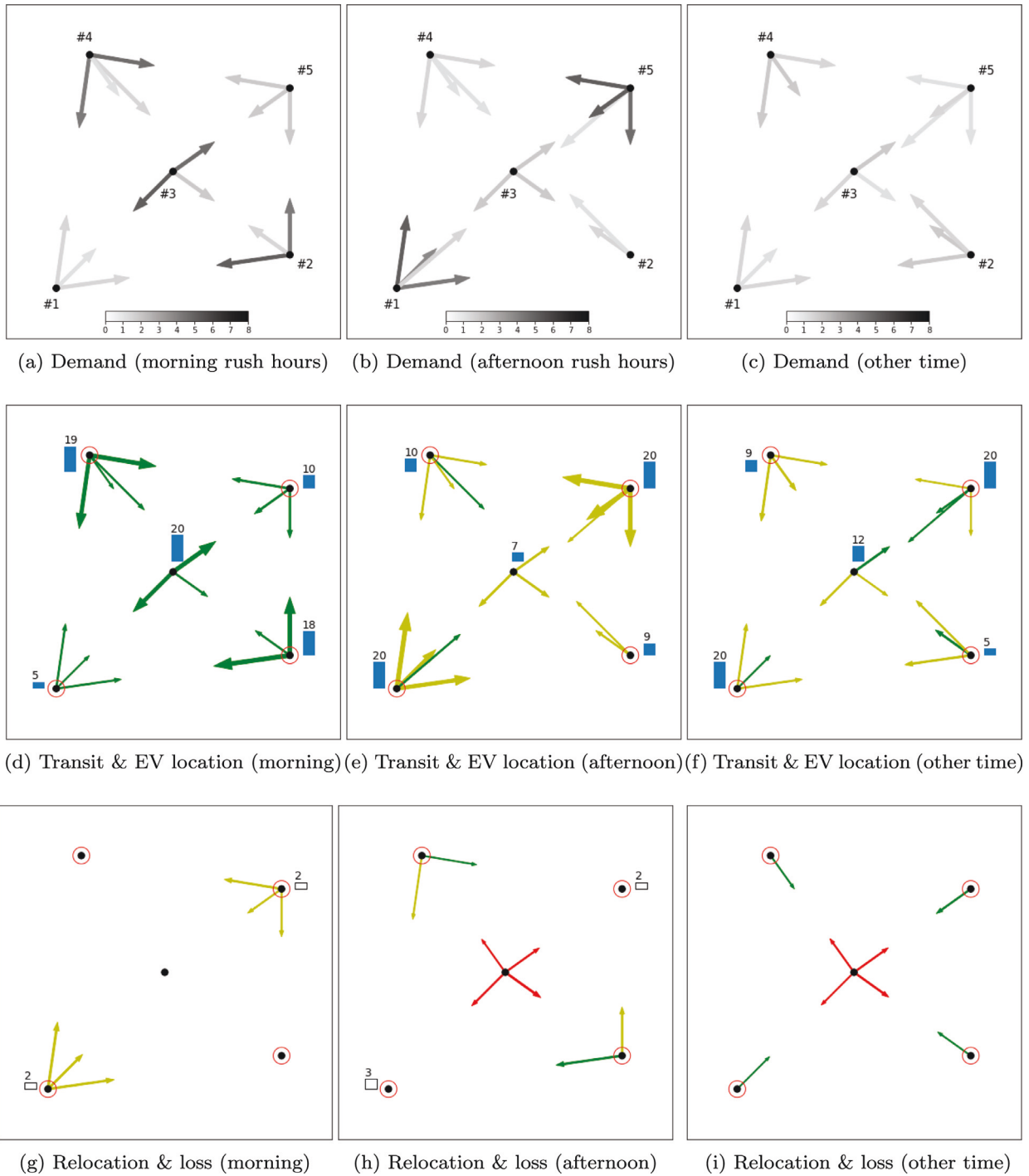


Fig. 7. Illustrations of results for daily commute pattern (twin-center city).

