

Distributionally Robust Optimization Approaches for a Stochastic Mobile Facility Routing and Scheduling Problem

Karmel S. Shehadeh^{1,*}

^aDepartment of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

Abstract

We study a mobile facility (MF) routing and scheduling problem in which probability distributions of the time-dependent demand for MF services is unknown. To address distributional ambiguity, we propose and analyze two distributionally robust MF routing and scheduling (DMFRS) models that seek to minimize the fixed cost of establishing the MF fleet and maximum expected transportation and unmet demand costs over all possible demand distributions residing within an ambiguity set. In the first model, we use a moment-based ambiguity set. In the second model, we use an ambiguity set that incorporates all distributions within a 1-Wasserstein distance from a reference distribution. To solve DMFRS models, we propose a decomposition-based algorithm and derive lower bound and two-families of symmetry breaking inequalities to strengthen the master problem and speed up convergence. Finally, we present extensive computational experiments comparing the operational and computational performance of the proposed distributionally robust models and a stochastic programming model and drive insights into DMFRS.

Keywords: (R) Facilities planning and design, mobile facility, demand uncertainty, scheduling and routing, distributionally robust optimization

1. Introduction

A mobile facility (MF) is a facility capable of moving from one place to another, providing real-time service to customers in the vicinity of its location when it is stationary (Halper and Raghavan, 2011). In this paper, we study a mobile facility (MF) routing and scheduling problem with stochastic demand first introduced by Lei et al. (2014). Specifically, we seek to find the number of MFs to use in a given service region over a specified planning horizon and the route and schedule for each selected MF. Customers' demand for MF service at each demand node is time-dependent and random. In contrast to Lei et al. (2014), we consider the case when the probability distribution of the demand is unknown, and only a possibly small data on the demand may be available. The objective is to find MFs routing and scheduling decisions that minimize the fixed cost of establishing the MF fleet

*Corresponding author.

Email address: kshehadeh@lehigh.edu, kas720@lehigh.edu (Karmel S. Shehadeh)

and cost of assigning demands to the MFs (e.g., transportation), and cost of unsatisfied demand.

The concept of MF routing and scheduling is very different than conventional static facility location (FL) problem and conventional vehicle routing (VR) problems. In static FL problems, we usually consider opening facilities at fixed locations. Conventional VR problems aims at handling the movement of items between facilities (e.g., depots) and customers. MF is a “*facility-like-vehicle*” that behaves in a way similar to traditional facilities when they are stationary except that they can move from one place to another if necessary (Lei et al., 2014). Thus, the most evident advantage of MF over fixed facilities is their flexibility in moving to accommodate the change in the demand over time and location (Halper and Raghavan, 2011; Lei et al., 2014, 2016).

MFs are used in many applications ranging from cellular services, healthcare services, to humanitarian relief logistics. For example, light trucks with portable cellular stations can provide cellular service in areas where existing cellular network of base stations temporarily fails (Halper and Raghavan, 2011). Mobile clinics (i.e., customized MFs fitted with medical equipment and staffed by health professionals) can travel to rural and urban areas to provide various (prevention, testing, diagnostic) health services. Mobile clinics also offer an alternative healthcare delivery option when a disaster, conflict, or other events cause stationary healthcare facilities to close or stop operations. (Blackwell and Bosse, 2007; Brown-Connolly et al., 2014; Du Mortier and Coninx, 2007; Gibson et al., 2011; Oriol et al., 2009; Song et al., 2013). We have recently seen how mobile clinics played a significant role in providing drive-through COVID-19 testing sites or triage locations in response to the COVID-19 pandemic. In humanitarian relief logistics, MFs give relief organizations the ability to provide aid to populations dispersed in remote and dense areas.

MF operators often seek a tactical plan, including the MF fleet’s routes and time schedules, that reduces their fixed operating costs and maximizes demand satisfaction. The MF deployment, routing, and scheduling problem (MFRSP) is a challenging optimization problem for two primary reasons. First, customers demand is random and hard to predict, especially with limited data during the planning process. Second, even in a perfect world in which we know with certainty the amount of demand in each period, the deterministic MFRSP is NP-hard as it is reduced to the classic FL problem (Halper and Raghavan, 2011; Lei et al., 2014). Thus, the incorporation of demand variability increases the overall complexity of MFRSP.

Ignoring demand uncertainty may lead to sup-optimal MF routing and scheduling decisions and the inability to meet customers’ demand (shortage). Failure to meet customer demand may lead to adverse outcomes, especially in healthcare, as it impacts population health. It also impacts customers’ satisfaction and thus the reputation of MF service providers and may increase their operational cost (due to, e.g., outsourcing the excess demands to other providers). To model uncertainty, Lei et al. (2014) proposed the first a *priori* two-stage stochastic optimization model (SP) for MFRSP, which seeks optimal routing and scheduling decisions to minimize the total expected system-wide cost, where the expectation is taken with respect to an assumed known

probability distribution of customers' demand. Although attractive, the applicability of the SP approach is limited to the case in which we know the distribution of the demand or we have sufficient data to model it. In practice, it is unlikely that decision-makers can estimate the actual distributions of random demand accurately, especially with limited data during the planning process. (Basciftci et al., 2020; Lei et al., 2016; Liu et al., 2019). If we solve a model with a data sample from a biased distribution, then the resulting decisions may have a disappointing performance when implemented under the true distribution.

Distributionally robust optimization (DRO) is an another approach to model uncertainty when distributions of random parameters are hard to estimate and subject to uncertainty (i.e., ambiguous). In DRO, we assume that distribution resides in an ambiguity set (i.e., a family of all possible distributions that share common properties). We then optimize based on the distribution within this set, i.e., the probability distribution is a decision variable (Rahimian and Mehrotra, 2019). There are two primary advantageous of DRO. First, DRO allows uncertain variables to follow an arbitrary distribution defined in the ambiguity set. As such, DRO alleviates the unrealistic assumption of the decision-maker's complete knowledge of distributions. Second, various techniques have been developed to derive tractable DRO models of real-world problems such as static facility location and healthcare scheduling problems (see, e.g., Basciftci et al. (2020); Luo and Mehrotra (2018); Saif and Delage (2020); Shehadeh and Sanci (2021); Shehadeh and Tucker (2020); Wang et al. (2020); Wu et al. (2015)).

The ambiguity set a key ingredient of DRO models and has two primary types: moment-based and distance-based ambiguity sets. Most DRO studies employ moment-based ambiguity sets, consisting of all distributions sharing certain moments (e.g., mean). Asymptotic properties of moment-based DRO model cannot be guaranteed because the moment information represents descriptive statistics. Other recent DRO approaches define the ambiguity set by choosing a distance metric (e.g., Wasserstein metric) to describe the deviation from a reference (often empirical) distribution. The main advantage of distance-based DRO approaches is that they enable incorporating possibly small-size data in the ambiguity set and optimization, and they often enjoy asymptotic properties (Mevisen et al., 2013). Despite the potential advantages, there are no DRO approaches for the specific MFRSP that we study in this paper (see Section 2). Therefore, we aim to investigate and compare the value and performance of moment-based and Wasserstein distance-based DRO models to address distributional ambiguity of demand in this specific MFRSP.

1.1. Contributions

We propose an analyze two *distributionally robust mobile facility routing and scheduling* (DMFRS) models that search for optimal (1) number of MFs to use within a planning horizon, (2) a routing plan and a schedule for the selected MFs, i.e., the node that each MF is located at in each time period, (3) assignment of MFs to customers. Decisions (1)-(2) are planning (first-stage) decisions,

which cannot be changed in the short run. Conversely, the assignments of the demand are decided based on demand realization, and thus are second-stage decisions. The objective is to minimize fixed operating costs (i.e., cost of using MFs and traveling inconvenience cost) and the maximum expectation of transportation and unsatisfied demand costs over all distributions residing within an ambiguity set. In the first model (MAD-DMFRS), we use an ambiguity set that incorporate the mean, support, and mean absolute deviation (MAD). In the second model (W-DMFRS), we use an ambiguity set that incorporates all distributions within a 1-Wasserstein distance from a reference (e.g., empirical) distribution. Our main contributions are:

- To the best of our knowledge, and according to our literature review in Section 2, our paper is the first to addresses the distributional ambiguity of the demand in this MFRSP using DRO.
- We propose a decomposition-based algorithm to solve W-DMFRS and MAD-DMFRS. We derive valid lower bound inequalities to strengthen the master problem and improve convergence. We also derive two-families of symmetry breaking inequalities, which break symmetries in the solution space of the first-stage routing and scheduling decisions and thus improve computational time.
- We conduct extensive computational experiments comparing the computational and operational performances of W-DMFRS, MAD-DMFRS, and the classical SP approach. Our results demonstrate (1) how the DRO approaches have superior operational performance in terms of satisfying customers demand as compared to the SP approach, (2) MAD-DMFRS is more computationally efficient than W-DMFRS and SP, (3) MAD-DMFRS yield more conservative decisions than W-DMFRS, which often have a higher fixed cost but significantly lower operational (unmet demand and transportation) costs, (4) efficiency of the symmetry breaking and lower bound inequalities, (5) the trade-off between cost, number of MFs, MF capacity, and operational performance.

1.2. Structure of the paper

The remainder of the paper is structured as follows. In Section 2, we review the relevant literature. In Section 3, we formally define DMFRS and its formulations. In Section 4, we present our DMFRS-decomposition algorithm and strategies to improve convergence. In Section 5, we present our computational results. Finally, we draw conclusions and discuss future directions in Section 6.

2. Relevant Literature

There is limited literature on MF as compared to stationary facilities. However, as pointed out by Lei et al. (2014), MFRSP share some features with several well-studied problems, including Dynamic Facility Location Problem (DFLP), Vehicle Routing Problem (VRP), and the Covering Tour Problem (CTP). In this review, we briefly discuss the similarities and differences between

MFRSP and these problems. Given that we consider making decisions over a planning period, then MFRSP is somewhat similar to DFLP, which seeks to locate/re-locate facilities over a planning horizon. To mitigate the impact of demand fluctuation along the planning period, decision-makers may open new facilities and close or relocate existing facilities at a relocation cost (Albareda-Sambola et al. (2009); Antunes et al. (2009); Contreras et al. (2011); Drezner and Wesolowsky (1991); Jena et al. (2015, 2017); Van Roy and Erlenkotter (1982)). Most DFLPs assume that the relocation time is relatively short as compared to the planning horizon. In contrast, MFRSP takes into account the relocation time of MFs. In addition, each MF needs to follow a specific route during the entire planning horizon, which is not a requirement in DFLP.

In CTP, one seeks to select a subset of nodes to visit that can cover other nodes within a particular coverage (Current et al. (1985); Flores-Garza et al. (2017); Gendreau et al. (1997); Hachicha et al. (2000); Tricoire et al. (2012)). In contrast to MFRSP, CTP does not consider the variations of demand over time and assumes that the amount of demand to be met by vehicles is not related to the length of time the MF is spending at the stop. The VRP is one of the most extensively studied problems in operations research that has numerous applications and variants (Subramanyam et al., 2020). Both MFRSP and VRP consider the routing decisions of vehicles. However, MFRSP is different than VRP in the following ways (Lei et al., 2014). First, in MFRSP, we can meet customer demand by a nearby MF (e.g., cellular stations). In VRP, vehicles visit customers to meet their demands. Second, the amount of demand that an MF can serve at each location depends on the duration of the MF stay at each, which is a decision variable. In contrast, VRPs often assume a fixed service time. Finally, most VRPs require that each customer be visited exactly once in each route. In contrast, in MFRSP, customers may be visited zero or multiple times.

Next, we review MFRSP studies closely related to our work. Halper and Raghavan (2011) introduced the concept of MF and proposed a continuous-time formulation to model the maximum covering mobile facility routing problem for multiple facilities under deterministic settings. To solve their model, Halper and Raghavan (2011) proposes several computationally effective heuristic algorithms. In DMFRS, we consider uncertainty of demand distribution toward minimizing the average cost of the entire system. To avoid the challenges of dynamic and re-optimization approaches, Lei et al. (2014) propose the first *a priori* two-stage SP for MFRSP. Lei et al. (2014)’s SP seeks optimal first-stage routing and scheduling decisions to minimize the total expected system-wide cost, where the expectation is taken with respect to a known distribution of the demand. A priori optimization has a managerial advantage since it guarantees the regularity of service, which is beneficial for both customer and service provider. That is, a prior plane allow the customers to know when and where to obtain service and enable MF service providers to be familiar with routes and better manage their time schedule during the day. The applicability of the SP approach is limited to the case in which the distribution of the demand is fully known, or we have sufficient data to model it.

Paper	Modeling Uncertainty	Optimization Approach	Decisions		Objectives		Symmetry breaking
			1st-stage	2nd-stage	1st-stage	2nd-stage	
Lei et al. (2014)	Demand distribution is assumed known	SP: optimization is based on an assumed (known) distribution	(1) # of MFs to activate (2) routes and time schedule of active MFs	(1) amount of customer demand coveblack (2) unmet demand in each interval	(1) fixed cost of establishing MF fleet (2) traveling inconvenience cost	(1) cost of unmet demand (2) transportation or shipping cost	
Lei et al. (2016)	polyhedral uncertainty set based on maximum "positive" deviation from mean value	RO: optimization is based on worst-case scenario, i.e., maximum positive deviation of demand from its mean value in each time interval	(1) size of MF fleet (2) flow of MFs over arcs in each interval	(1) amount of customer demand coveblack in each time interval (2) unmet demand in each interval	(1) fixed cost of establishing MF fleet	cost of unmet demand	
Our paper	ambiguity set: a family of all possible demand probability distributions that share the same properties	DRO: optimization is based on the worst-case distribution residing in ambiguity set, i.e., demand's distribution is a decision variable	(1) # of MFs to "activate" (2) routes and time schedule of active MFs	(1) amount of customer demand served by MFs in each time period (2) unmet demand in each period	(1) fixed cost of establishing MF fleet (2) traveling inconvenience cost	(1) cost of unmet demand (2) transportation or shipping cost	✓

Notation: SP is stochastic programming, RO is robust optimization, DRO is distributionally robust optimization, MF is mobile facility

Figure 1: Theoretical Comparison between Lei et al. (2014), Lei et al. (2016), and our approach.

Robust optimization (RO) and distributionally robust optimization (DRO) are alternative techniques to model, analyze and optimize decisions under uncertainty and ambiguity (where the underlying distributions are unknown) (Zhen et al., 2021). RO assumes that the uncertain parameters can take on any value from within a pre-specified uncertainty set of possible outcomes with some structure (Bertsimas and Sim, 2004; Ben-Tal et al., 2015; Soyster, 1973). In RO, optimization is based on the worst-case scenario occurring within the uncertainty set. Notably, Lei et al. (2016) proposed the first and only two-stage RO approach for MF feet sizing and routing problem with demand uncertainty. Lei et al. (2016)'s model aims to find the MF fleet's size and routing decisions that minimize fixed cost of establishing the MF fleet (first stage), and a penalty cost for unmet demands (second stage). Optimization in Lei et al. (2016)'s RO model is based on the worst-case scenario of the demand occurring within a polyhedral uncertainty set. By focusing the optimization on the worst-case scenario, RO may lead to overly conservative and suboptimal decisions for other more-likely scenarios (Chen et al., 2020; Delage and Saif, 2018).

DRO models the uncertain parameters as random variables whose underlying probability distribution can be any distribution within a pre-defined ambiguity set. The ambiguity set is a family of all possible distributions characterized by some known properties of random parameters (Esfahani and Kuhn, 2018). In DRO, optimization is based on the worst-case distribution within this set. DRO is an attractive approach to model uncertainty with ambiguous distributions because: (1) it alleviates the unrealistic assumption of the decision-makers' complete knowledge of the distribution governing the uncertain parameters, (2) it is usually more computationally tractable than its SP and RO counterparts (Delage and Saif, 2018; Rahimian and Mehrotra, 2019), and (3) one can use minimal distributional information or small sample to construct the ambiguity set and then build DRO models. Rahimian and Mehrotra (2019) provide comprehensive survey of DRO literature.

