Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Reliable Facility Location Design Under Uncertain Correlated Disruptions

Mengshi Lu, Lun Ran, Zuo-Jun Max Shen

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 17, No. 4, Fall 2015, pp. 445–455
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)
http://dx.doi.org/10.1287/msom.2015.0541

Reliable Facility Location Design Under Uncertain Correlated Disruptions

Mengshi Lu
Krannert School of Management, Purdue University, West Lafayette, Indiana 47907, mengshilu@purdue.edu

Lun Ran
School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China, ranlun@bit.edu.cn

Zuo-Jun Max Shen
Department of Industrial Engineering and Operations Research, and Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, California 94720, maxshen@berkeley.edu

Most previous studies on reliable facility location design assume that disruptions at different locations are independent. In this paper, we present a model that allows disruptions to be correlated with an uncertain joint distribution, and we apply distributionally robust optimization to minimize the expected cost under the worst-case distribution with given marginal disruption probabilities. The worst-case distribution has a practical interpretation with disruption propagation, and its sparse structure allows solving the problem efficiently. Our numerical results show that ignoring disruption correlation could lead to significant loss that increases dramatically in key factors such as source disaster probability, disruption propagation effect, and service interruption penalty. On the other hand, the robust model results in very low regret, even when disruptions are independent, and starts to outperform the model assuming independence when disruptions are mildly correlated. Most of the benefit of the robust model can be captured with a very low additional cost, which makes it easy to implement. Given these advantages, we believe that the robust model can serve as a promising alternative approach for solving reliable facility location problems.

Keywords: facility location; supply chain disruption; distributional uncertainty
History: Received: April 4, 2013; accepted: March 13, 2015. Published online in *Articles in Advance* May 19, 2015.

1. Introduction

Recently, severe supply chain disruptions have have resulted in significant losses because of facility damage and production or service interruption. Designing reliable supply chains when facilities are subject to random disruptions has gained unprecedented attention from industry and academia. For example, IBM has launched the Business Continuity and Resilience Services to help companies evaluate their disruption risk and improve their resilience using optimized planning and design.[1] The Ford Motor Company has implemented a quantitative model that evaluates disruption risks in procurement and supplier selection (Simchi-Levi et al. 2014). In operations research and management sciences, reliable facility location design has been studied extensively (e.g., Snyder and Daskin 2005, Cui et al. 2010, Lim et al. 2010).

In most of the existing reliable facility location literature, disruptions at different locations are assumed to be independent. However, in practice, positively correlated disruptions are widely observed. First, large-scale natural disasters such as earthquakes, tsunamis, and hurricanes usually cause damages in vast geographic regions. For example, in October 2012, Hurricane Sandy caused power outages to 7.9 million businesses and households in 15 different states (*CNN* 2014). Second, severe weather conditions, such as tornadoes and storms, tend to have widespread outbreaks within a short period of time. For example, in May 2013, 61 tornadoes occurred in eight different states within three days, and the total damage was estimated to amount to three billion dollars.[2] Under these circumstances, multiple facilities can be disrupted simultaneously either by one large-scale disaster or by multiple severe weather hazards that occur within a short period of time.

Disruption correlation can significantly affect the magnitude of the disruption risk faced by the supply chain. As we shall see later, it also affects opti-

[1] http://www.ibm.com/services/continuity (accessed May 9, 2015).

[2] National Oceanic and Atmospheric Administration. Billion-dollar weather and climate disasters. Accessed December 15, 2014, http://www.ncdc.noaa.gov/billions/events.

mal facility location design. However, because of the difficulty in estimation, modeling, and optimization, most existing literature in reliable facility location design only considered independent disruptions. In this paper, we present a distributionally robust optimization model to incorporate correlated disruptions. We assume that the disruptions have an unknown joint distribution and minimize the expected cost under the worst-case distribution with given marginal disruption probabilities. Using the structural property of a class of widely studied reliable facility location problems, we derive the worst-case distribution in a closed form, which has a practical interpretation. The sparse structure of the worst-case distribution also allows us to transform this seemingly complex problem into a much simpler equivalent problem and solve it efficiently.