mainly with EVs with high and medium SOC levels. In Fig. 6(g)–(i), we can observe that many relocation flows are usually in the opposite directions from the transit flow for fleet balancing. In addition, EVs with low SOC levels are relocated to stations with chargers for charging. In the morning rush hours, due to huge commute flows towards Station 3, there are some customers unable to get service (about 2 in total per half hour as shown in Fig. 6(g)). Similarly, we also observed some demand loss during afternoon rush hours at Station 3 (2 per half hour as shown in Fig. 6(h)), which indicates a lower service level for customers returning home after working. Demand loss happens mainly due to the concentrated appearance

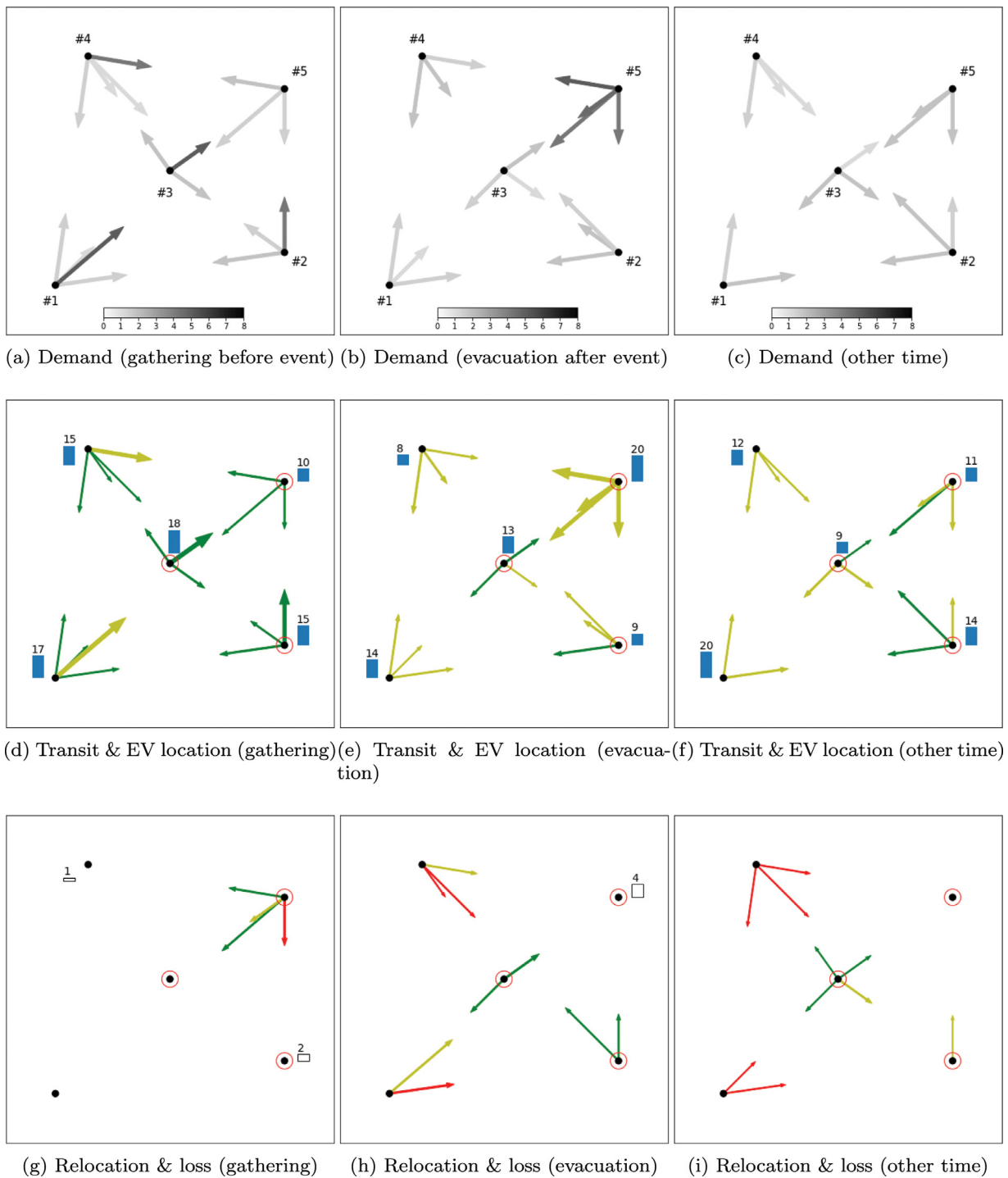


Fig. 8. Illustrations of results for event pattern.

of demand at stations with limited capacity, as well as lack of charging time. We also find that the system can provide a good service outside rush hours, where there is no delay observed, as shown in Fig. 6(i).

Now we consider the daily commute pattern in a twin-center city. In particular, we set Stations 1 and 5 as the working places and the rest as residential areas. Similar to the single-center city case, we show the result in Fig. 7. The revised demand pattern can be observed in Fig. 7(a)–(c).

Different from the single-center city, the optimal number of charging stations increases to four. The charging station previously installed at Station 3 are replaced by the new centers Station 1 and 5. Due to the dispersed flow pattern, more charging stations are necessary. From Fig. 7(d)–(e), we observe similar insights for transit flow and distribution of the fleet as the single-center city. However, we find an interesting result for relocation operations from Fig. 7(h) and (i), where Station 3 plays a role of hub due to its location as the geometric center.