Despite the potential advantages, there are no DRO approaches for the specific MFRSP that we study in this paper. Therefore, our paper is the first to propose and analyze two DRO models

based on a Wasserstein ambiguity set and mean, support, and MAD ambiguity set. In Figure 1, we provide a comparison between our approach and that of Lei et al. (2014) and Lei et al. (2016), which are the only papers that considered a MFRSP closely related to us. In contrast to Lei et al. (2016), we additionally incorporate the MF traveling inconvenience cost in the first-stage objective and the random transportation cost in the second-stage objective. Additionally, we optimize the system performance over all demand distributions residing within the ambiguity sets. Our master and sub- problems and lower bound inequalities have a different structure than those of Lei et al. (2016) due to the differences in the decision variables and objectives. Finally, we propose two-families of symmetry-breaking inequalities, which break symmetries in the solution space of the routing and scheduling decisions. These inequalities can improve the solvability of any formulation that uses the same routing and scheduling decisions. Lei et al. (2014) and Lei et al. (2016) did not address the issue of symmetry in MFRSP.

Finally, it worth mentioning that our work uses similar reformulation techniques in recent DRO static FL literature (see, e.g., Basciftci et al. (2020); Luo and Mehrotra (2018); Saif and Delage (2020); Shehadeh and Sanci (2021); Shehadeh and Tucker (2020); Wang et al. (2020); Wu et al. (2015) and references therein).

3. DMFRS Formulation and Analysis

In this section, we formally define DMFRS and its formulations. In Section 3.1, we present a two-stage formulation. In Section 3.3 and 3.4, we respectively present and analyze the MAD-DMFRS and W-DMFRS models

3.1. Definitions and Assumptions

As in Lei et al. (2014), we consider a fleet of M mobile facilities and define DMFRS on a directed network $G(V, E)$ with node set $V := \{v_1, \dots, v_n\}$ and edge set $E := \{e_1, \dots, e_m\}$. The sets $I \subseteq V$ and $J \subseteq V$ are the set of all customers points and the subset of nodes where MFs can be located, respectively. The distance matrix $D = (d_{i,j})$ is defined on E and satisfies the triangle inequality, where $d_{i,j}$ is a deterministic and time-invariant distance between any pair of nodes i and j . For simplicity in modeling, we consider a planning horizon of T identical time periods, and we assume that the length of each period $t \in T$ is sufficiently short such that, without loss of generality, all input parameters are the same from one time period to another (this is the same assumption made in Lei et al. (2014, 2016); Halper and Raghavan (2011)). The demand, $W_{i,t}$, of each customer i in each time period t is random. The probability distribution of the demand is unknown, and only a possibly small data on the demand may be available. We assume that we know the the mean μ and range $[\underline{W}, \overline{W}]$ of \mathbf{W} . Mathematically, we make the following assumption on the support of \mathbf{W} .

Assumption 1. The support set \mathcal{S} of \mathbf{W} in (1) is nonempty, convex, and compact.

$$\mathcal{S} := \left\{ \mathbf{W} \geq 0 : \underline{W}_{i,t} \leq W_{i,t} \leq \overline{W}_{i,t}, \forall i \in I, \forall t \in T. \right\} \quad (1)$$

We consider the following basic features of the DMFRS as in Lei et al. (2014): (1) each MF has all the necessary service equipment and can move from one place to another, (2) all MFs are homogeneous, providing the same service, and traveling at the same speed, (3) we explicitly account for the travel time of the MF in the model, and service time are only incurred when the MF is not in motion, (4) the travel time $t_{j,j'}$ from location j to j' is an integer multiplier of a single time period (Lei et al. (2014, 2016), and (5) the amount of demand to be served is proportional to the duration of the service time at the location serving the demand.

We consider a cost f for using an MF, which represents the expenses associated with purchasing or renting an MF, staffing cost, equipment, etc. Each MF has a capacity limit C , which represents the amount of demand that an MF can serve in a single time unit. Due to the random fluctuations of demand and limited capacity, there is a possibility that the MF fleet fail to satisfy customers' demand fully. We consider a penalty cost γ for each unit of unmet demand. This penalty cost can represent the opportunity cost for the loss of demands or expense for outsourcing the excess demands to other companies (Basciftci et al., 2020; Lei et al., 2016). Thus, maximizing demand satisfaction is an important objective that we incorporate in our model (Lei et al., 2014).

Given that an MF cannot provide service when in motion, it is not desirable to keep it moving for a long time to avoid losing potential benefits. On the other hand, it is not desirable to keep the MF stationary all the time, as this may also lead to losing the potential benefits of making a strategic move to locations with higher demands. Thus, the trade-off of the problem includes the decision to move or keep the MF stationary. We consider a traveling inconvenience cost α to discourage unnecessary moving in cases where moving would neither improve or degrade the total performance. As in Lei et al. (2014), we assume that α is much lower than other costs such that its impact over the major trade-off is negligible.

We assume that the service quality a customer receives from an MF is inversely proportional to the distance between the two to account for the “access cost” (this assumption is common in practice and in the literature, see, e.g., Ahmadi-Javid et al. (2017); Reilly (1931); Drezner (2014); Lei et al. (2016); Berman et al. (2003); Lei et al. (2014)). Accordingly, we consider a demand assignment cost that is linearly proportional to the distance between the customer point and the location of an MF, i.e., $\beta d_{i,j}$, where $\beta \geq 0$ represents the assignment cost factor per demand unit and per distance unit.

Given a set of MF, M , we seek to find: (1) the number of MFs to use within T , (2) a routing plan and a schedule for the selected MFs, i.e., the node that each MF is located at in each time period, (3) assignment of MFs to customers. Decisions (1)–(2) are first-stage decisions that we make before realizing the demand patterns. The assignment decisions (3) represent the recourse (second-stage)

actions in response to the first-stage decisions and the realizations of demand patterns (i.e., you cannot assign the demand to MF's before realizing the demand). The quality the objective is to minimize (1) the first-stage fixed operation costs, and (2) the expectation of the recourse assignment cost and unsatisfied demand penalty cost.

Notation: For $a, b \in \mathbb{Z}$, we define $[a] := \{1, 2, \dots, a\}$ and $[a, b]_{\mathbb{Z}} := \{c \in \mathbb{Z} : a \leq c \leq b\}$. The abbreviations “w.l.o.g.” and “w.l.o.o.” respectively represent “without loss of generality” and “without loss of optimality.” Table 1 summarizes all notation.

3.2. Stochastic Programming Model

For all $m \in M$, let binary variable $y_m = 1$ if MF m is permitted to use, and zero otherwise. For all $j \in J$, $m \in M$, and $t \in T$, let binary variable $x_{j,m}^t = 1$ if MF m stays at location j at period t . The feasible region \mathcal{X} of variables x and y is defined in (2).

$$\mathcal{X} = \left\{ x, y : \begin{array}{l} x_{j,m}^t + x_{j',m}^{t'} \leq y_m, \quad \forall t, m, j, j' \neq j, t' \in \{t, \dots, \min\{t + t_{j,j'}, T\}\} \\ x_{j,m}^t \in \{0, 1\}, y_m \in \{0, 1\}, \quad \forall j, m, t \end{array} \right\} \quad (2)$$

As defined in Lei et al. (2014), region \mathcal{X} represent: (1) the requirement that an MF can only be in service when it is stationary, (2) MF m at location j in period t can only be available at location $j' \neq j$ after a certain period of time depending on the time it takes to travel from location j to j' , $t_{j,j'}$, i.e., $x_{j',m}^{t'} = 0$ for all $j' \neq j$ and $t' \in \{t, \dots, \min\{t + t_{j,j'}, T\}\}$, and (3) MF m has to be in an active condition before providing service. We refer the reader to Appendix A for a detailed derivation of region \mathcal{X} . For all (i, j, m, t) , let variable $z_{i,j,m}^t$ represents the amount of demand of customer point i being served by MF m located at j in period t . Let variables $u_{i,t}$ represent the amount of unmet demand of node i in period t . Assuming that the distribution of demand $\xi := [W_{1,1}, \dots, W_{I,T}]^\top$, denoted by \mathbb{P}_ξ , is known, we formulate the following SP:

$$(\text{SP}) \quad Z^* = \min_{(y,x) \in \mathcal{X}} \left\{ \sum_{m \in M} f y_m - \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} \alpha x_{j,m}^t + \mathbb{E}_{\mathbb{P}}[Q(x, W)] \right\} \quad (3a)$$

where for a feasible $(y, x) \in \mathcal{X}$ and a realization of uncertain demand $\xi := [W_{1,1}, \dots, W_{I,T}]^\top$

$$Q(x, \xi) := \min_{z, u} \left(\sum_{j \in J} \sum_{i \in I} \sum_{m \in M} \sum_{t \in T} \beta d_{i,j} z_{i,j,m}^t + \gamma \sum_{t \in T} \sum_{i \in I} u_{i,t} \right) \quad (4a)$$

$$\text{s.t.} \quad \sum_{j \in J} \sum_{m \in M} z_{i,j,m}^t + u_{i,t} = W_{i,t}, \quad \forall i \in I, t \in T \quad (4b)$$

$$\sum_{i \in I} z_{i,j,m}^t \leq C x_{j,m}^t \quad \forall j \in J, m \in M, t \in T \quad (4c)$$

$$u_t \geq 0, z_{i,j,m}^t \geq 0, \quad \forall i \in I, j \in J, m \in M, t \in T \quad (4d)$$

Formulation (3) seeks first-stage decisions $(x, y) \in \mathcal{X}$ that minimize the sum of (1) fixed operating cost for using MF (first-term in (F.1a)), (2) traveling inconvenience cost (which is equivalent

Table 1: Notation.

Indices	
m	index of MF, $m = 1, \dots, M$
i	index of customer location, $i = 1, \dots, I$
j	index of MF location, $j = 1, \dots, J$
Parameters and sets	
T	planning horizon
M	number, or set, of MFs
J	number, or, set of locations
f	fixed operating cost
$d_{i,j}$	distance between any pair of nodes i and j
$t_{i,j'}$	travel time from j to j'
C	the amount of demand that can be served by an MF in a single time unit
$W_{i,t}$	demand at customer site i for each period t
$\underline{W}_{i,t}/\overline{W}_{i,t}$	lower/upper bound of demand at customer location i for each period t
γ	penalty of not satisfying demand
First-stage decision variables	
y_m	$\begin{cases} 1, & \text{if MF } m \text{ is permitted to use,} \\ 0, & \text{otherwise.} \end{cases}$
x_{jm}^t	$\begin{cases} 1, & \text{if MF } m \text{ stays at location } j \text{ at period } t, \\ 0, & \text{otherwise.} \end{cases}$
Second-stage decision variables	
$z_{i,j,m}^t$	amount of demand of customer point i being served by MF m located at j in period t
$u_{i,t}$	total amount of unmet demand of node i in period t

to maximization of profit of keeping MFs stationary whenever possible. α is the profit weight factor.), and (3) expected cost of recourse activity (assignment cost and penalty cost in (4a)). Constraints (4b) account for the amount of demand from each customer in each time period that is satisfied and the amount of demand that is not satisfied. Constraints (4c) respect the capacity of each MF. Finally, constraints (4d) specify feasible ranges of the decision variables.

3.3. DMFRS Over MAD Ambiguity

In this section, we assume that the distribution \mathbb{P}_ξ of ξ is unknown but belongs to an ambiguity set of possible distributions, which incorporates demand's mean values, Mean Absolute Deviations (MADs), and support set. As in Wang et al. (2019) and Wang et al. (2020), we use MAD as the dispersion or variability measure because it enables a linear and computationally attractive reformulation. Specifically, we let $\mu_{i,t} = \mathbb{E}_{\mathbb{P}}[W_{i,t}]$ and $\eta_{i,t} = \mathbb{E}_{\mathbb{P}}(|W_{i,t} - \mu_{i,t}|)$ respectively represent the mean value and MAD of the demand $W_{i,t}$ at node i in period t , for all $i \in I$ and $t \in T$. Using these notation and the support \mathcal{S} defined in (1), we define the the following ambiguity set

$$\mathcal{F}(\mathcal{S}, \mu, \eta) := \left\{ \mathbb{P} \in \mathcal{P}(\mathcal{S}) : \begin{aligned} & \int_{\mathcal{S}} d\mathbb{P} = 1 \\ & \mathbb{E}_{\mathbb{P}}[W_{i,t}] = \mu_{i,t}, \quad \forall i \in I, t \in T \\ & \mathbb{E}_{\mathbb{P}}(|W_{i,t} - \mu_{i,t}|) \leq \eta_{i,t}, \quad \forall i \in I, t \in T \end{aligned} \right\} \quad (5)$$

Where $\mathcal{P}(\mathcal{S})$ in $\mathcal{F}(\mathcal{S}, \mu)$ represents the set of distributions supported on \mathcal{S} with mean value μ and dispersion measure η . Using $\mathcal{F}(\mathcal{S}, \mu, \eta)$, we formulate MAD-DMFRS as

$$(\text{MAD-DMFRS}) \min_{(y,x) \in \mathcal{X}} \left\{ \sum_{m \in M} f y_m - \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} \alpha x_{jm}^t + \left[\sup_{\mathbb{P}_\xi \in \mathcal{F}(\mathcal{S}, \mu, \eta)} \mathbb{E}_{\mathbb{P}}[Q(x, \xi)] \right] \right\} \quad (6)$$

Formulation MAD-DMFRS in (6) seeks first stage decisions $(x, y) \in \mathcal{X}$ that minimizes the first stage cost and the maximum expectation of the second-stage recourse cost, over a family of distributions characterized by the ambiguity set $\mathcal{F}(\mathcal{S}, \mu, \eta)$. Here, \mathbb{P}_ξ is a decision variable.