We compare the optimal solutions of the robust model with those of the traditional model, which is based on the assumption of independent disruptions. We are particularly interested in the regret or loss from model misspecification, which is the cost increase when the optimal solution of one model is erroneously used in the other model. We find that ignoring disruption correlation can result in significant losses. On the other hand, applying the robust model under independent disruptions results in much lower cost increases. We study the impact of key factors such as source disaster probability, disruption propagation effect, and service interruption penalty on the regret of the two models. We find that as these factors increase, the regret of the traditional model increases dramatically, whereas the regret of the robust model only increases slightly or largely stays the same. We also compare the two models under different degrees of correlation and find that even though the robust model is based on the worst-case correlation, it still outperforms the traditional model when disruptions are only mildly correlated. By considering a weighted-average objective consisting of the worst-case expected cost and the normal operating cost with no disruption, we find that most of the benefit of the robust model can be captured with a very low additional cost.

Given these advantages, we believe this robust model can serve as a promising alternative for solving reliable facility location problems. It does not require any additional model input and thus can be applied directly to real-world problems that are currently being solved by the traditional approach that assumes independent disruptions. The robust model also requires much less computational effort. Thus, it can be used to solve large-scale problems efficiently.

The rest of the paper is organized as follows. In §2 we review related literature. In §3 we present the distributionally robust reliable facility location model and its equivalent formulation. In §4 we present an example for supply chain network design using real-world data and a numerical study using simulated data. In §5 we summarize the results and discuss directions for future work.

2. Literature Review

Snyder et al. (2014) identified two major streams of reliable facility location models: stochastic (S) models and robust (R) models. Stochastic models further fall into four main categories: scenario-based (SB) models, implicit formulation (IF) models, reliable backup (RB) models, and continuum approximation (CA) models. For robust models, most of the literature is based on the interdiction median (IM) model. Table 1 summarizes some of the literature in these categories. For a more comprehensive and detailed review, refer to Snyder et al. (2014).

Next, we discuss why most models in the literature are not applicable or suitable for correlated disruptions. The IF model is based on implicitly calculating the probability that a customer will be served by each facility, which requires the assumption of independent disruptions. The RB model assumes each customer is backed up by a fixed perfectly reliable facility under all disruption scenarios. Disruption correlation will not affect the probability that a customer must be rerouted to a more distant backup facility. In fact, if the backup facilities have infinite capacity (which to our knowledge is assumed by all literature using the RB model), disruption correlation will not affect the expected cost. The IM model is concerned with the worst-case disruption scenario among all possible scenarios, but it does not consider any probabilistic distribution. Thus, it cannot model disruption correlation. When designing supply chains under the threat of natural disasters and severe weather hazards, disruptions typically follow a probabilistic distribution with positive correlations. Also, when multiple suppliers or distribution centers are disrupted, supply usually must be shipped from more distant

Table 1 Summary of Literature on Reliable Facility Location

Category	Literature
Stochastic	
SB	Shen et al. (2011)
IF	Snyder and Daskin (2005), Berman et al. (2007), Cui et al. (2010), Chen et al. (2011), Shen et al. (2011), Li and Ouyang (2012), Aboolian et al. (2013), Li et al. (2013b)
RB	Lim et al. (2010), An et al. (2015), Li et al. (2013a)
CA	Cui et al. (2010), Li and Ouyang (2010), Berman et al. (2013), Lim et al. (2013)
Robust	
IM	Church and Scaparra (2007), Scaparra and Church (2008), Liberatore et al. (2012), An et al. (2014)

sources. Thus, the aforementioned models may not be applicable.

The SB model can incorporate correlated disruptions using sample average approximation (SAA). To our knowledge, the only paper that considered the SAA approach for reliable facility location is Shen et al. (2011), which assumes independent disruptions. Their results showed that the SAA approach performs rather poorly compared with a greedy heuristic algorithm. For correlated disruptions, one can expect the performance of SAA to be as poor as for independent disruptions because the joint distribution of correlated Bernoulli random variables is not well defined, and simulating from this distribution is harder than simulating from the independent distribution. Another drawback of the SAA approach is that it requires knowing the disruption correlation. In practice, the exact correlation may be unknown, and only the marginal disruption probability is known. For example, in §4, we present an example for supply chain network design in which the marginal disruption probability is estimated using severe weather hazard probability data from the National Oceanic and Atmospheric Administration (NOAA). The disruption correlation, on the other hand, is not available.