Lastly, we consider the event pattern. To be more specific, assume that an event is held from 11:00 am to 1:00 pm near Station 5. So people tend to gather to Station 5 during 10:00 am to 11:00 am and evacuate from 1:00 pm to 2:00 pm. Therefore, we decompose the entire day into gathering, evacuation, and the rest, where the corresponding demand is designed as shown in Fig. 8(a)–(c).

Similarly, the results are shown in Fig. 8(d)–(i), where three charging stations are built at Stations 2, 3, and 5. The EV transit and relocation flows follow a similar distribution as the single-city commute pattern, hence the location of the event plays a similar role as the center of the city. However, since the event causes huge unbalanced demand at a very short time period, we observe a large demand loss. During the event gathering, the total demand loss is about 3 at two stations (as shown in Fig. 8(g)), while at the evacuation horizon, the demand loss at Station 5 is horribly 4 (as shown in Fig. 8(h)). This somewhat shows that the EV sharing system can handle relatively smooth demand fluctuation, but is vulnerable to huge demand surge, where there is a lack of time and space to relocate or charge the EV.

We summarize some useful insights based on the three demand patterns. First, more charging stations are necessary when the demand surge is more dispersed rather than concentrated. Second, EV flows for transit should be with higher SOC levels, while those EVs with low SOC levels should be relocated or charged. Last, with a known demand surge, the fleet should be redistributed along with the demand to avoid demand loss as much as possible.