3.3.1. Reformulation of MAD-DMFRS

In this section, we derive a solvable reformulation of the MAD-DMFRS formulation in (6). As we show in the proof of Proposition 1 in Appendix Appendix B, problem in (B.1) is equivalent to problem (7)

Proposition 1. *For any fixed $(y, x) \in \mathcal{X}$, problem (B.1) is equivalent to*

$$\min_{\rho, \psi \geq 0} \left\{ \sum_{t \in T} \sum_{i \in I} (\mu_{i,t} \rho_{i,t} + \eta_{i,t} \psi_{i,t}) + \max_{\mathbf{W} \in \mathcal{S}} \left\{ Q(x, \xi) + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + |W_{i,t} - \mu_{i,t}| \psi_{i,t}) \right\} \right\} \quad (7)$$

Note that $Q(x, \xi)$ is a minimization problem, and thus in (7) we have an inner max-min problem, which is not suitable to solve in standard solution methods. For a given first-stage solution $x \in \mathcal{X}$ and realization of ξ , $Q(x, \xi)$ is a feasible linear program (LP). The dual of $Q(x, \xi)$ is as follows

$$Q(x, \xi) = \max_{\lambda, v} \sum_{t \in T} \sum_{i \in I} \lambda_{i,t} W_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t \quad (8a)$$

$$\text{s.t. } \lambda_{i,t} + v_{j,m}^t \leq \beta d_{i,j}, \quad \forall i \in I, j \in J, m \in M, t \in T \quad (8b)$$

$$\lambda_{i,t} \leq \gamma, \quad \forall i \in I, t \in T \quad (8c)$$

$$v_{j,m}^t \leq 0, \quad \forall j \in J, t \in T \quad (8d)$$

where λ and v are the dual variables associated with constraints (4b) and (4c), respectively. It is easy to see that w.l.o.o $\lambda_{i,t} \geq 0$ for all $i \in I$ due to constraints (8c) and the objective of maximizing a positive and bounded variables ($W_{i,t}$ times $\lambda_{i,t}$). Additionally, $v_{j,m}^t \leq \min\{\min_i\{\beta d_{i,j} - \lambda_{i,t}\}, 0\}$ by constraints (8b) and (8d). Given the objective of maximizing a positive term $C x_{j,m}^t$ multiplied by $v_{j,m}^t$, then $v_{j,m}^t = \min\{\min_i\{\beta d_{i,j} - \lambda_{i,t}\}, 0\}$ in the optimal solution. Given that β , d , and λ are finite, then $v_{j,m}^t$ is finite. It follows that problem(8) is a feasible and bounded LP. Given that $W \in [\underline{W}_{i,t}, \overline{W}_{i,t}]$ by definition, in view of dual formulation (8), we can rewrite the inner maximization problem $\max\{\cdot\}$ in (7) as

$$\max_{\lambda, v, W, k} \left\{ \sum_{t \in T} \sum_{i \in I} \lambda_{i,t} W_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) \right\} \quad (9a)$$

$$\text{s.t. (8b) - (8d), } W_{i,t} \in [\underline{W}_{i,t}, \overline{W}_{i,t}], \quad \forall i \in I, \forall t \in T \quad (9b)$$

$$k_{i,t} \geq W_{i,t} - \mu_{i,t}, \quad k_{i,t} \geq \mu_{i,t} - W_{i,t}, \quad \forall i \in I, \forall t \in T \quad (9c)$$

Note that the objective function in (9) contains the interaction term $\lambda_{i,t} W_{i,t}$. To linearize formulation (9), we define $\pi_{i,t} = \lambda_{i,t} W_{i,t}$ for all $i \in I$ and $t \in T$. Also, we introduce the following

McCormick inequalities for variables $\pi_{i,t}$:

$$\pi_{i,t} \geq \lambda_{i,t} \underline{W}_{i,t}, \quad \pi_{i,t} \leq \lambda_{i,t} \overline{W}_{i,t}, \quad \forall i \in I, \forall t \in T \quad (10a)$$

$$\pi_{i,t} \geq \gamma \underline{W}_{i,t} + \overline{W}_{i,t}(\lambda_{i,t} - \gamma) \quad \pi_{i,t} \leq \gamma \overline{W}_{i,t} + \underline{W}_{i,t}(\lambda_{i,t} - \gamma), \quad \forall i \in I, \forall t \in T \quad (10b)$$

Given that $W_{i,t}$ and $\lambda_{i,t}$ are bounded, then (10a)-(10b) yield an exact reformulation of the term $\lambda_{i,t} W_{i,t}$, for all $i \in I$ and $t \in T$. Accordingly, for a fixed $x \in \mathcal{X}$, problem (9) is equivalent to the following mixed-integer linear program

$$\max_{\lambda, v, W, \pi, k} \left\{ \sum_{t \in T} \sum_{i \in I} \pi_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) \right\} \quad (11a)$$

$$\text{s.t. } (\lambda_{i,t}, v_{i,t}) \in \{(8b) - (8d)\}, \quad \pi_{i,t} \in \{(10a) - (10b)\} \quad (11b)$$

$$W_{i,t} \in [\underline{W}_{i,t}, \overline{W}_{i,t}], \quad k_{i,t} \geq W_{i,t} - \mu_{i,t}, \quad k_{i,t} \geq \mu_{i,t} - W_{i,t}, \quad \forall i \in I, \forall t \in T \quad (11c)$$

Combining the inner problem in the form of (11) with the outer minimization problems in (7) and (6), we derive the following equivalent reformulation of the MAD-DMRFS model in (6):

$$\min \left\{ \sum_{m \in M} f y_m - \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} \alpha x_{j,m}^t + \sum_{t \in T} \sum_{i \in I} [\mu_{i,t} \rho_{i,t} + \eta_{i,t} \psi_{i,t}] + \delta \right\} \quad (12a)$$

$$\text{s. t. } (y, x) \in \mathcal{X}, \quad \psi \geq 0 \quad (12b)$$

$$\delta \geq h(x, W) \quad (12c)$$

where $h(x, W) := \max_{\lambda, v, W, \pi, k} \left\{ \sum_{t \in T} \sum_{i \in I} \pi_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) : (8b) - (8d), (10a) - (10b), (11c) \right\}$.

Proposition 2. *For any feasible $x \in \mathcal{X}$, $h(x, W) < \infty$. Furthermore, $h(x, W)$ is a convex piecewise linear function in x (We refer to Appendix Appendix C for a detailed proof.)*

3.4. DMFRS Over 1-Wasserstein Ambiguity

In this section, we assume that the distribution \mathbb{P}_ξ of ξ belongs to an ambiguity sets based on the Wasserstein distance. Specifically, we construct an ambiguity set based on 1-Wasserstein distance, which often admits tractable reformulation in most real-world applications (see, e.g., Duque and Morton (2020); Hanasusanto and Kuhn (2018); Jiang et al. (2019); Saif and Delage (2020)). Suppose that random vectors ξ_1 and ξ_2 follow \mathbb{F}_1 and \mathbb{F}_2 , respectively, where probability distributions \mathbb{F}_1 and \mathbb{F}_2 are defined over the common support \mathcal{S} . The Wasserstein distance between \mathbb{F}_1 and \mathbb{F}_2 , $d(\mathbb{F}_1, \mathbb{F}_2)$ represents the cost of an optimal transportation plan for moving from \mathbb{F}_1 to \mathbb{F}_2 , where the cost of moving from ξ_1 to ξ_2 equals to the norm $\|\xi_1 - \xi_2\|$. Mathematically,

$$d(\mathbb{F}_1, \mathbb{F}_2) := \inf_{\Pi \in \mathcal{P}(\mathbb{F}_1, \mathbb{F}_2)} \left\{ \int_{\mathcal{S}^2} \|\xi_1 - \xi_2\| \Pi(d\xi_1, d\xi_2) \right\} \quad \left. \begin{array}{l} \Pi \text{ is a joint distribution of } \xi_1 \text{ and } \xi_2 \\ \text{with marginals } \mathbb{F}_1 \text{ and } \mathbb{F}_2, \text{ respectively} \end{array} \right\} \quad (13)$$

where $\mathcal{P}(\mathbb{F}_1, \mathbb{F}_2)$ is the set of all probability distributions supported on \mathcal{S} with marginals \mathbb{F}_1 and \mathbb{F}_2 . We assume that \mathbb{P}_ξ is only observed via a possibly small finite set $\{\hat{\xi}^1, \dots, \hat{\xi}^N\}$ of N i.i.d. samples, which may come from the historical realizations of the demand. Accordingly, we consider the following 1-Wasserstein ambiguity set

$$\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon) = \left\{ \mathbb{P}_\xi \in \mathcal{P}(\mathcal{S}) : d(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N) \leq \epsilon \right\} \quad (14)$$

$\mathcal{P}(\mathcal{S})$ is the set of all distributions supported on \mathcal{S} , $\hat{\mathbb{P}}_\xi^N = \frac{1}{N} \sum_{n=1}^N \delta_{\xi^n}$ is the empirical distribution of ξ based on the N i.i.d samples, and $\epsilon > 0$ is the radius of the ambiguity set. The set $\mathcal{F}_p(\hat{\mathbb{P}}_\xi^N, \epsilon)$ represent a Wasserstein ball of radius ϵ centered at the empirical distribution $\hat{\mathbb{P}}_\xi^N$. When $\epsilon = 0$, the ambiguity set contains the empirical distribution and the DRO problem reduces to an SP. A larger radius ϵ indicates that we seek more robust solutions. Using $\mathcal{F}_p(\hat{\mathbb{P}}_\xi^N, \epsilon)$, we formulate W-DMFRS:

$$(\text{W-DMFRS}) \hat{Z}(N, \epsilon) = \min_{(y, x) \in \mathcal{X}} \left\{ \sum_{m \in M} f y_m - \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} \alpha x_{j,m}^t + \left[\sup_{\mathbb{P}_\xi \in \mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)} \mathbb{E}_{\mathbb{P}_\xi} [Q(x, \xi)] \right] \right\} \quad (15)$$

Formulation (15) searches for first stage decisions $(x, y) \in \mathcal{X}$ that minimize the first stage cost and the maximum expectation of the second stage cost over all distributions residing in $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)$. The Wasserstein metric measures the distance between true distribution and empirical distribution and can recover the true distribution when the data sample's size goes to infinity.

3.4.1. Reformulation of W-DMFRS

In the this section, we derive an equivalent solvable reformulation of the W-DMFRS model in (15). First, we consider the inner maximization problem $\sup [\cdot]$ in (15) for a fixed $x \in \mathcal{X}$. In Proposition 3, we present an equivalent dual formulation of this problem (See Appendix Appendix D for a detailed proof).

Proposition 3. *Problem $\sup_{\mathbb{P}_\xi \in \mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)} \mathbb{E}_{\mathbb{P}_\xi} [Q(x, \xi)]$ in (15) is equivalent to*

$$\inf_{\rho \geq 0} \left\{ \epsilon \rho + \left[\frac{1}{N} \sum_{n=1}^N \sup_{\xi \in \mathcal{S}} \{Q(x, \xi) - \rho |\xi - \hat{\xi}^n|\} \right] \right\} \quad (16a)$$

Formulation (16) is potentially challenging to solve because it require solving N non-convex optimization problems. Fortunately, given that the support of ξ is rectangular and finite (Assumption 1) and $Q(x, \xi)$ is feasible and bounded for every x and ξ , we next show that we can recast these problems as LPs for each feasible ρ and $x \in \mathcal{X}$. First, using the dual formulation of $Q(x, \xi)$, we rewrite the inner problem $\sup\{\cdot\}$ in (16) for each n as follows

$$\max_{\lambda, v, W} \left\{ \sum_{t \in T} \sum_{i \in I} \lambda_{i,t} W_{i,t} - \rho |W_{i,t} - W_{i,t}^n| + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t \right\} \quad (17a)$$

$$\text{s.t. } (\boldsymbol{\lambda}, \mathbf{v}) \in \{(8b) - (8d)\}, \mathbf{W} \in [\underline{\mathbf{W}}, \overline{\mathbf{W}}] \quad (17b)$$

Seconded, using the same techniques in Section 3.3.1 and Wang et al. (2019), we define an epigraphical random variable $\eta_{i,t}^n$ for the term $|W_{i,t} - W_{i,t}^n|$. Then, using variables $\boldsymbol{\eta}$, $\pi_{i,t} = \lambda_{i,t} W_{i,t}$, and inequalities (10a)-(10b) for variables $\pi_{i,t}$, we derive the following equivalent reformulation of (17)

$$\max_{\lambda, v, W, \pi} \left\{ \sum_{t \in T} \sum_{i \in I} \pi_{i,t} - \rho \eta_{i,t}^n + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t \right\} \quad (18a)$$

$$\text{s.t. } (\boldsymbol{\lambda}, \mathbf{v}) \in \{(8b) - (8d)\} \quad (18b)$$

$$\pi_{i,t} \in \{(10a) - (10a)\}, \quad W_{i,t} \in [\underline{W}_{i,t}, \overline{W}_{i,t}] \quad \forall i, \forall t \quad (18c)$$

$$\eta_{i,t}^n \geq W_{i,t} - W_{i,t}^n, \quad \eta_{i,t}^n \geq W_{i,t}^n - W_{i,t} \quad \forall i, \forall t \quad (18d)$$

Third, combining the inner problem in the form of (18) with the outer minimization problems in (16) and (15), we derive the following equivalent reformulation of the W-DMFRS model in (15)

$$\begin{aligned} \hat{Z}(N, \epsilon) = \min_{(y, x) \in \mathcal{X}, \rho \geq 0} & \left\{ \sum_{m \in M} f y_m - \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} \alpha x_{j,m}^t + \epsilon \rho \right. \\ & \left. + \frac{1}{N} \sum_{n \in N} \max \left\{ \sum_{t \in T} \sum_{i \in I} \pi_{i,t} - \rho \eta_{i,t}^n + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t : (18b) - (18d) \right\} \right\} \quad (19a) \end{aligned}$$

Using the same techniques in the proof of Proposition 2, one can easily verify that function $[\max_{\lambda, v, W} \{ \sum_{t \in T} \sum_{i \in I} \pi_{i,t} - \rho \eta_{i,t}^n + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t \}] < \infty$ and is a convex piecewise linear function in x .

4. Solution Method

In this section, we present a decomposition-based algorithm to solve the two-stage MAD-DMFRS formulation in (12), and strategies to improve the solvability of the formulation. The algorithmic steps for solving the W-DMFRS in (19) are similar. In Section 4.1, we present our decomposition Algorithm. In sections 4.2 and 4.3, we respectively derive lower bound and two-families of symmetry breaking inequalities to strengthen the master problem and speed up convergence.

4.1. DMFRS-Decomposition Algorithm

Proposition 2 suggests that constraint (12c) describes the epigraph of a convex and piecewise linear function of decision variables in formulation (12). Therefore, given the two-stage characteristics of MAD-DMFRS in (12), it is natural to attempt to solve problem (12) via a separation-based decomposition algorithm. Algorithm 1 presents DMFRS-decomposition algorithm, and the algorithm for the W-DMFRS in 19 has the same steps. Algorithm 1 is finite because we identify a new piece of the function $\max_{\lambda, v, W, \pi, k} \{ \sum_{t \in T} \sum_{i \in I} \pi_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) \}$ each time when the set $\{L(x, \delta) \geq 0\}$ is augmented in step 4, and the function has a finite number of pieces according to Proposition 2.

Algorithm 1: DMFRS-decomposition algorithm.

1. Input. Feasible region \mathcal{X} ; set of cuts $\{L(x, \delta) \geq 0\} = \emptyset$; $LB = -\infty$ and $UB = \infty$.

2. Master Problem. Solve the following master problem

$$Z = \min \left\{ \sum_{m \in M} f y_m - \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} \alpha x_{j,m}^t + \sum_{t \in T} \sum_{i \in I} [\mu_{i,t} \rho_{i,t} + \eta_{i,t} \psi_{i,t}] + \delta \right\} \quad (20a)$$

$$\text{s. t.} \quad (x, y) \in \mathcal{X}, \quad \psi \geq 0, \quad \{L(x, \delta) \geq 0\} \quad (20b)$$

and record an optimal solution (x^*, ρ^*, ψ^*) and optimal value Z^* . Set $LB = Z^*$.

3. Sub-problem. With (x, ρ, ψ) fixed to (x^*, ρ^*, ψ^*) , solve the following problem

$$h(x, W) = \max_{\lambda, v, W, \pi, k} \left\{ \sum_{t \in T} \sum_{i \in I} \pi_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) \right\} \quad (21a)$$

$$\text{s.t.} \quad (8b) - (8d), (10a) - (10b), (11c) \quad (21b)$$

record optimal solution $(\pi^*, \lambda^*, W^*, v^*, k^*)$ and $h(x, W)^*$. Set $UB = \min\{UB, h(x, W)^* + (LB - \delta^*)\}$

4. if $\delta^* \geq \sum_{t \in T} \sum_{i \in I} \pi_{i,t}^* + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^{t*} v_{j,m}^{t*} + \sum_{t \in T} \sum_{i \in I} -(W_{i,t}^* \rho_{i,t}^* + k_{i,t}^* \psi_{i,t}^*)$ **then**

stop and return x^* and y^* as the optimal solution to DMFRS formulation (20)

else add the cut $\delta \geq \sum_{t \in T} \sum_{i \in I} \pi_{i,t}^* + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m} v_{j,m}^{t*} + \sum_{t \in T} \sum_{i \in I} -(W_{i,t}^* \rho_{i,t} + k_{i,t}^* \psi_{i,t})$ to the set of cuts $\{L(x, \delta) \geq 0\}$ and go to step 2.

end if

4.2. Multiple Optimality Cuts and Lower Bound Inequalities

In this section, we aim to incorporate more second-stage information into the first stage without adding optimality cuts into the master problem by exploiting the structural properties of the recourse problem. We first observe that once the first-stage solutions and the demand are known, the second-stage problem can be decomposed into independent sub-problems with respect to time periods. Accordingly, we can construct cuts for each sub-problem in step 4. Let δ_t represent the optimality cut for each period t , we replace δ in (20a) with $\sum_t \delta_t$, where

$$\delta_t \geq \sum_{i \in I} \pi_{i,t}^* + \sum_{j \in J} \sum_{m \in M} C x_{j,m} v_{j,m}^{t*} + \sum_{i \in I} -(W_{i,t}^* \rho_{i,t} + k_{i,t}^* \psi_{i,t}) \quad \forall t \in T \quad (22)$$

The original single cut is the summation of multiple cuts of the form, i.e., $\delta = \sum_{t \in T} \delta_t$. Hence, in each iteration, we incorporate more or at least an equal amount of information into the master problem using (22) as compared with the original single cut approach. In this manner, the optimality cuts become more specific, which may result in better lower bounds and, therefore, a faster convergence. In Proposition 4, we further identify valid inequalities for each time period to tighten the master problem (see Appendix E for a proof).