To our knowledge, CA is the only approach that has been successfully applied to incorporate correlated disruptions. Li and Ouyang (2010) considered the CA counterpart of the IF model given the conditional disruption probability. They found that the expected cost is higher when disruptions are positively correlated. Their numerical study shows that the impact of correlation on the expected cost can be significant when both the disruption probability and the service interruption penalty are high. Lim et al. (2013) considered the CA counterpart of the RB model with capacitated backup facilities. Their main purpose is to study the effect of misspecified disruption probabilities and/or correlations on the relative regret. They found that the expected cost increases in correlation and decreases in capacity. Their numerical result shows that joint underestimation of disruption probability and correlation results in higher loss compared to joint overestimation. Berman et al. (2013) considered the continuous 2-median and 2-center problems restricted to a unit line segment. They derived the trajectory of optimal locations as a function of disruption probability and correlation.

The major difference between our model and the CA-based models is that ours is a discrete location model, whereas the CA model is a continuous location model. Continuous location models require that the demand points can be properly approximated by a continuous function and that potential locations are not restricted to a set of given candidate sites. Whereas these conditions may hold under certain circumstances (for example, individual customers within an urban area can be approximated by a continuous function), they may not hold under many other circumstances. We consider a detailed supply chain design problem in which discrete demand points are distributed across a large area and potential locations for warehouses and distribution centers are restricted to a number of candidate sites. Thus, we believe a discrete model is more suitable for this setting.

Given the difference in the nature and the specific settings of the models, a simple comparison of the results and insights from this paper and those from the CA-based papers may not be completely appropriate. Nonetheless, we notice the following key differences. First, in contrast to Li and Ouyang (2010), who found that the regret of ignoring correlation is usually not significant, we find that such regret is significant in our supply chain network design example using real-world data and our numerical study using simulated data. Also, we find that such regret is much higher than the regret from using the robust design under independent disruptions. Also, Li and Ouyang (2010) found that the number of open facilities is smaller when disruptions are correlated, whereas we find the opposite. Second, Lim et al. (2013) found that the effect of misspecification in disruption correlation alone is very limited. We find that misspecification in correlation alone can also result in significant losses, and overestimating the correlation (i.e., assuming worst-case correlation) is, in general, better than underestimating (i.e., assuming independence).

More recently, several papers studied discrete location models with certain deterministic interdependence structures between locations. Liberatore et al. (2012) considered one type of interdependence known as the "ripple effect" in which disruptions at one location will cause nearby facilities to lose a fixed amount of capacity. They incorporated the ripple effect in the IM model with fortification decisions. Our model differs from that of Liberatore et al. (2012) in that the IM model is for determining the worst-case disruption scenario for a given location design, whereas our model is for determining the optimal design. Another difference is that we consider correlated random disruptions, whereas Liberatore et al. (2012) considered deterministic interdependence structures between locations. Li et al. (2013b) considered a different type of interdependence known as "supporting station" in which the facilities require resources provided by several supporting stations. Independent disruptions to the supporting stations will thus lead to correlated disruptions to the facilities. Our model does not require the special structure of supporting stations and thus can be applied under more general settings.

In summary, our model significantly differs from those explored in existing literature. In contrast to the CA-based models, our model is a discrete model that is applicable under more general problem settings. New insights are drawn from our numerical results. In contrast to the models of Liberatore et al. (2012) and Li et al. (2013b), our model is based on correlated random disruptions rather than deterministic interdependence structures.

3. Model and Formulation

In this section we examine the reliable uncapacitated fixed-charge location (RUFL) problem to illustrate the distributionally robust optimization model for reliable facility location problems. The same approach can be applied to other widely studied reliable facility location problems, including the p-median problem, the capacitated fixed-charge location problem, and the multiallocation hub location problem. Details of the generalization are available in the online appendix (available as supplemental material at http://dx.doi.org/10.1287/msom.2015.0541).