5.4. Case study

In this section, we apply our model to solve a realistic problem, an EV sharing service planning in the Brooklyn area at the New York City. We obtain the demand information from released historical data of TLC green taxi trip records in December, 2017 (NYC Taxi and Limousine Commission, 2017), which includes the taxi service with pickup/drop-off time and location of each customer (differentiate via zones). We choose 20 zones as the distributed candidate stations I and their geographic locations are shown in Fig. 9(a) (zones marked in orange). To meet the service range of the EV sharing system, we select the subset of the trip data with both the pickup and drop-off locations among the 20 candidate stations as the potential demand. We similarly aggregate the planning horizon into 15 h from 7:00 am to 10:00 pm in each day, with a 15-min time step, i.e., $T = \{0, 1, 2, \dots, 60\}$. Further, we divide T into three equal stages, i.e., 7:00 am ~ 12:00 pm, 12:00 pm ~ 5:00 pm, 5:00 pm ~ 10:00 pm, and assume the demand is of perfect information within each stage. The hourly average demands between each origin-destination station pairs are shown in Fig. 9(b). It can be seen that Stations 181, 189 and 61 have relatively higher pickup demands, especially for Station 181. We consider a 3-h charging cycle from empty to full based on level II charging time (Yilmaz and Krein, 2012), which implies 13 different SOC levels based on the 15-min time step (including the empty level). However, we impose a safety (minimal) SOC level of 1-h charging over the system during operation as prepared for emergencies, and in this case we only consider a total of SOC levels of nine, i.e., $E = \{0, 1, 2, \dots, 8\}$, where SOC level 0 indicates the amount of electricity obtained by charging for 1 h. The average travel time between each two stations can be obtained from the data, and we rounded it above to fit it in our discrete framework. Other parameters

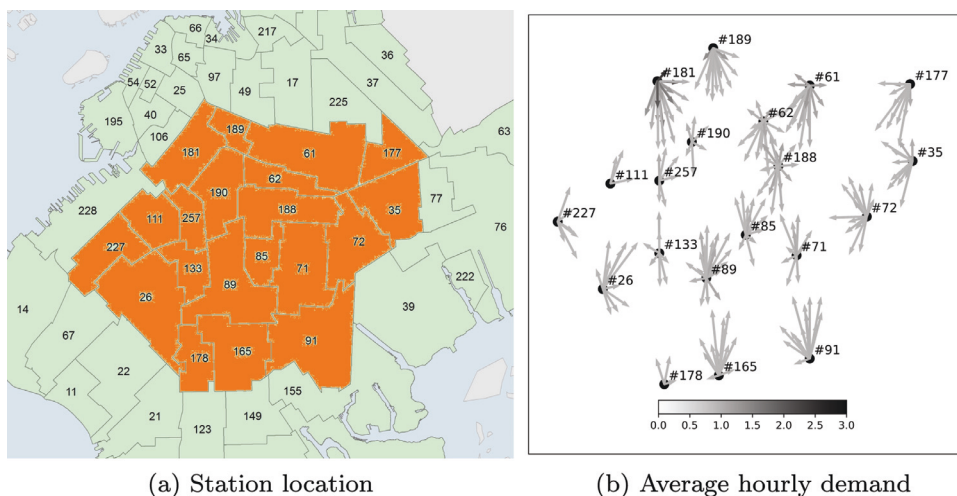


Fig. 9. Data input.

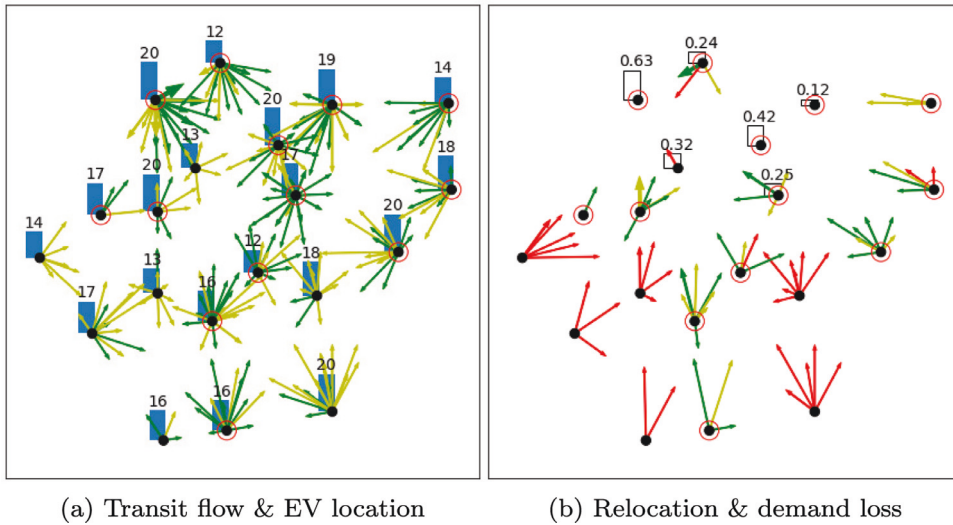


Fig. 10. Infrastructure and flow.

setting are assumed to remain the same as in Section 5.1. For each stage, we create 5 demand scenarios based on randomly selected historical data from different days at the same time.