Proposition 4. *Inequalities (23) are valid lower bound inequalities for DMFRS.*

$$\delta_t \geq \sum_{i \in I} \min\{\gamma, \min_{j \in J}\{\beta d_{i,j}\}\} \underline{W}_{i,t}, \quad \forall t \in T \quad (23)$$

4.3. Symmetry-Breaking Inequalities

Suppose there are three homogeneous MFs. As such, solutions $y = [1, 1, 0]^\top$, $y = [0, 1, 1]^\top$, and $y = [1, 0, 1]^\top$ are equivalent (i.e., yield the same objective) in the sense that they all permit 2 out of 3 MFs to be used in the planning period. To avoid wasting time exploring such equivalent solutions, we assume that MFs are numbered sequentially and add inequalities (24) to the the first stage.

$$y_{m+1} \leq y_m \quad \forall m < M. \quad (24)$$

enforcing arbitrary ordering or scheduling of MFs. Second, recall that in the first period, $t = 1$, we decide the initial locations of the MFs. Therefore, it doesn't matter which MF to assign to location j . For example, suppose that we have three candidate locations, and MFs 1 and 2 are active. Then, feasible solutions $(x_{1,1}^1 = 1, x_{3,2}^1 = 1)$ and $(x_{1,2}^1 = 1, x_{3,1}^1 = 1)$ yield the same objectives. To avoid exploring such equivalent solutions, we define a dummy location $J + 1$ and add the following inequalities to the the first stage.

$$x_{j,m}^1 - \sum_{j' > j}^{J+1} x_{j',m+1}^1 \leq 0, \quad \forall m < M, \forall j \in J \quad (25a)$$

$$x_{J+1,m}^1 = 1 - \sum_{j=1}^J x_{j,m}^1, \quad \forall m \in M. \quad (25b)$$

Inequalities are valid for any formulation that uses the same sets of first-stage routing and scheduling decisions. We derived inequalities (24)–(25) based on similar symmetry breaking principles in Denton et al. (2010), Ostrowski et al. (2011), and Shehadeh et al. (2019).

5. Computational Experiments

The primary objective of our computational study is to compare the operational and computational performance of the proposed DRO models and a sample average approximation (SAA) model of the SP. The sample average model solves model (3) with \mathbb{P}_ξ replaced by an empirical distribution based on N samples of ξ (see Appendix F for the formulation). For notational convenience, hereafter, we use M-DRO and W-DRO to denote the W-DMFRS and MAD-DMFRS models, respectively. In Section 5.1, we describe the set of test instances and discuss other experimental setups. In Section 5.2, we compare solution time of the three models. In Section 5.3, we demonstrate efficiency of the lower bound and symmetry breaking inequalities (23)–(25). We compare the performance of the optimal solutions of the models in Section 5.4. In Section 5.5, we analyze the reliability of the three models. We close by analyzing the sensitivity of the DRO models to different parameter settings in Section 5.6.

Table 2: DMFRS instances. Notation: I is # of customers, J is # of locations, and T is # of periods.

Inst	I	J	T
1	10	10	10
2	10	10	20
3	15	15	10
4	15	15	20
5	20	20	10
6	20	20	20
7	25	25	10
8	25	25	20
9	30	30	10
10	30	30	20

5.1. *Experimental Setup*

We construct 10 DMFRS instances, in part based on the same parameters settings and assumption made in Lei et al. (2014) and Lei et al. (2016), who addressed MF location problems. We summarize our test instances in Table 2. Each of the 11 DMFRS instances is characterized by the number of customers locations I , number of candidate locations J , and the number of periods T . Instances 1–4 are from Lei et al. (2014) and instances 5–10 are from Lei et al. (2016).

For each DMFRS instance, we generate a total of I vertices as uniformly distributed random numbers on a 100 by 100 plane and compute the distance between nodes in Euclidean sense as in Lei et al. (2014). We follow the same procedures in the DRO scheduling and facility location literature to generate random parameters as follows. For most instances, we randomly generate the mean values $\mu_{i,t}$ of the demand of each customer i in period t from a uniform distribution $U[\underline{W}, \overline{W}] = [20, 60]$ and the standard deviation $\sigma_{i,t} = 0.5\mu_{i,t}$. We sample N realizations $W_{i,t}^n, \dots, W_{I,T}^n$, for $n = 1, \dots, N$, by following lognormal (LogN) distributions with the generated $\mu_{i,t}$ and $\sigma_{i,t}$. We round each parameter to the nearest integer. We solve the SAA and W-DRO models using the N sample and the MAD-DRO model with the corresponding mean, MAD, and range. We compute the best radius in W-DRO model using the same iterative procedure in Jiang et al. (2019).

We assume that all cost parameters are calculated in terms of present monetary value. Specifically, as in Lei et al. (2014), for each instance we randomly generate (1) the fixed cost from a uniform distribution $U[a, b]$ with $a = 1000$ and $b = 1500$, (2) the assignment cost factor per unit distance per unit demand β from $U[0.0001a, 0.0001b]$, and (3) the penalty cost per unit demand γ from $U[0.01a, 0.01b]$. Finally, we set the traveling inconvenience cost factor to $0.0001a$, and unless stated otherwise, we use a capacity parameter $C = 100$. We implemented the SP model, DRO models, and decomposition algorithm using AMPL2016 Programming language calling CPLEX V12.6.2 as a solver with default settings. We run all experiments on MacBook Pro with Intel Core i7 processor, 2.6 GHz CPU, and 16 GB (2667 MHz DDR4) of memory. We imposed a solver time limit of 1 hour.

5.2. *CPU Time*

In this section, we compare solution times of the models. We focus on DMFRS instances where the sample size is possibly small, which is often seen in most real-world applications (especially in

Table 3: Computational details of solving MAD-DRO model. Results marked with * are obtained with $\epsilon = 2\%$.

Inst	I	T	$W \in [20, 60], C = 100$						$W \in [50, 100], C = 100$					
			CPU time			iteration			CPU time			iteration		
			Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
1	10	10	1	4	6	3	12	16	2	2	3	3	4	4
2	10	20	4	10	13	5	13	18	3	5	7	3	4	5
3	15	10	20	30	51	9	15	24	9	14	19	4	5	7
4	15	20	29	35	44	8	40	122	26	42	55	4	5	7
5	20	10	39	74	147	9	26	69	5	55	82	4	5	6
6	20	20	147	352	872	11	30	75	61	85	134	3	4	4
7	25	10	267	624	933	22	47	60	58	140	249	4	6	9
8	25	20	802	1253	1790	12	23	40	512	1024	1776	4	5	7
9*	30	10	277	1259	1802	12	27	33	237	456	737	5	6	8
10*	30	20	133	1354	1775	2	13	38	500	1246	1631	1	4	6

Table 4: Computational details of solving W-DRO model.

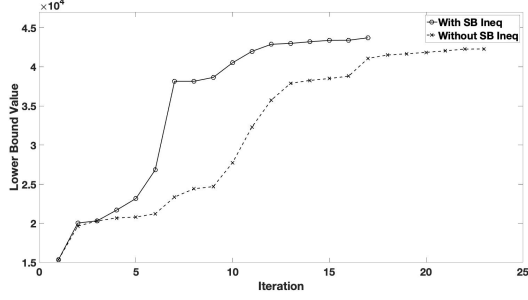
Inst	I	T	$W \in [20, 60], C = 100$						$W \in [50, 100], C = 100$					
			CPU time			iteration			CPU time			iteration		
			Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
1	10	10	13	21	32	13	16	19	9	14	19	8	9	10
2	10	20	26	57	87	15	16	17	42	81	151	10	17	24
3	15	10	72	91	137	22	27	38	27	40	47	7	12	15
4	15	20	77	90	105	17	20	22	117	195	287	14	22	27
5	20	10	68	97	138	14	19	26	92	169	252	10	16	20
6	20	20	201	432	745	14	16	18	212	339	474	5	7	8
7	25	10	646	1099	2080	43	54	69	185	256	330	6	8	10

healthcare) and is the main motivation of our DRO models. In addition to the base-case settings $\mathbf{W} \in [20, 60]$, we consider $\mathbf{W} \in [50, 100]$. For each of the 10 DMFRS instances and demand range, we randomly generate and solve 10 instances of MAD-DRO, W-DRO, and SAA as described in Section 5.1. That is, we generate and solve 200 MAD-DRO, W-DRO, and SAA instances.

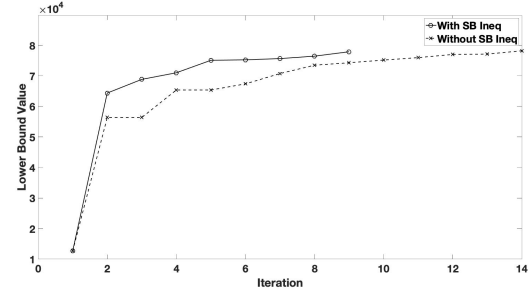
First, we present details of solving the MAD-DRO and W-DRO instances using our DMFRS-decomposition algorithm. Table 3 and Table 4 presents the minimum (Min), average (Avg), and maximum (Max) CPU time (in seconds) and number of iterations taken to solve these instances via the DMFRS-decomposition algorithm. From these tables, we first observe that the computational effort (i.e., time per iteration) tend to increase as the scale ($I \times J \times T$) of the instances increases.

Second, we observe that the W-DRO model takes a longer time to solve each instance than the MAD-DRO model. The average solution times of MAD-DRO instances 1–7 range from 2 second to 10 minutes. The average solution times of larger instances, instances 8–10, ranges from 8 to 23 minutes. In contrast, the average solution times of the W-DRO instances 1–7 range from 14 seconds to 18 minutes. And the W-DRO cannot solve instances 8-10 within the limit limit. It makes sense that solution times of the W-DRO model are longer than the MAD-DRO model because the MAD-DRO model is a deterministic and smaller (i.e., it has fewer variables and constraints) model.

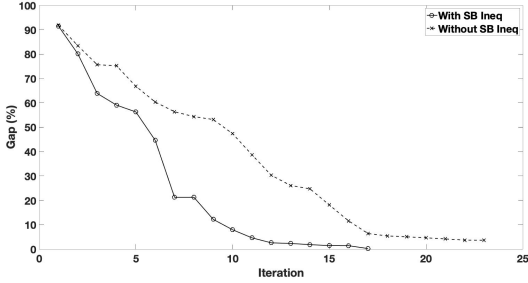
Finally, it is worth mentioning that solution times of W-DRO increase with N , and solutions times of the SAA are approximately the same as W-DRO’s solutions times. Due to page number limitations, we do not present these results.



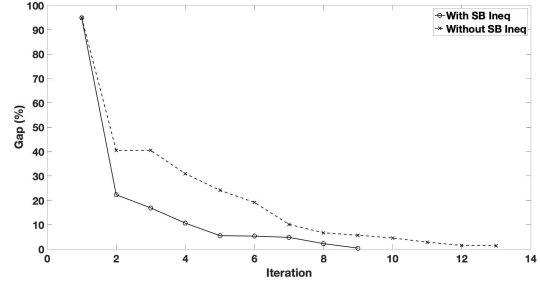
(a) LB values, W-DRO



(b) LB value, MAD-DRO



(c) Gap values, W-DRO



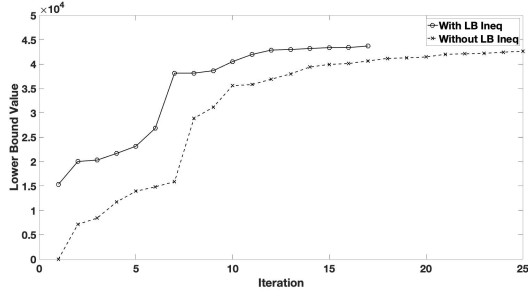
(d) Gap values, W-DRO

Figure 2: Comparisons of lower bound and gap values with and without SB inequalities (24)–(25).

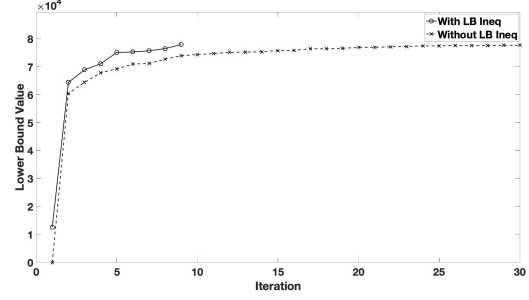
5.3. Efficiency of Inequalities (23)–(25)

In this Section, we study the efficiency of symmetry breaking inequalities (24)–(25) and lower bounding inequalities (23). Given the challenges of solving DMFRS instances without these inequalities, we use Instance 1 in this experiment for illustrative purposes. First, we separately solve the models with and without (w/o) these symmetry-breaking (SB) inequalities (24)–(25). We observe that without these SB inequalities, solution times of (W-DRO, MAD-DRO, SP) significantly increase from (15, 4, 9) to (1765, 1003, 3600) seconds. Instance 3–10 terminated with a large gap after one hour without these SB inequalities. As shown in Figure 2, both the lower bound and gap (i.e., the relative difference between the upper and lower bounds on the objective value) converge faster when we include inequalities (24)–(25) in the master problem. Moreover, inequalities (24)–(25) lead to a stronger bound in each iteration. These results demonstrate the importance of breaking the symmetry in the first-stage decisions and the effectiveness of our SB inequalities.

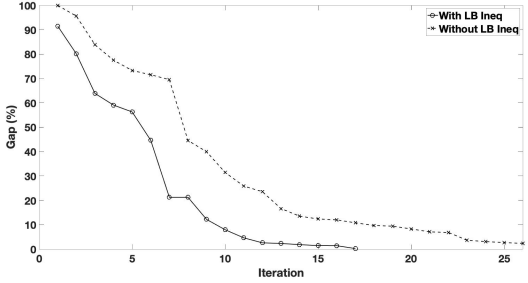
Next, we analyze the impact of including the lower bounding (LB) inequalities (23) in the master problem of DMFRS–decomposition algorithm. We first note the algorithm takes hundreds of iterations and a longer time until convergence without these LB inequalities. Therefore, in Figure 3, we present the LB and gap values with and without inequalities (23) from the first 25 iterations. It is obvious that both the lower bound and gap values converge faster when we introduce inequalities (23) into the master problem. Moreover, because of the better bonding effect, the algorithm converges to the optimal solution in fewer iterations and shorter solution times. For



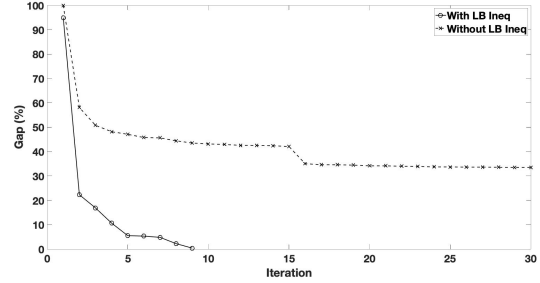
(a) LB values, W-DRO



(b) LB values, MAD-DRO



(c) Gap values, W-DRO



(d) Gap values, MAD-DRO

Figure 3: Comparisons of lower bound and gap values with and without valid LB inequalities (23).

example, the algorithm takes 10 seconds and 9 iterations to solve the MAD-DRO instance with these inequalities and terminates with a 33% gap after an hour without these inequalities. The results in this section demonstrate the importance and efficiency of inequalities (23)–(25).