Consider the problem of locating facilities at a set $\mathcal{J} = \{1, \ldots, J\}$ of candidate locations to serve a set $\mathcal{I} = \{1, \ldots, I\}$ of customers. Let d_i denote the demand of customer $i \in \mathcal{I}$ and f_j the fixed cost of opening a facility at location $j \in \mathcal{J}$. Serving customer i from a facility at location j incurs unit transportation cost c_{ij}. Let $\mathbf{x} = (x_0, x_1, \ldots, x_J)$ denote the facility location decision, where $x_j = 1$ if facility is opened at location j, and $x_j = 0$ otherwise. The facilities are subject to random disruptions. Let $\boldsymbol{\xi} = (\xi_0, \xi_1, \ldots, \xi_J)$ denote the disruption scenario, where $\xi_j = 0$ if location j is disrupted, and $\xi_j = 1$ if it is online, i.e., not disrupted. (We will sometimes, for convenience, use the set of online locations, S, to denote the disruption scenario, with the correspondence $S(\boldsymbol{\xi}) = \{j \in \mathcal{J}\colon \xi_j = 1\}$ and $\boldsymbol{\xi}(S) = (I(0 \in S), I(1 \in S), \ldots, I(J \in S))$, where $I(\cdot)$ is the indicator function.) Given $\mathbf{x}$ and $\boldsymbol{\xi}$, each customer is either assigned to an available (i.e., open and online) facility, or its service is interrupted. Let y_{ij} denote the customer assignment decision, with $y_{ij} = 1$, if customer i is assigned to facility j, and $y_{ij} = 0$ otherwise. To model service interruptions, a virtual facility 0 is added to $\mathcal{J}$. If $y_{i0} = 1$, customer i's service is interrupted, with c_{i0} being the unit penalty cost. The virtual facility is never disrupted, i.e., $\xi_0 \equiv 1$, and its fixed cost $f_0 = 0$.

Let $h(\mathbf{x}, \boldsymbol{\xi})$ denote the transportation and penalty cost under the optimal customer assignment/interruption decisions, given location design $\mathbf{x}$ and disruption scenario $\boldsymbol{\xi}$, i.e.,

$$h(\mathbf{x}, \boldsymbol{\xi}) = \min\left\{ \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} d_i c_{ij} y_{ij} \;\middle|\; \begin{array}{ll} \sum_{j\in\mathcal{J}} y_{ij} = 1, & \forall\, i\in\mathcal{I}; \\ 0 \le y_{ij} \le x_j \xi_j, & \forall\, i\in\mathcal{I}, \forall\, j\in\mathcal{J} \end{array} \right\}. \quad (1)$$

Let $p(\boldsymbol{\xi})$ be the joint distribution of the disruptions, i.e., $p(\boldsymbol{\xi})$ is the probability that disruption scenario $\boldsymbol{\xi}$ occurs. The RUFL problem is defined as

$$\text{(RUFL)} \quad \min_{\mathbf{x}\in\mathcal{X}} \left\{ \sum_{j\in\mathcal{J}} f_j x_j + \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})] \right\},$$

where $\mathcal{X} = \{\mathbf{x}\colon x_j \in \{0, 1\}, \forall\, j \in \mathcal{J}\}$. Traditional RUFL models (e.g., Snyder and Daskin 2005, Cui et al. 2010) consider the special case in which disruptions are independent, i.e., $p(\boldsymbol{\xi}) = \prod_{j\in\mathcal{J}} (1 - q_j)^{\xi_j} (q_j)^{1-\xi_j}$, where q_j is the marginal disruption probability of location j.