It takes 4532 s to obtain the initial stage planning by Algorithm 1, and another 22,340 s to get an optimality gap of 6.3% by Algorithm. Following the same style as Section 5.3, results are shown in Fig. 10. Four charging stations should be built in all zones except for Stations 190, 227, 178, 91, 26, 133, 71 with a total number of 332 EVs. This yields an expected total cost f of \$29,834, including a fixed cost $f^b = \$6,500$, EV holding costs $f^h = \$18,094$, expected loss penalty cost $f^l = \$632$, and expected vehicle relocation cost $f^r = \$4,608$. The result indicates that chargers should be installed in areas with high pickup demand. This helps reduce the relocation cost for charging and increase the EV utilization.

The average transit flow and the corresponding SOC level are shown in Fig. 10(a), while the average relocation flow and demand loss are shown in Fig. 10(b). It can be seen that major transit flows are concentrated in the northern part of the city. Meanwhile, the relocation flow consists of two types, one type directing from the south to the north for charging (with low SOC levels), and the other type with a contrary direction for demand balancing (with high SOC levels). We also see a bunch of demand loss in the north (especially, station 181 and 62) due to the high in and out demand. To further improve the service level in these zones, we may think about increasing their capacity or expanding the team for relocation operations.

6. Conclusion

Deploying an EV sharing service is extremely challenging, especially for a one-way sharing service, which requires to handle a complicated dynamic system involving huge spatial-temporal demand uncertainties. This paper fills such gap with a quantitative framework for planning a one-way EV sharing system in an urban area. Under reasonable assumptions, we capture the complicated decision process through a multistage stochastic programming problem, which can be solved through Monte-Carlo sampling techniques. Due to the multistage nature, we are able to jointly cope with the long-term charging infrastructure decision together with the real-time fleet operation policies. To conquer the curse of dimensionality, we provide an accelerating solution method based on Lagrangian relaxation and the SDDP algorithm to obtain the operation policy while check the optimality gap to the optimum. We compare the solution of our model with that of deterministic model and show that our solution is better at the cost of slight longer computational time. The scale of the problem can be further extended at the cost of more computing resources, or by taking advantage of parallel computing. Fortunately, we are able to deliver meaningful insights through a series of numerical experiments. From the sensitivity analysis, we understand how the change in some critical factors will affect the different compositions of the total cost. For example, We find that the total number of EVs deployed is dominated by demand, and the number of charging stations is not very sensitive to loss penalty cost and relocation cost. Following the sensitivity analysis, we explore different demand patterns to obtain detailed insights on demand surge. We find that chargers should be installed at stations where average demand is high, but more chargers are necessary when the demand surge is dispersed in a larger area. Finally, in a case study of the New York City EV sharing system planning, we provide an infrastructure design and further show how to find the bottleneck of the system by identifying energy flows and the occurrence of demand loss over the system.

Serving as a building block, this paper also motivates the research in many other directions. We have to acknowledge the scale of the problem is still challenging under the current framework, due to the joint decision of the fleet moving and charging. We are actively exploring more effective acceleration techniques to reduce the computational burden. On the other

hand, currently the model only considers a demand pattern independent of the location. It is very interesting to investigate how the location of charging station, power grid integration, or the pricing scheme would affect the demand pattern.