5.4. Solutions Quality

In this section, we compare how the DRO and SP models perform when the underlying uncertainty distributions are perfectly specified and misspecified. For illustrative purposes and brevity, we use instance 3 ($I = 15$, $J = 15$, and $T = 10$) as an example of an average DMFRS instance. We observe similar results for other instances. We test out-of-sample performance (i.e., the objective value obtained by simulating the optimal solution of a model under a larger unseen data) of these models as follows. First, we sample data sets of sizes $N \in \{10, 50, 100\}$ from the in-sample logN distribution as described in section 5.1. Second, using each data set N , we solve an instance of W-DRO, an instance of MAD-DRO, and an instance of SAA of SP. Third, we fix the optimal first-stage decisions x yielded by each model in the second-stage of the SP. Then, we solve the second-stage recourse problem in (4) using first-stage decisions and the following two sets of $N' = 10,000$ out-of-sample data of $W_{i,t}^n$, for all $i \in I, t \in T$, and $n \in [N']$, to compute the corresponding second-stage cost.

Set 1. *Perfect distributional information*: we use the same settings and distributions (LogN) that we use for generating the N optimization sample to generate the N' data points. This is to

Table 5: Optimal Solutions

Model	$N = 10$	$N = 50$	$N = 100$
SP	7	8	8
W-DRO	9	9	8
MAD-DRO	10	10	10

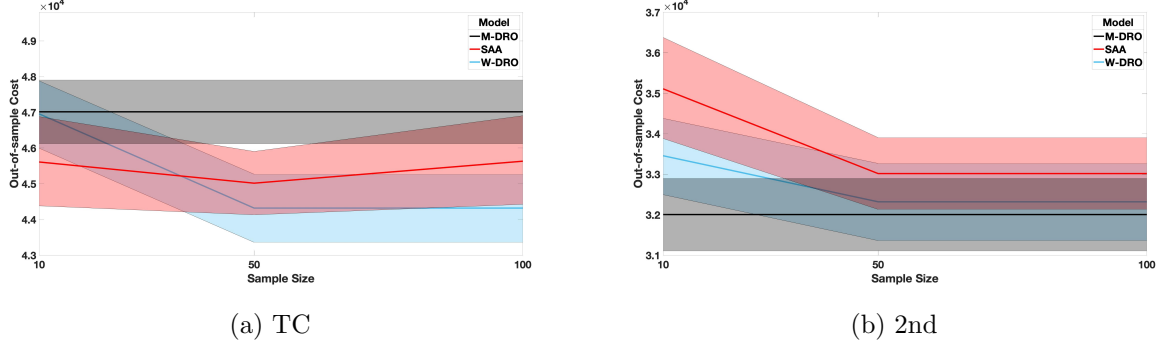


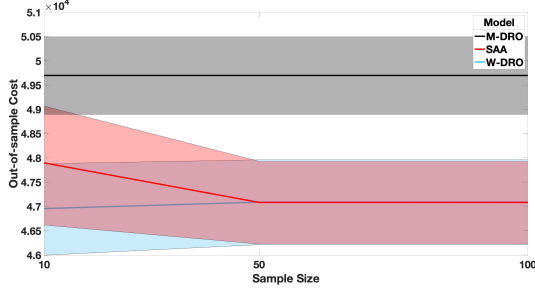
Figure 4: Out-of-sample performance under perfect information, LogN

simulate model’s performances when the true distribution is the same as the one used in the optimization.

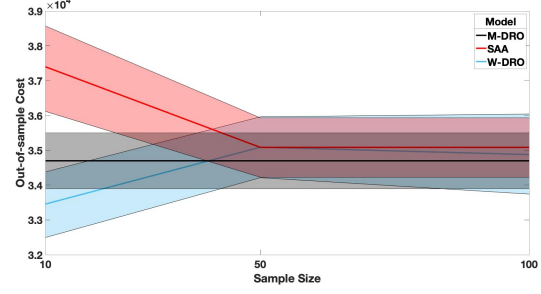
Set 2. *Misspecified distributional information*: we follow the same out-of-sample simulation testing procedure described in Wang et al. (2020) and employed in Shehadeh and Tucker (2020) to generate the $N' = 10,000$ data. Specifically, we perturb the distribution of the demand by a parameter Δ and use a parameterized uniform distribution $U[(1 - \Delta)\underline{W}, (1 + \Delta)\overline{W}]$, where a higher value of Δ corresponds to a higher variation level. We apply $\Delta \in \{0, 0.25, 0.5\}$ with $\Delta = 0$ indicating that we only vary the demand distribution. This is to simulate performance when the true distribution is different from the one we used on the optimization.

Table 6 presents the optimal solution yielded by each model. We first observe from this table that MAD-DRO always activates (schedules) a higher number of MFs to serve customers during the planning horizon. By scheduling more MFs, MAD-DRO tends to conservatively mitigate the ambiguity of the demand (reflected by better out-of-sample unmet demand costs). Second, the W-DRO model schedules a higher number of MFs than the SP model, and the difference is significant when the sample size is small. This makes sense as a small sample does not have sufficient distributional information, and thus the W-DRO makes robust/conservative decisions.

Next, we analyze how these optimal solutions perform via out-of-sample simulation using the N' data sets. Figures 4 presents the mean (line) and the (shaded) area between the 20% and 80% quantiles of the out-of-sample costs as a function of N with perfect distributional information (under Set 1). We make the following observations from this figure. The MAD-DRO model yields a higher average and upper quantile total cost (TC) than the W-DRO and SP because it schedules more MF’s and thus yields a higher fixed first-stage (and one time) cost. When the sample size is small (i.e., $N = 10$), the W-DRO and MAD-DRO models have a higher total cost than the SP due

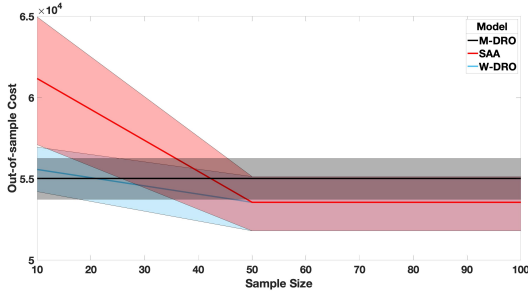


(a) TC

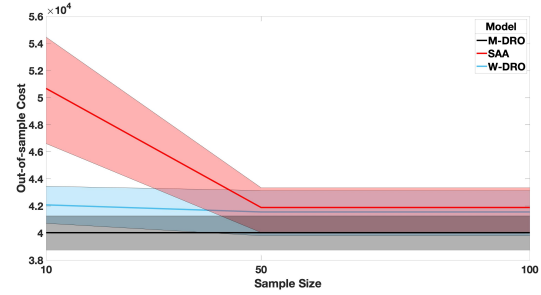


(b) 2nd

Figure 5: Out-of-sample performance under Unfirom, $\Delta = 0$



(a) TC



(b) 2nd

Figure 6: Out-of-sample performance under Unfirom $\Delta = 0.25$

to the higher fixed cost associated with activating more MFs (10 and 9 versus 7). However, both the W-DRO and MAD-DRO yield significantly lower second-stage operational costs (transportation and unmet demand) on average and at all quantiles. Thus, they offer a better quality of service. As N increases (more information becomes available), the W-DRO and SP model's out-of-sample costs decrease, with the former yielding significantly lower total and second-stage costs. The decrease in the out-of-sample cost of the SP makes sense because the decisions are made with more information.

Figures 5–7 present the results from misspecified distributions for each choice of variation, Δ . It is quite apparent that the W-DRO model solutions consistently outperform the SP solutions under all levels of variation (Δ) and across the criteria of mean and quantiles of total and second stage costs, especially when N is small. This demonstrates that the W-DRO approach is effective in an environment where the distribution quickly changes or when there is small data on demand variability. The MAD-DRO solutions yield lower second-stage operational costs than the SP and W-DRO solutions under all levels of variation, $\Delta \in \{0, 0.25, 0.5\}$. When $\Delta = 0$ (i.e., small variation), the MAD-DRO model yields a higher total cost (due to higher fixed cost) and the lowest mean and lower quantiles of the second-stage cost. Under higher variations ($\Delta = 0.25, 0.5$), the MAD-DRO models yield significantly lower total cost than the SP and W-DRO models. The DRO model's superior performance reflects the value of modeling distributional ambiguity of the demand.

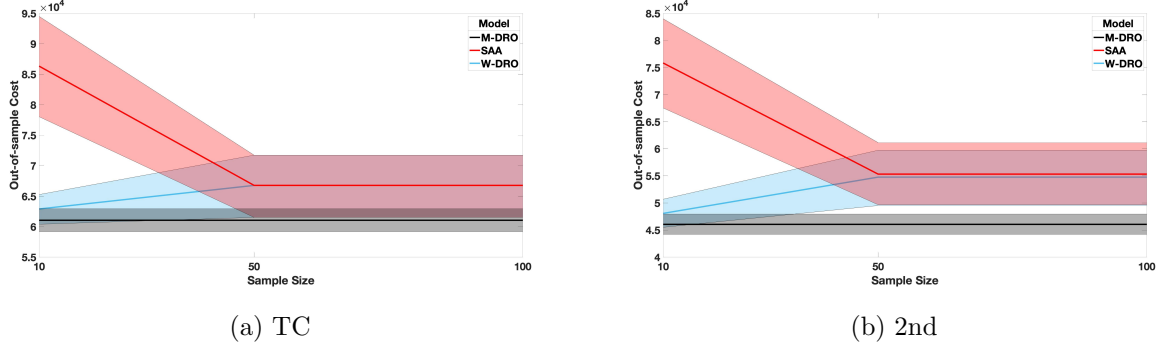


Figure 7: Out-of-sample performance under Unifrom, $\Delta = 0.5$

Table 6: Reliability of the models.

Perfect Distributional Information			
Model	$N = 10$	$N = 50$	$N = 100$
SP	0.48	0.48	0.59
W-DRO	1	1	1
MAD-DRO	1	1	1
Misspecified distributional information, Uniform with $\Delta = 0.5$			
Model	$N = 10$	$N = 50$	$N = 100$
SP	0	0	0
W-DRO	0.71	0.98	1
MAD-DRO	1	1	1

5.5. Reliability of the models

MFs operators often need to decide on the budget on their operational cost ahead of the actual MF service provision. The optimal value of a DMFRS model serves as an estimate of the cost of implementing the corresponding optimal solution (x, y) . When the actual cost is larger than the estimated cost, the MF operator may run into financial problems related to a budget deficit. Thus, risk-averse operators may prefer implementing solutions that have a higher reliability (i.e., estimated cost by model is higher than the actual cost). Table 6, present the reliability (i.e., the probability that the optimal value of a model exceed the out-of-sample performance of the corresponding optimal solutions) results for for instance 3 under perfect (logN) and misspecified distribution (uniform). We observe that the reliability of the SP increases with N . This is reasonable since we have more information from a larger data sample. However, the reliability of the SP model is the lowest among the three models and is zero under misspecified distribution. In contrast, the reliability of the DRO models is always 1 under perfect distribution. Under misspecified distribution, the reliability of W-DRO ranges from 0.71 (with $N = 10$) and 1 (with $N = 100$). The MAD-DRO consistently gives the best reliability result of 1. Overall, this demonstrates that DRO models have better reliability compared to SP.

5.6. Sensitivity Analysis

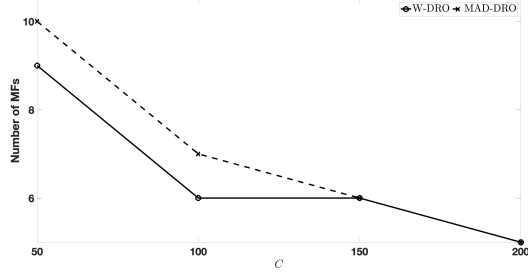
In this section, we study the sensitivity of DRO models to different parameter settings. Given that we observe similar results for all of the 11 DMFRS instances, for presentation brevity and illustrative purposes, we present results for instance 1 $(I, J, T)=(10, 10, 10)$ and instance 5 $(I, J, T)=(20, 20, 10)$ as examples of small and relatively large DMFRS instances.

First, we analyze the optimal number of active MFs as a function of the fixed cost, f , MF capacity, C , and range of demand. We fix all parameters as described in Section 5.1 and solve W-DRO and MAD-DRO with $C \in \{50, 100, 150, 200\}$ and $f \in \{1, 500 \text{ (low)}, 6, 000 \text{ (average)}, 10, 000 \text{ (high)}\}$ under the base range $\mathbf{W} \in [20, 60]$ and $\mathbf{W} \in [50, 100]$ (a higher volume of the demand). Figure 8 and 9 presents the optimal number of active MFs and the associated total cost for instance 1 and 5, respectively, $\mathbf{W} \in [20, 60]$. Since we observe similar results, we present the results under $\mathbf{W} \in [50, 100]$ in Figures G.12–G.13 in Appendix G.

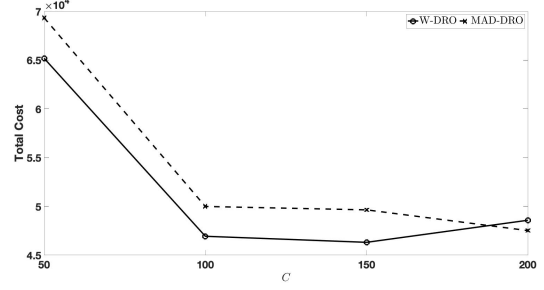
We observe the following from these Figures. First, the optimal number of scheduled MFs decreases as C increases irrespective of f . This makes sense because, with higher capacity, each MF can serve a larger amount of demand in each period. Second, both models schedule more MFs under $\mathbf{W} \in [50, 100]$, i.e., a higher volume of the demand. For example, consider instance 5, when $f = 1, 500$ and $C = 100$ the (W-DRO, MAD-DRO) models schedule (10, 13) and (18, 19) MFs under $\mathbf{W} \in [20, 60]$ and $\mathbf{W} \in [50, 100]$, respectively. Third, MAD-DRO, which is more conservative than W-DRO, always schedules a higher number of MFs, especially when C is tight. As such, MAD-DRO often has a slightly higher total cost (due to higher fixed) and better second-stage cost, i.e., better operational performance (see Figures G.14–G.16). For example, consider instance 1 under $f = 6, 000$ and $C = 50$, the W-DRO and MAD-DRO schedule 8 and 10 MFs, respectively. The associated (total, second-stage) costs of these solutions are respectively (74,495, 26,495) and (83,004, 23,004). Finally, both models schedule less MFs under a higher fixed cost f .

Second, we fix $M = 20$ and solve the models with unmet demand penalty $\gamma \in \{0.10\gamma_o, 0.25\gamma_o, 0.35\gamma_o, 0.50\gamma_o\}$ (where γ_o is the base case penalty in Section 5.1) and $f \in \{1, 500, 6, 000, 10, 000\}$. Figure 10 presents the number of MFs as a function of γ and f , and Figure 11 presents the associated second-stage cost. It is not surprising that as γ increases (i.e., satisfying customer demand becomes more important), the number of scheduled MFs increases, and accordingly, a larger amount of customer demand is satisfied and thus a lower second-stage cost (see Figure 11). However, we observe that for fixed γ , the MAD-DRO model schedules more MFs and thus yield a lower unmet demand cost (i.e., MAD-DRO solutions satisfy a higher amount of customers demand). For example, consider instance 1 and $f = 1, 500$, when γ decreases from $0.5\gamma_o$ to $0.1\gamma_o$ the optimal number of scheduled MFs of (W-DRO, MAD-DRO) decreases from (6, 6) to (1, 3) and average unmet demand cost increases from (9, 0) to (16, 117, 10, 827).