In distributionally robust optimization, instead of assuming some specific joint distribution, we assume that $p(\boldsymbol{\xi})$ is unknown but lies within a distributional uncertainty set. Specifically, we consider the set of all joint distributions such that the marginal disruption probability of location j is equal to q_j, i.e.,

$$\mathcal{P} = \left\{ p \;\middle|\; \begin{array}{ll} \sum_{S\colon j\in S} p(S) = 1 - q_j, & \forall\, j\in\mathcal{J}; \\ p(S) \ge 0, & \forall\, S \subseteq \mathcal{J}; \\ p(S) = 0, & \forall\, S, 0 \notin S \end{array} \right\}.$$

Recall that the virtual facility is never disrupted, i.e., $q_0 = 0$, which guarantees that p is a probability distribution. The choice of this uncertainty set allows direct comparison with the traditional model, since it does not require any additional model input.

The distributionally robust reliable uncapacitated fixed-charge location (DR-RUFL) problem minimizes the expected cost under the worst-case distribution in $\mathcal{P}$:

$$\text{(DR-RUFL)} \quad \min_{\mathbf{x}\in\mathcal{X}} \left\{ \sum_{j\in\mathcal{J}} f_j x_j + \max_{p\in\mathcal{P}} \mathbb{E}_p[h(\mathbf{x}, \boldsymbol{\xi})] \right\}. \quad (2)$$

Distributionally robust optimization has been extensively studied and applied to various operations management problems. More specifically, our model falls into the category of marginal moment models (Bertsimas et al. 2004). Agrawal et al. (2012) also studied marginal moment models. Their focus is to derive an upper bound on the regret that arises from ignoring correlation for a class of problems. Most reliable facility location models are not in this class, which means ignoring correlation can result in substantial regret.

Although considering the worst-case distribution is conservative, we believe it can usually be justified. First, previous studies suggest that in supply chain risk management, managers are more concerned about the "maximum exposure," i.e., the worst case (Tang 2006). Second, as we will discuss later, the worst-case distribution for the DR-RUFL problem has a practical interpretation. For certain types of disruptions that propagate from a central source, e.g., earthquakes, the worst-case distribution resembles the actual distribution more closely than the independent distribution. Third, since the actual distribution is typically unknown, given only the marginal probability, one can either apply the traditional models that assume the disruptions are independent or apply the DR-RUFL model that considers the worst case. Our numerical results in §4 show that the latter option outperforms the former even when disruptions are only mildly correlated. Furthermore, the optimal solution under the worst-case distribution is not expensive to implement, and most of its benefit can be achieved with a very low additional cost.

3.1. Equivalent Formulation of DR-RUFL

The DR-RUFL problem in (2) is a mini-max formulation. The inner problem has the objective of choosing the worst disruption distribution p for a given location decision $\mathbf{x}$, which can be formulated as a linear program:

$$\max \mathbb{E}_p[h(\mathbf{x}, S)] = \max_{p\in\mathcal{P}} \sum_{S\subseteq\mathcal{J}} p(S)h(\mathbf{x}, S).$$

This linear program has 2^J variables, which could still make the DR-RUFL problem computationally intractable. However, as we will show later, because of the structural property of RUFL, the worst-case distribution has a closed-form solution that does not depend on $\mathbf{x}$ or $h(\mathbf{x}, S)$. The DR-RUFL problem can then be transformed into a much simpler equivalent problem and solved efficiently.

First, we need to show that with any given $\mathbf{x}$, the cost function $h(\mathbf{x}, S)$ in (1) is supermodular in S. A set function g is said to be supermodular if for any $S, T \subseteq \mathcal{J}$, $g(S\cap T) + g(S\cup T) \geq g(S) + g(T)$. It can be shown that g is supermodular if and only if

$$g(S\cup\{j\}) - g(S) \leq g(T\cup\{j\}) - g(T), \quad \forall S\subset T\subset\mathcal{J},\ \forall j\in\mathcal{J}\backslash T. \tag{3}$$

We have the following lemma.

Lemma 1 (Supermodularity). *For any $\mathbf{x}\in\mathcal{X}$, the cost function $h(\mathbf{x}, S)$ given in (1) is supermodular in S.*

The intuition is that having additional available facilities has diminishing marginal return, which corresponds to condition (3). The same result holds for several other reliable facility location problems, as we mentioned at the beginning of this section.