Acknowledgments

This research is supported in part by the U.S. National Science Foundation through Grants CNS# 1637772 and CNS#1638355. The authors gratefully acknowledge the valuable comments from the editor and three anonymous reviewers that helped improve the paper.

References

- de Almeida Correia, G.H., Antunes, A.P., 2012. Optimization approach to depot location and trip selection in one-way carsharing systems. *Transp. Res. Part E* 48 (1), 233–247.
- Barrios, J., Godier, J., 2014. Fleet sizing for flexible carsharing systems: simulation-based approach. *Transp. Res. Record* (2416) 1–9.
- Berman, O., Larson, R. C., Fouska, N., 1992. Optimal location of discretionary service facilities. *Transp. Sci.* 26 (3), 201–211.
- Boyaci, B., Zografos, K.G., Geroliminis, N., 2015. An optimization framework for the development of efficient one-way car-sharing systems. *Eur. J. Oper. Res.* 240 (3), 718–733.
- Boyaci, B., Zografos, K.G., Geroliminis, N., 2017. An integrated optimization-simulation framework for vehicle and personnel relocations of electric carsharing systems with reservations. *Transp. Res. Part B* 95, 214–237.
- Boyd, S., Xiao, L., Mutapcic, A., 2003. Subgradient methods. In: *Lecture notes of EE392o*, Stanford University, Autumn Quarter, 2004, pp. 2004–2005.
- Brandstätter, G., Kahr, M., Leitner, M., 2017. Determining optimal locations for charging stations of electric car-sharing systems under stochastic demand. *Transp. Res. Part B* 104, 17–35.
- Bruglieri, M., Colorni, A., Luè, A., 2014. The vehicle relocation problem for the one-way electric vehicle sharing: an application to the milan case. *Procedia-Social Behav. Sci.* 111, 18–27.
- Carino, D.R., Kent, T., Myers, D.H., Stacy, C., Sylvanus, M., Turner, A.L., Watanabe, K., Ziemba, W.T., 1994. The russell-yasuda kasai model: an asset/liability model for a japanese insurance company using multistage stochastic programming. *Interfaces* 24 (1), 29–49.
- Chen, Z.-L., Powell, W.B., 1999. Convergent cutting-plane and partial-sampling algorithm for multistage stochastic linear programs with recourse. *J. Optim. Theory Appl.* 102 (3), 497–524.
- Correia, G.H.D.A., Jorge, D.R., Antunes, D.M., 2014. The added value of accounting for users flexibility and information on the potential of a station-based one-way car-sharing system: an application in lisbon, portugal. *J. Intell. Transp. Syst.* 18 (3), 299–308.
- Donohue, C.J., Birge, J.R., 2006. The abridged nested decomposition method for multistage stochastic linear programs with relatively complete recourse. *Algorithmic Oper. Res.* 1 (1).
- Fleten, S.-E., Kristoffersen, T.K., 2008. Short-term hydropower production planning by stochastic programming. *Comput. Oper. Res.* 35 (8), 2656–2671.
- Frade, I., Ribeiro, A., Gonçalves, G., Antunes, A., 2011. Optimal location of charging stations for electric vehicles in a neighborhood in lisbon, portugal. *Transp. Res. Record* (2252) 91–98.
- Garrick, D., 2016. Car2go switching electric cars to gas. <http://www.sandiegouniontribune.com/news/politics/sdut-car-share-car2go-fleet-gas-electric-2016mar16-story.html>. [Online; accessed 19-July-2017].
- George, D.K., Xia, C.H., 2011. Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. *Eur. J. Oper. Res.* 211 (1), 198–207.
- Hodgson, M.J., 1990. A flow-capturing location-allocation model. *Geog. Anal.* 22 (3), 270–279.
- Kall, P., Wallace, S.W., Kall, P., 1994. *Stochastic Programming*. Springer.