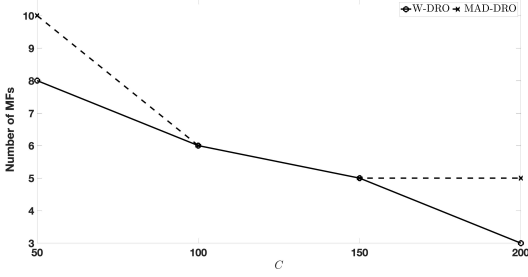
Our experiments in this section provide an example of how decision-makers can use our DRO approaches to generate DMFRS solutions under different parameter settings. Practitioners can use



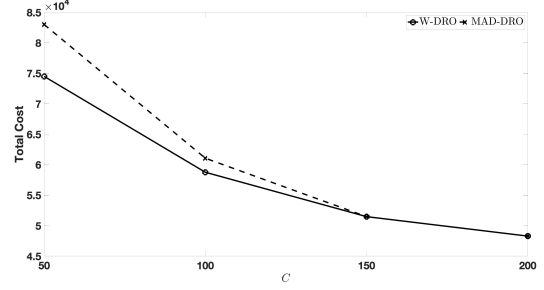
(a) Number of MFs, $f = 1,500$



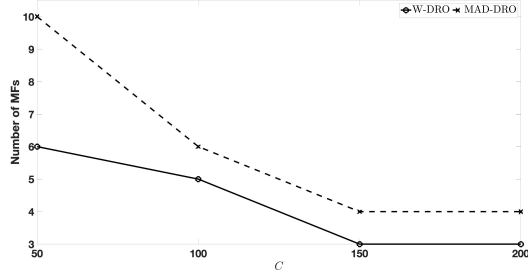
(b) Total cost, $f = 1,500$



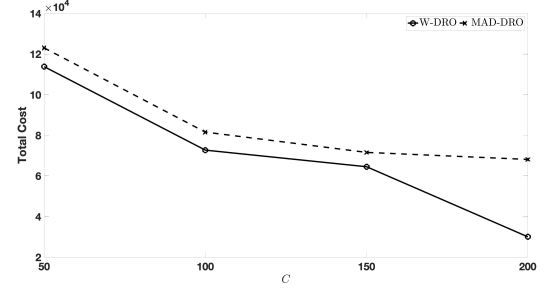
(c) Number of MFs, $f = 6,000$



(d) Total cost, $f = 6,000$



(e) Number of MFs, $f = 6,000$



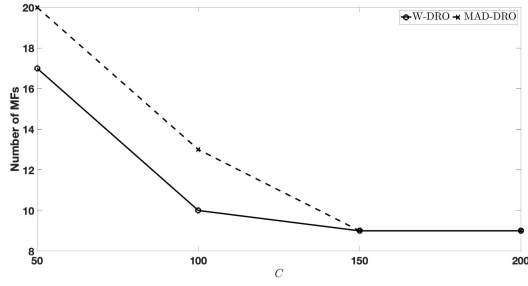
(f) Total cost, $f = 1,000$

Figure 8: Comparison of the results for different values of C and f under $\mathbf{W} \in [20, 60]$. Instance 1

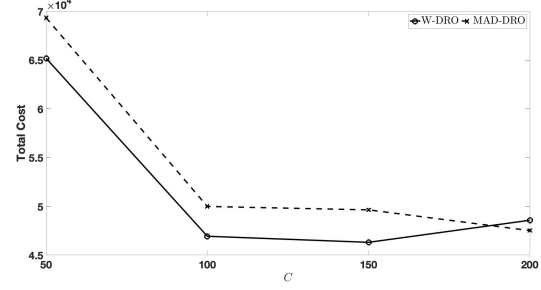
these results to decide whether to adopt the MAD-DRO model (which provides better operational and computational performance and cost) or the W-DRO model (which provides a lower fixed one-time cost of setting up the MF fleet).

6. Conclusion

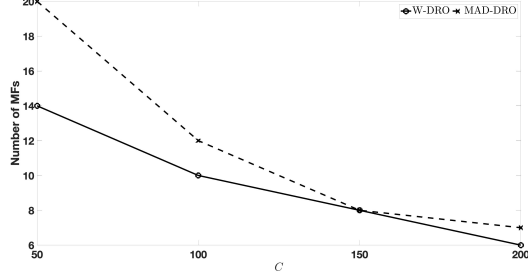
In this paper, we study a DMFRS problem, which seeks to find the number of MFs to serve customers in a given region during a planning horizon and their routing and scheduling decisions. The probability distribution of the demand is ambiguous. To address distributional ambiguity, we propose two DRO models. In the first (MAD-DMFRS), we use a MAD ambiguity set. In the second (W-DMFRS), we use an ambiguity set that incorporates all distributions within a 1-Wasserstein distance from a reference distribution. We derive equivalent mixed-integer non-linear programming



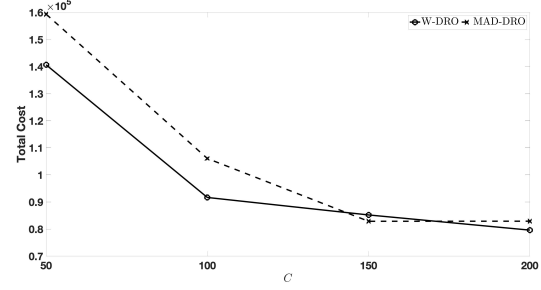
(a) Number of MFs, $f = 1,500$



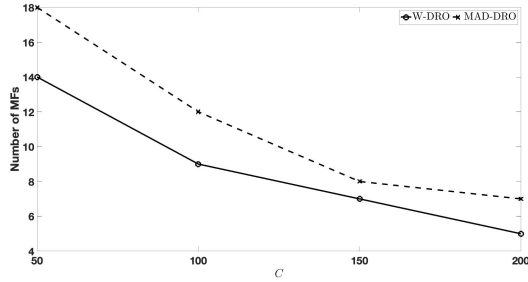
(b) Total cost, $f = 1,500$



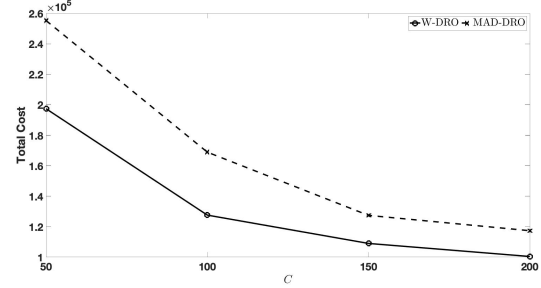
(c) Number of MFs, $f = 6,000$



(d) Total cost, $f = 6,000$



(e) Number of MFs, $f = 10,000$



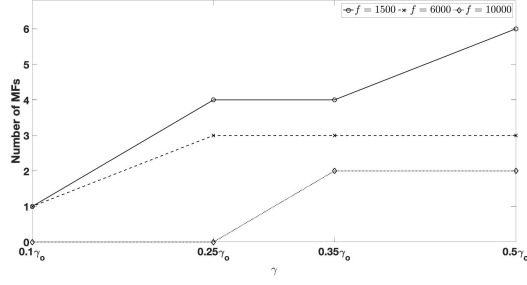
(f) Total cost, $f = 1,000$

Figure 9: Comparison of the results for different values of C and f under $\mathbf{W} \in [20, 60]$. Instance 5

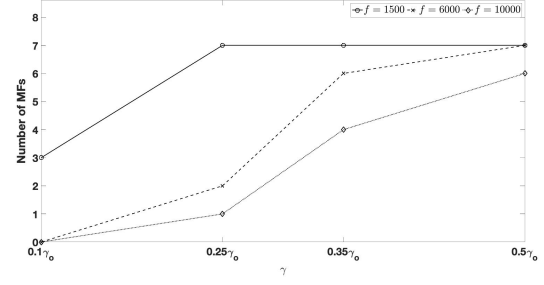
reformulations of these models. We linearize and propose a decomposition algorithm to solve the reformulations. We also derive lower bound and two-families of symmetry breaking inequalities to strengthen the master problem and speed up convergence.

Our computational results demonstrate (1) how the DRO approaches have superior operational performance in terms of satisfying customers demand as compared to the SP approach, (2) MAD-DMFRS is more computationally efficient than W-DMFRS, (3) MAD-DMFRS yield more conservative decisions than W-DMFRS, which often have a higher fixed cost but significantly lower operational (unmet demand and transportation) costs, (4) efficiency of the symmetry breaking and lower bound inequalities, (5) the trade-off between cost, number of MFs, MF capacity, and operational performance.

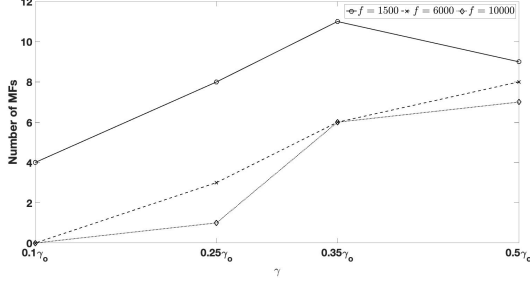
We suggest the following areas for future research. First, we aim to extend our models to opti-



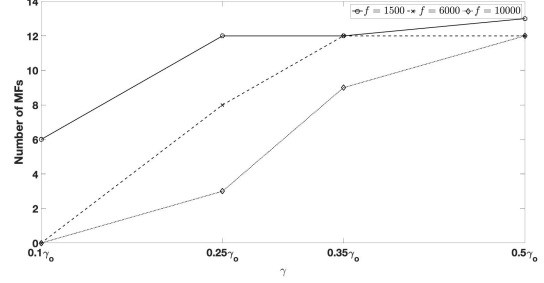
(a) W-DRO, Instance 1



(b) MAD-DRO, Instance 1



(c) W-DRO, Instance 5



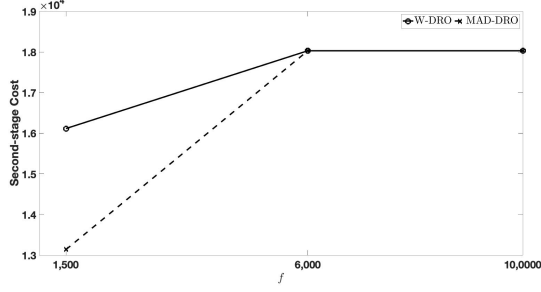
(d) MAD-DRO, Instance 5

Figure 10: Comparison of the results for different values of γ

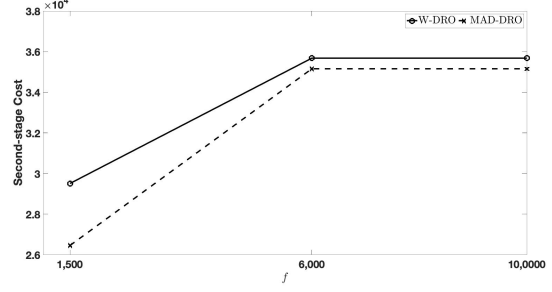
mize the capacity and size of the MF fleet. Second, want to extend our approach by incorporating multi-modal probability distributions. Third, we aim to incorporate the uncertainty of travel time in a data-driven DRO model. Fourth, we aim to incorporate other performance metrics such as access to MF service, equity, and fairness.

Acknowledgment

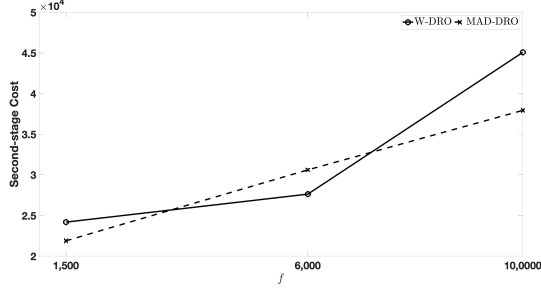
We want to thank all colleagues who have contributed significantly to the related literature. Dr. Karmel S. Shehadeh dedicates her effort in this paper to every little dreamer in the whole world who has a dream so big and so exciting. Believe in your dreams and do whatever it takes to achieve them—the best is yet to come for you



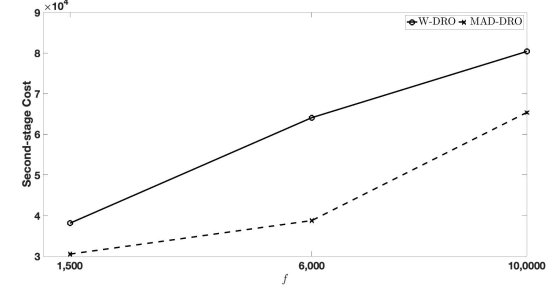
(a) $\gamma = 0.1\gamma_o$ Instance 1



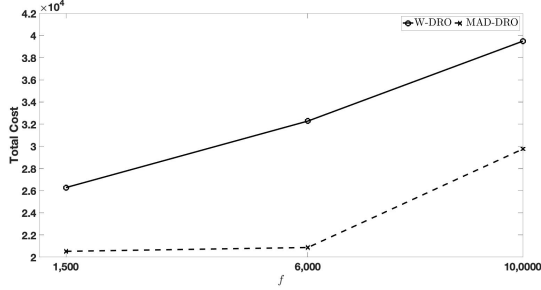
(b) $\gamma = 0.1\gamma_o$, Instance 5



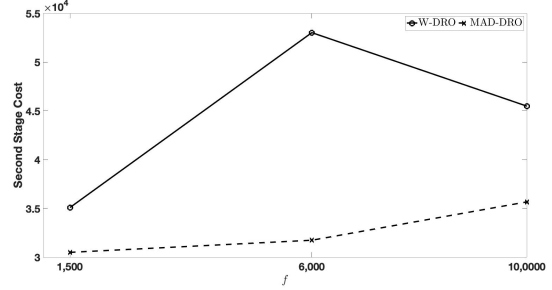
(c) $\gamma = 0.25\gamma_o$, Instance 1



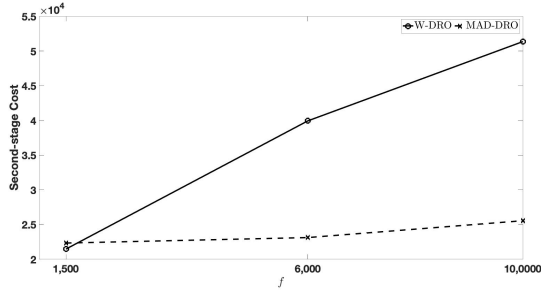
(d) $\gamma = 0.25\gamma_o$, Instance 5



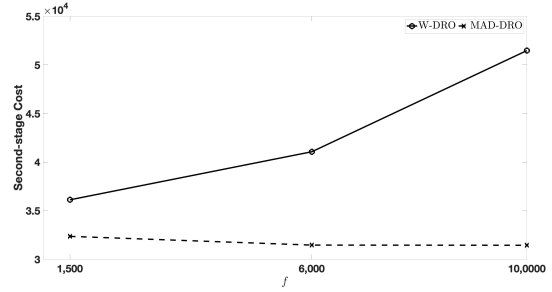
(e) $\gamma = 0.35\gamma_o$, Instance 1



(f) $\gamma = 0.35\gamma_o$, Instance 5



(g) $\gamma = 0.50\gamma_o$, Instance 1



(h) $\gamma = 0.50\gamma_o$, Instance 5

Figure 11: Comparison of the second-stage cost for different values of γ .