- Li, X., Ma, J., Cui, J., Ghiasi, A., Zhou, F., 2016. Design framework of large-scale one-way electric vehicle sharing systems: a continuum approximation model. *Transp. Res. Part B* 88, 21–45.
- Lim, M.K., Mak, H.-Y., Rong, Y., 2014. Toward mass adoption of electric vehicles: impact of the range and resale anxieties. *Manuf. Serv. Oper. Manage.* 17 (1), 101–119.
- Luè, A., Colorni, A., Nocerino, R., Paruscio, V., 2012. Green move: an innovative electric vehicle-sharing system. *Procedia-Social Behav. Sci.* 48, 2978–2987.
- Mak, H.-Y., Rong, Y., Shen, Z.-J.M., 2013. Infrastructure planning for electric vehicles with battery swapping. *Manage. Sci.* 59 (7), 1557–1575.
- Melkote, S., Daskin, M.S., 2001. An integrated model of facility location and transportation network design. *Transp. Res. Part A* 35 (6), 515–538.
- Millard-Ball, A., 2005. *Car-Sharing: Where and How it Succeeds*, Vol. 108. Transportation Research Board.
- Mulvey, J.M., Shetty, B., 2004. Financial planning via multi-stage stochastic optimization. *Comput. Oper. Res.* 31 (1), 1–20.
- Nair, R., Miller-Hooks, E., 2011. Fleet management for vehicle sharing operations. *Transp. Sci.* 45 (4), 524–540.
- NYC Taxi and Limousine Commission, 2017. *TLC Trip Record Data*.
- Pereira, M.V., Pinto, L.M., 1991. Multi-stage stochastic optimization applied to energy planning. *Math. Program.* 52 (1–3), 359–375.
- Shapiro, A., 2011. Analysis of stochastic dual dynamic programming method. *Eur. J. Oper. Res.* 209 (1), 63–72.
- Shapiro, A., Dentcheva, D., Ruszczyński, A., 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.
- Shen, Z.-J.M., Coullard, C., Daskin, M.S., 2003. A joint location-inventory model. *Transp. Sci.* 37 (1), 40–55.
- Van Slyke, R.M., Wets, R., 1969. L-Shaped linear programs with applications to optimal control and stochastic programming. *SIAM J. Appl. Math.* 17 (4), 638–663.
- Voelcker, J., 2016. BlueIndy electric car-sharing: after 9 months, how's it doing? http://www.greencarreports.com/news/1104668_blueindy-electric-car-sharing-after-9-months-hows-it-doing. [Online; accessed 19-July-2017].
- Wang, M., Weber, T., Darlington, T., 2005. Well-to-wheels analysis of advanced fuel/vehicle systems—a north american study of energy use, greenhouse gas emissions, and criteria pollutant emissions. *Well-to-Wheels Analysis of Advanced Fuel/Vehicle Systems: A North American Study of Energy Use, Greenhouse Gas Emissions, and Criteria Pollutant Emissions*.
- Wong, K., Wong, S.C., Yang, H., 2001. Modeling urban taxi services in congested road networks with elastic demand. *Transp. Res. Part B* 35 (9), 819–842.
- Worley, O., Klabjan, D., Sweda, T.M., 2012. Simultaneous vehicle routing and charging station siting for commercial electric vehicles. In: *Electric Vehicle Conference (IEVC), 2012 IEEE International*. IEEE, pp. 1–3.
- Xu, M., Meng, Q., Liu, Z., 2018. Electric vehicle fleet size and trip pricing for one-way carsharing services considering vehicle relocation and personnel assignment. *Transp. Res. Part B* 111, 60–82.
- Yao, Z., Lee, L.H., Jaruphongs, W., Tan, V., Hui, C.F., 2010. Multi-source facility location-allocation and inventory problem. *Eur. J. Oper. Res.* 207 (2), 750–762.
- Yilmaz, M., Krein, P., 2012. Review of charging power levels and infrastructure for plug-in electric and hybrid vehicles and commentary on unidirectional charging. In: *IEEE International Electrical Vehicle Conference*.