Appendix A. Derivation of feasible region \mathcal{X} in (2)

In this Appendix, we provide additional details on the derivation of the constraints defining the feasible region \mathcal{X} of variables (x, y) . As described in Lei et al. (2014), we can enforce the requirement

that an MF can only be in service when it is stationary using the following expressions:

$$x_{j,m}^t + x_{j',m}^t \leq 1 \quad \forall t, m, j, j \neq j', t' \in \{t, \dots, \min\{t + t_{j,j'}, T\}\} \quad (\text{A.1})$$

If $x_{j,m}^t = 1$ (i.e., MF m is stationary at some location j in period t , it can only be available at location $j' \neq j$ after a certain period, depending on the time it takes to travel from location j to location j' . It follows by (A.1) that $x_{j',m}^{t'} = 0$ for all $j \neq j', t' \in \{t, \dots, \min\{t + t_{j,j'}, T\}\}$. As pointed out by Lei et al. (2014), this indicated that *an earlier decision of deploying an MF at one candidate location would directly affect future decisions both temporally and spatially. In fact, this correlation is a major source of complexity for optimizing MFRSP.*

To enforce the condition that the MF has to be in an active condition before providing service it is necessary to include the following constraints:

$$x_{j,m}^t \leq y_m \quad \forall j, m, t \quad (\text{A.2})$$

It is straightforward to verify that constraint sets (A.1) and (A.2) can be combined into the following compact form

$$\mathcal{X} = \left\{ x, y : \begin{array}{l} x_{j,m}^t + x_{j',m}^{t'} \leq y_m, \quad \forall t, m, j, j \neq j', t' \in \{t, \dots, \min\{t + t_{j,j'}, T\}\} \\ x_{j,m}^t \in \{0, 1\}, y_m \in \{0, 1\}, \quad \forall j, m, t \end{array} \right\} \quad (\text{A.3})$$

Appendix B. Proof of Proposition 1

Proof. For a fixed $x \in \mathcal{X}$, we can explicitly write the inner problem $\sup[\cdot]$ in (6) as the following functional linear optimization problem.

$$\max_{\mathbb{P} \geq 0} \int_S Q(x, \xi) d\mathbb{P} \quad (\text{B.1a})$$

$$\text{s.t.} \quad \int_S W_{i,t} d\mathbb{P} = \mu_{i,t} \quad \forall i \in I, t \in T \quad (\text{B.1b})$$

$$\int_S |W_{i,t} - \mu_{i,t}| d\mathbb{P} \leq \eta_{i,t} \quad \forall i \in I, t \in T \quad (\text{B.1c})$$

$$\int_S d\mathbb{P} = 1 \quad (\text{B.1d})$$

Letting $\rho_{i,t}, \psi_{i,t}$ and θ be the dual variables associated with constraints (B.1b), (B.1c), (B.1d), respectively, we present problem (B.1) (problem (9) in the main manuscript) in its dual form:

$$\min_{\rho, \theta, \psi \geq 0} \sum_{t \in T} \sum_{i \in I} (\mu_{i,t} \rho_{i,t} + \eta_{i,t} \psi_{i,t}) + \theta \quad (\text{B.2a})$$

$$\text{s.t.} \quad \sum_{t \in T} \sum_{i \in I} (W_{i,t} \rho_{i,t} + |W_{i,t} - \mu_{i,t}| \psi_{i,t}) + \theta \geq Q(x, \xi) \quad \forall \mathbf{W} \in \mathcal{S} \quad (\text{B.2b})$$

where ρ and θ are unrestricted in sign, $\psi \geq 0$, and constraint (B.2b) is associated with the primal variable \mathbb{P} . Under the standard assumptions that $\mu_{i,t}$ lies in the interior of the set $\{\int_S W_{i,t} d\mathbb{Q} : \mathbb{Q} \in \mathcal{P}\}$,

is a probability distribution over S , strong duality hold between (B.1) and (B.2) (Bertsimas and Popescu (2005); Jiang et al. (2017); Shehadeh and Sanci (2021)). Note that for fixed (ρ, θ, ψ) , constraint (B.2b) is equivalent to

$$\theta \geq \max_{W \in \mathcal{S}} \left\{ Q(x, \xi) + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + |W_{i,t} - \mu_{i,t}| \psi_{i,t}) \right\}$$

Since we are minimizing θ in (B.2), the dual formulation of (B.1) is equivalent to:

$$\min_{\rho, \psi \geq 0} \left\{ \sum_{t \in T} \sum_{i \in I} (\mu_{i,t} \rho_{i,t} + \eta_{i,t} \psi_{i,t}) + \max_{W \in \mathcal{S}} \left\{ Q(x, \xi) + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + |W_{i,t} - \mu_{i,t}| \psi_{i,t}) \right\} \right\}$$

Appendix C. Proof of Proposition 2

First, note that the feasible regions $\Omega = \{(8b) - (8d), (10a) - (10b), (11c)\}$ and \mathcal{S} are independent of x and bounded. In addition, DMFRS has a complete recourse (i.e., feasible for any feasible $(x, y) \in \mathcal{X}$). Therefore, $\max_{\lambda, v, W, \pi, k} \left[\sum_{t \in T} \sum_{i \in I} \pi_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) \right] < \infty$. Second, for any fixed π, v, W, k , $\left[\sum_{t \in T} \sum_{i \in I} \pi_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) \right]$ is a linear function of x . It follow that $\max_{\lambda, v, W, \pi, k} \left[\sum_{t \in T} \sum_{i \in I} \pi_{i,t} + \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} C x_{j,m}^t v_{j,m}^t + \sum_{t \in T} \sum_{i \in I} -(W_{i,t} \rho_{i,t} + k_{i,t} \psi_{i,t}) \right]$ is the maximum of a linear functions of x , and hence convex and piecewise linear. Finally, it is easy to see that each linear piece of this function is associated with one distinct extreme point of Ω and \mathcal{S} . Given that each of these polyhedra has a finite number of extreme points, the number of pieces of this function is finite. This completes the proof.

Appendix D. Proof of Proposition 3

Recall that $\hat{\mathbb{P}}_{\xi}^N = \frac{1}{N} \sum_{n=1}^N \delta_{\xi^n}$. The definition of Wasserstein distance indicates that there exist a joint distribution Π of $(\xi, \hat{\xi})$ such that $\mathbb{E}_{\Pi}[\|\xi - \hat{\xi}\|] \leq \epsilon^p$. In other words, for any $\mathbb{P}_{\xi} \in \mathcal{P}(\mathcal{S})$, we can rewrite any joint distribution $\Pi \in \mathcal{P}(\mathbb{P}_{\xi}, \hat{\mathbb{P}}_{\xi}^N)$ by the conditional distribution of ξ given $\hat{\xi} = \hat{\xi}^n$ for $n = 1, \dots, N$, denoted as \mathbb{F}_{ξ}^n . That is, $\Pi = \frac{1}{N} \sum_{n=1}^N \mathbb{F}_{\xi}^n$. Notice that if we find one joint distribution $\Pi \in \mathcal{P}(\mathbb{P}_{\xi}, \hat{\mathbb{P}}_{\xi}^N)$ such that $\int \|\xi - \hat{\xi}\|_p^p d\Pi \leq \epsilon$, then $d(\mathbb{P}_{\xi}, \hat{\mathbb{P}}_{\xi}^N) \leq \epsilon$. Hence, we can drop the infimum operator in Wasserstein distance and arrive at the following equivalent problem

$$\sup_{\mathbb{F}_{\xi}^n \in \mathcal{P}(\mathcal{S}), n \in [N]} \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} Q(x, \xi) d\mathbb{F}_{\xi}^n \quad (\text{D.1a})$$

$$\text{s.t.} \quad \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} \|\xi - \hat{\xi}^n\|_p^p d\mathbb{F}_{\xi}^n \leq \epsilon^p \quad (\text{D.1b})$$

Using a standard strong duality argument and letting $\rho \geq 0$ be the dual multiplier, we can reformulate problem (D.1) by its dual, i.e.

$$\begin{aligned}
& \inf_{\rho \geq 0} \sup_{\mathbb{F}_\xi^n \in \mathcal{P}(\mathcal{S}), n \in [N]} \left\{ \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} Q(x, \xi) d\mathbb{F}_\xi^n + \rho \left[\epsilon^p - \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} \|\xi - \hat{\xi}^n\|_p^p d\mathbb{F}_\xi^n \right] \right\} \\
&= \inf_{\rho \geq 0} \left\{ \epsilon^p \rho + \frac{1}{N} \sum_{n=1}^N \sup_{\mathbb{F}_\xi^n \in \mathcal{P}(\mathcal{S})} \int_{\mathcal{S}} \left[Q(x, \xi) - \rho \|\xi - \hat{\xi}^n\|_p^p \right] d\mathbb{F}_\xi^n \right\} \\
&= \inf_{\rho \geq 0} \left\{ \epsilon^p \rho + \frac{1}{N} \sum_{n=1}^N \sup_{\xi \in \mathcal{S}} \{ Q(x, \xi) - \rho \|\xi - \hat{\xi}^n\|_p^p \} \right\}. \tag{D.2}
\end{aligned}$$

The results follow with $p = 1$.

Appendix E. Proof of Proposition 4

Recall from the definition of the ambiguity set that the lowest demand of each customer i in period t equals to the integer parameter $\underline{W}_{i,t}$. Now, if we treat the MFs as uncapacitated facilities, then we can fully satisfy $\underline{W}_{i,t}$ at the lowest assignment cost from the nearest location $j \in J'$, where $J' := \{j : x_{j,m}^t = 1\}$. Note that $J' \subseteq J$. Thus, the lowest assignment cost, δ_t , must be at least equal to or larger than $\sum_{i \in I} \min_{j \in J'} \{\beta d_{i,j}\} \underline{W}_{i,t}$. If $\gamma < \min_{j \in J} \{\beta d_{i,j}\}$ then δ_t must be at least equal to or larger than $\sum_{i \in I} \gamma \underline{W}_{i,t}$. Accordingly, $\delta_t \geq \sum_{i \in I} \min_{j \in J} \{\gamma, \min_{j \in J} \{\beta d_{i,j}\}\} \underline{W}_{i,t}$ is a valid lower bound δ_t , $\forall t \in T$.

Appendix F. Sample Average Approximation

$$\min \left[\sum_{m \in M} f y_m - \sum_{t \in T} \sum_{j \in J} \sum_{m \in M} \alpha x_{j,m}^t + \frac{1}{N} \sum_{n=1}^N \left(\sum_{j \in J} \sum_{i \in I} \sum_{m \in M} \sum_{t \in T} \beta d_{i,j} z_{i,j,m}^{t,n} + \gamma \sum_{t \in T} \sum_{i \in I} u_{i,t}^n \right) \right] \tag{F.1a}$$

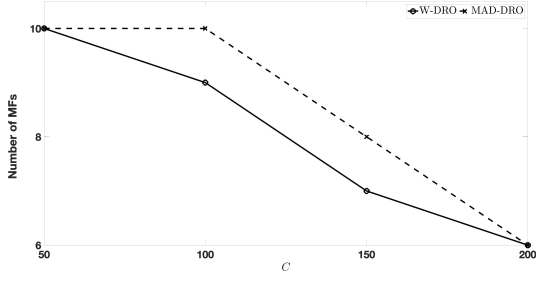
$$\text{s.t. } (x, y) \in \mathcal{X} \tag{F.1b}$$

$$\sum_{j \in J} \sum_{m \in M} z_{i,j,m}^{t,n} + u_{i,t}^n = W_{i,t}^n, \quad \forall i \in I, t \in T, n \in [N] \tag{F.1c}$$

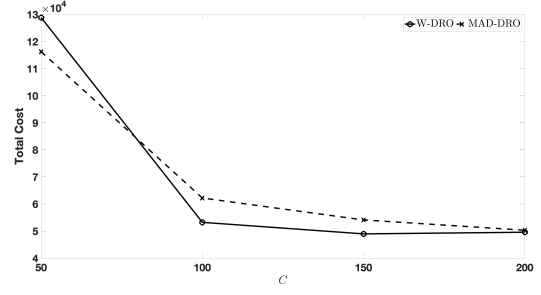
$$\sum_{i \in I} z_{i,j,m}^{t,n} \leq C x_{j,m}^t \quad \forall j \in J, m \in M, t \in T, n \in [N] \tag{F.1d}$$

$$u_{i,t}^n \geq 0, z_{i,j,m}^{t,n} \geq 0, \quad \forall i \in I, j \in J, m \in M, t \in [T], n \in [N] \tag{F.1e}$$

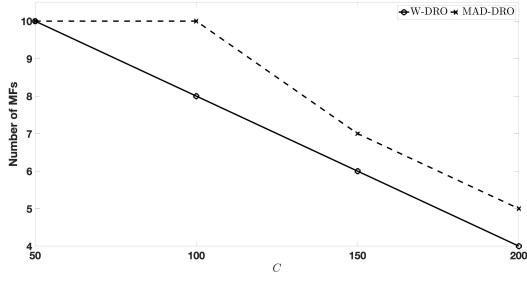
Appendix G. Additional Sensitivity Results



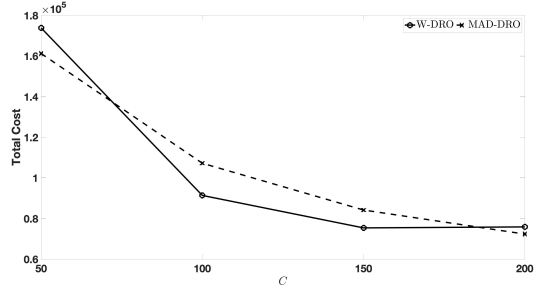
(a) Number of MFs, $f = 1,500$



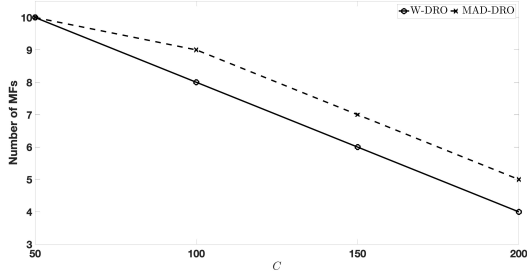
(b) Total cost, $f = 1,500$



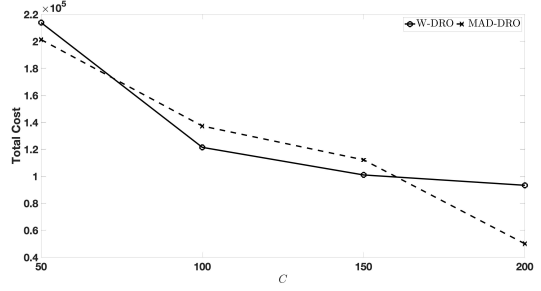
(c) Number of MFs, $f = 6,000$



(d) Total cost, $f = 6,000$

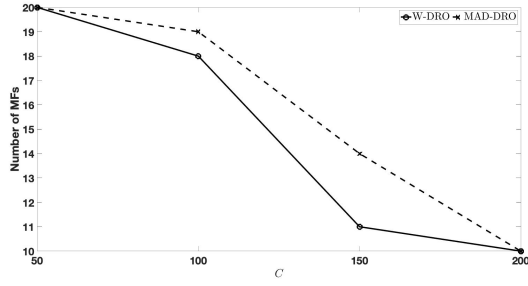


(e) Number of MFs, $f = 1,000$

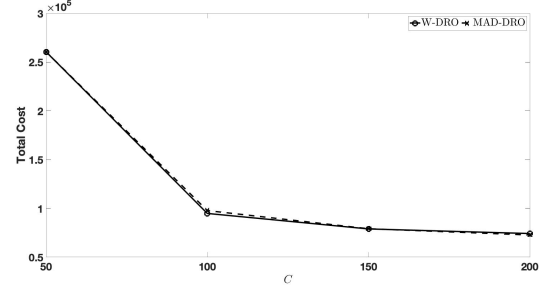


(f) Total cost, $f = 1,000$

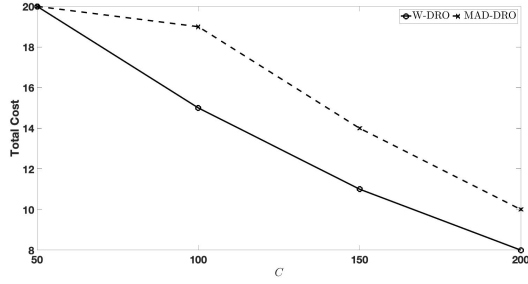
Figure G.12: Comparison of the results for different values of C and f under $\mathbf{W} \in [50, 100]$. Instance 1



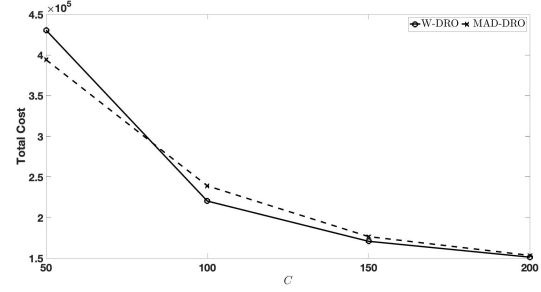
(a) Number of MFs, $f = 1,500$



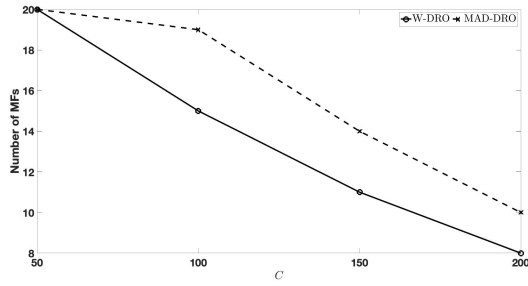
(b) Total cost, $f = 1,500$



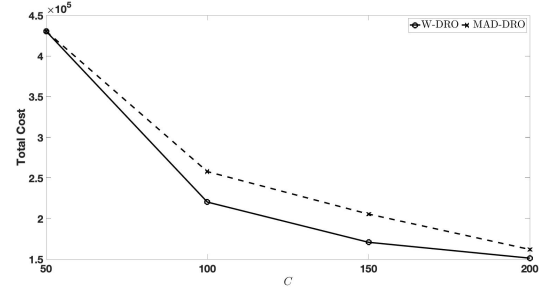
(c) Number of MFs, $f = 6,000$



(d) Total cost, $f = 6,000$



(e) Number of MFs, $f = 10,000$

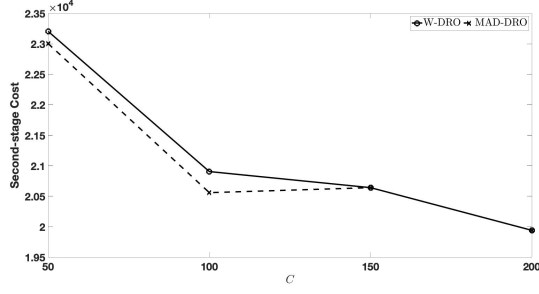


(f) Total cost, $f = 10,000$

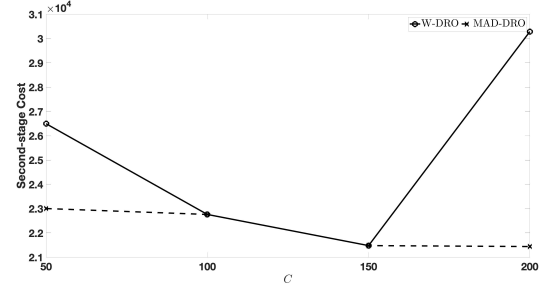
Figure G.13: Comparison of the results for different values of C and f under $\mathbf{W} \in [50, 100]$. Instance 5

References

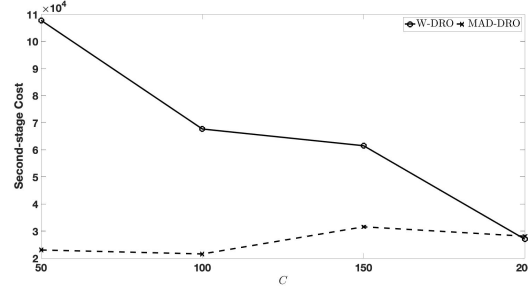
- Ahmadi-Javid, A., Seyedi, P., Syam, S. S., 2017. A survey of healthcare facility location. *Computers & Operations Research* 79, 223–263.
- Albareda-Sambola, M., Fernández, E., Hinojosa, Y., Puerto, J., 2009. The multi-period incremental service facility location problem. *Computers & Operations Research* 36 (5), 1356–1375.
- Antunes, A., Berman, O., Bigotte, J., Krass, D., 2009. A location model for urban hierarchy planning with population dynamics. *Environment and Planning A* 41 (4), 996–1016.
- Basciftci, B., Ahmed, S., Shen, S., 2020. Distributionally robust facility location problem under decision-dependent stochastic demand. *European Journal of Operational Research*.



(a) $f = 1,500$

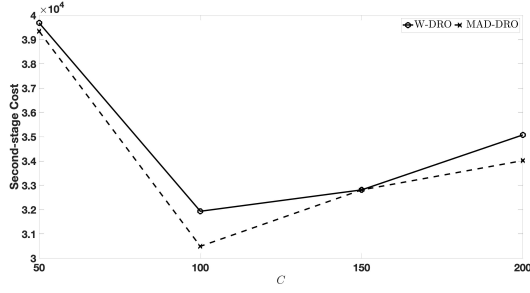


(b) $f = 6,000$

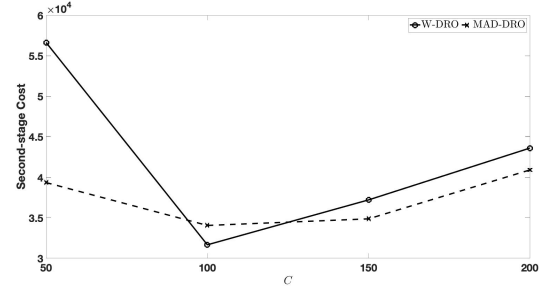


(c) $f = 10,000$

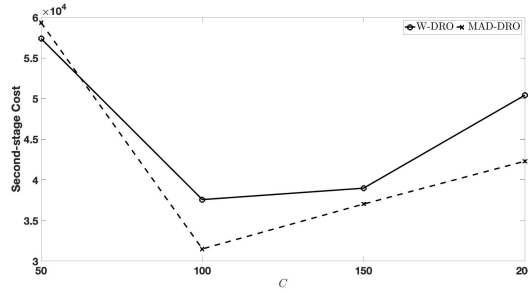
Figure G.14: Comparison of second-stage cost for different values of C and f under $\mathbf{W} \in [20, 60]$. Instance 1



(a) $f = 1,500$



(b) $f = 6,000$



(c) $f = 10,000$

Figure G.15: Comparison of second-stage cost for different values of C and f under $\mathbf{W} \in [20, 60]$. Instance 5

Ben-Tal, A., Den Hertog, D., Vial, J.-P., 2015. Deriving robust counterparts of nonlinear uncertain inequalities. Mathematical programming 149 (1-2), 265–299.

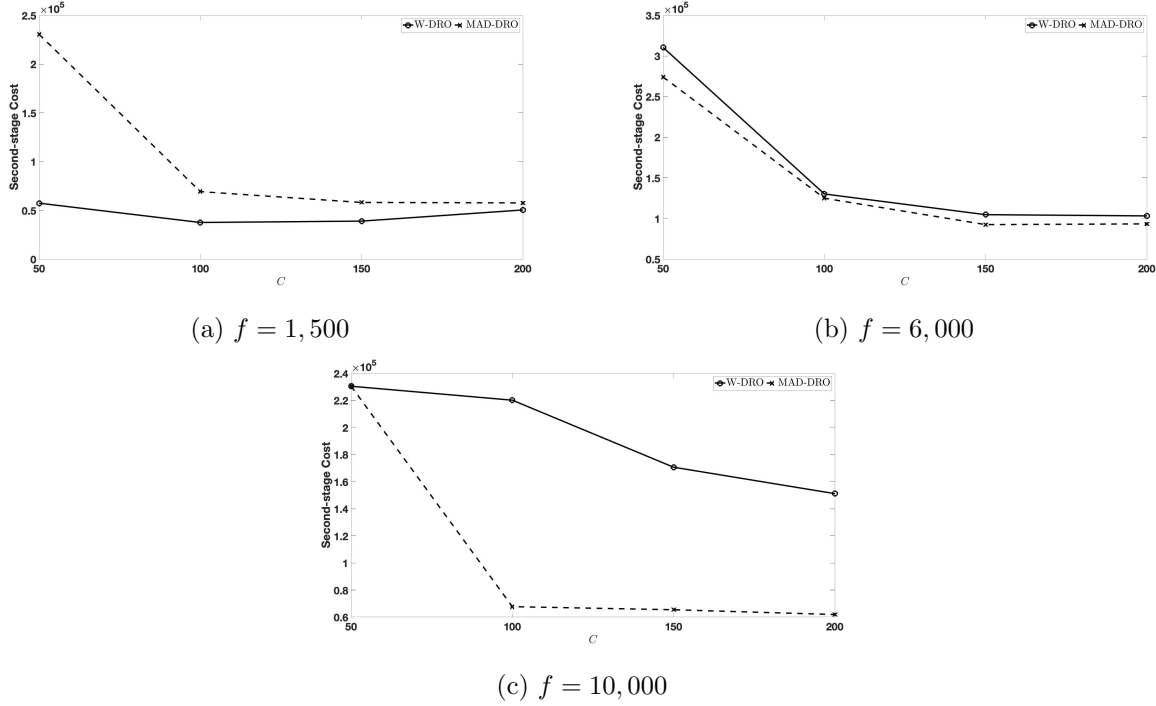


Figure G.16: Comparison of second-stage cost for different values of C and f under $\mathbf{W} \in [50, 100]$. Instance 5

Berman, O., Drezner, Z., Wesolowsky, G. O., 2003. Locating service facilities whose reliability is distance dependent. *Computers & Operations Research* 30 (11), 1683–1695.

Bertsimas, D., Popescu, I., 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* 15 (3), 780–804.

Bertsimas, D., Sim, M., 2004. The price of robustness. *Operations research* 52 (1), 35–53.

Blackwell, T., Bosse, M., 2007. Use of an innovative design mobile hospital in the medical response to hurricane katrina. *Annals of emergency medicine* 49 (5), 580–588.

Brown-Connolly, N. E., Concha, J. B., English, J., 2014. Mobile health is worth it! economic benefit and impact on health of a population-based mobile screening program in new mexico. *Telemedicine and e-Health* 20 (1), 18–23.

Chen, Z., Sim, M., Xiong, P., 2020. Robust stochastic optimization made easy with rsome. *Management Science*.

Contreras, I., Cordeau, J.-F., Laporte, G., 2011. The dynamic uncapacitated hub location problem. *Transportation Science* 45 (1), 18–32.

Current, J. R., Velle, C. R., Cohon, J. L., 1985. The maximum covering/shortest path problem:

- A multiobjective network design and routing formulation. *European Journal of Operational Research* 21 (2), 189–199.
- Delage, E., Saif, A., 2018. The Value of Randomized Solutions in Mixed-Integer Distributionally Robust Optimization Problems. GERAD HEC Montréal.
- Denton, B. T., Miller, A. J., Balasubramanian, H. J., Huschka, T. R., 2010. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations research* 58 (4-part-1), 802–816.
- Drezner, T., 2014. A review of competitive facility location in the plane. *Logistics Research* 7 (1), 114.
- Drezner, Z., Wesolowsky, G., 1991. Facility location when demand is time dependent. *Naval Research Logistics (NRL)* 38 (5), 763–777.
- Du Mortier, S., Coninx, R., 2007. Mobile health units in emergency operations: a methodological approach. Humanitarian Practice Network, Overseas Development Inst.
- Duque, Daniel, M. S., Morton, P. D., 2020. Distributionally robust two-stage stochastic programming. *Optimization Online* E-print.
- Esfahani, P. M., Kuhn, D., 2018. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171 (1-2), 115–166.
- Flores-Garza, D. A., Salazar-Aguilar, M. A., Ngueveu, S. U., Laporte, G., 2017. The multi-vehicle cumulative covering tour problem. *Annals of Operations Research* 258 (2), 761–780.
- Gendreau, M., Laporte, G., Semet, F., 1997. The covering tour problem. *Operations Research* 45 (4), 568–576.
- Gibson, J., Deng, X., Boe-Gibson, G., Rozelle, S., Huang, J., 2011. Which households are most distant from health centers in rural china? evidence from a gis network analysis. *GeoJournal* 76 (3), 245–255.
- Hachicha, M., Hodgson, M. J., Laporte, G., Semet, F., 2000. Heuristics for the multi-vehicle covering tour problem. *Computers & Operations Research* 27 (1), 29–42.
- Halper, R., Raghavan, S., 2011. The mobile facility routing problem. *Transportation Science* 45 (3), 413–434.
- Hanasusanto, G. A., Kuhn, D., 2018. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research* 66 (3), 849–869.

- Jena, S. D., Cordeau, J.-F., Gendron, B., 2015. Dynamic facility location with generalized modular capacities. *Transportation Science* 49 (3), 484–499.
- Jena, S. D., Cordeau, J.-F., Gendron, B., 2017. Lagrangian heuristics for large-scale dynamic facility location with generalized modular capacities. *INFORMS Journal on Computing* 29 (3), 388–404.
- Jiang, R., Ryu, M., Xu, G., 2019. Data-driven distributionally robust appointment scheduling over wasserstein balls. *arXiv preprint arXiv:1907.03219*.
- Jiang, R., Shen, S., Zhang, Y., 2017. Integer programming approaches for appointment scheduling with random no-shows and service durations. *Operations Research* 65 (6), 1638–1656.
- Lei, C., Lin, W.-H., Miao, L., 2014. A multicut l-shaped based algorithm to solve a stochastic programming model for the mobile facility routing and scheduling problem. *European Journal of operational research* 238 (3), 699–710.
- Lei, C., Lin, W.-H., Miao, L., 2016. A two-stage robust optimization approach for the mobile facility fleet sizing and routing problem under uncertainty. *Computers & Operations Research* 67, 75–89.
- Liu, K., Li, Q., Zhang, Z.-H., 2019. Distributionally robust optimization of an emergency medical service station location and sizing problem with joint chance constraints. *Transportation research part B: methodological* 119, 79–101.
- Luo, F., Mehrotra, S., 2018. Distributionally robust optimization with decision dependent ambiguity sets. *arXiv preprint arXiv:1806.09215*.
- Mevissen, M., Ragnoli, E., Yu, J. Y., 2013. Data-driven distributionally robust polynomial optimization. In: *Advances in Neural Information Processing Systems*. pp. 37–45.
- Oriol, N. E., Cote, P. J., Vavasis, A. P., Bennet, J., DeLorenzo, D., Blanc, P., Kohane, I., 2009. Calculating the return on investment of mobile healthcare. *BMC medicine* 7 (1), 1–6.
- Ostrowski, J., Linderoth, J., Rossi, F., Smriglio, S., 2011. Orbital branching. *Mathematical Programming* 126 (1), 147–178.
- Rahimian, H., Mehrotra, S., 2019. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*.
- Reilly, W. J., 1931. The law of retail gravitation. WJ Reilly.
- Saif, A., Delage, E., 2020. Data-driven distributionally robust capacitated facility location problem. *European Journal of Operational Research*.

- Shehadeh, K. S., Cohn, A. E., Epelman, M. A., 2019. Analysis of models for the stochastic outpatient procedure scheduling problem. *European Journal of Operational Research* 279 (3), 721–731.
- Shehadeh, K. S., Sanci, E., 2021. Distributionally robust facility location with bimodal random demand. *Computer and Operations Research*.
- Shehadeh, K. S., Tucker, E. L., 2020. A distributionally robust optimization approach for location and inventory prepositioning of disaster relief supplies. *arXiv preprint arXiv:2012.05387*.
- Song, Z., Hill, C., Bennet, J., Vavasis, A., Oriol, N. E., 2013. Mobile clinic in massachusetts associated with cost savings from lowering blood pressure and emergency department use. *Health affairs* 32 (1), 36–44.
- Soyster, A. L., 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research* 21 (5), 1154–1157.
- Subramanyam, A., Repoussis, P. P., Gounaris, C. E., 2020. Robust optimization of a broad class of heterogeneous vehicle routing problems under demand uncertainty. *INFORMS Journal on Computing*.
- Tricoire, F., Graf, A., Gutjahr, W. J., 2012. The bi-objective stochastic covering tour problem. *Computers & operations research* 39 (7), 1582–1592.
- Van Roy, T. J., Erlenkotter, D., 1982. A dual-based procedure for dynamic facility location. *Management Science* 28 (10), 1091–1105.
- Wang, S., Chen, Z., Liu, T., 2020. Distributionally robust hub location. *Transportation Science* 54 (5), 1189–1210.
- Wang, Y., Zhang, Y., Tang, J., 2019. A distributionally robust optimization approach for surgery block allocation. *European Journal of Operational Research* 273 (2), 740–753.
- Wu, C., Du, D., Xu, D., 2015. An approximation algorithm for the two-stage distributionally robust facility location problem. In: *Advances in Global Optimization*. Springer, pp. 99–107.
- Zhen, J., Kuhn, D., Wiesemann, W., 2021. Mathematical foundations of robust and distributionally robust optimization. *arXiv preprint arXiv:2105.00760*.