Using supermodularity, we can derive the worst-case distribution. Without loss of generality, assume the facilities are indexed in increasing order of marginal disruption probabilities, i.e., $0\equiv q_0 \leq q_1 \leq \cdots \leq q_J \leq q_{J+1} \equiv 1$. Consider $J+1$ disruption scenarios denoted by $\boldsymbol{\xi}^0, \boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^J$. In scenario $\boldsymbol{\xi}^s$, $\xi_j^s = I\ (j \leq s)$ for all $j\in\mathcal{J}$, where $I(\cdot)$ is the indicator function. In other words, locations $s+1, \ldots, J$ are disrupted, and locations $0, 1, \ldots, s$ are online. We then have the following lemma by Edmonds (1971) and Agrawal et al. (2010).

Lemma 2 (Worst-Case Disruption Distribution). *In the worst-case disruption distribution for DR-RUFL, only disruption scenarios $\boldsymbol{\xi}^0, \boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^J$ may have nonzero probabilities, and the probability of scenario $\boldsymbol{\xi}^s$ is equal to $q_{s+1} - q_s$ for all $s = 0, 1, \ldots, J$.*

To better understand Lemma 2, consider the case in which disruptions are caused by an earthquake (e.g., Liberatore et al. 2012). As the earthquake propagates from the epicenter, the strength of its impact will decrease. As a result, there exists a region known as the impact region in which the impact of the earthquake is strong enough to damage facilities. The marginal disruption probability of facility j is nothing but the probability that the impact region is large enough to include facility j. In scenario $\boldsymbol{\xi}^s$, facilities $s+1, \ldots, J$ are disrupted, whereas facilities $1, \ldots, s$ are online. This means that the impact region is large enough to include facility $s+1$ but not large enough to include facility s. Since the earthquake can only propagate continuously, i.e., it cannot reach a facility without impacting facilities that are closer to the epicenter, it is easy to see that only $\boldsymbol{\xi}^s$, $s = 0, 1, \ldots, J$ are possible disruption scenarios and the probability of senario $\boldsymbol{\xi}^s$ is equal to $q_{s+1} - q_s$.

A direct result from Lemma 2 is the worst-case correlation. Let ρ_{jk}^* be the worst-case correlation between locations j and k, with $j<k$. It is easy to verify that

$$\rho_{jk}^* = \sqrt{\frac{q_j(1-q_k)}{q_k(1-q_j)}}. \tag{4}$$

Two observations can be made. First, as a result of supermodularity, the worst-case correlation achieves the maximum correlation with the given marginal disruption probability. Second, the correlation is stronger between locations with similar marginal disruption probabilities. We think this partially reflects practical situations in which disruptions are caused by natural disasters and severe weather hazards. Facilities that are geographically close to each other tend to have similar disruption probabilities, and they are also more likely to be disrupted simultaneously by common hazards.

Another observation from Lemma 2 is that the worst-case disruption distribution only depends on the marginal disruption probability but not on transportation cost. In the traditional implicit formulation (IR) model for RUFL, customers are assigned to multiple backup facilities with different backup levels. The level r backup facility will only be used if level 1 through level $r-1$ backup facilities are disrupted. Under independent disruptions, it is optimal to assign backup facilities level by level in increasing order of transportation cost without considering reliability, as long as the number of backup levels is sufficiently large (Cui et al. 2010). However, under the worst-case correlated distribution, if the level r backup facility is less reliable than the level $r-1$ backup facility, it will be disrupted whenever the level $r-1$ facility is disrupted. Thus, assigning a less reliable facility as a higher level backup is meaningless. This shows that when disruptions are correlated, one needs to consider both transportation cost and reliability in determining backup levels.

Using the worst-case disruption distribution, we obtain an equivalent formulation of the DR-RUFL problem, which we refer to as the worst-case reliable uncapacitated fixed-charge location (WC-RUFL) problem.

Proposition 1 (Equivalent Formulation). *The DR-RUFL problem is equivalent to*

$$\text{(WC-RUFL)} \quad \min_{\mathbf{x}\in\mathscr{X}} \left\{ \sum_{j\in\mathcal{J}} f_j x_j + \sum_{s\in\mathcal{J}} (q_{s+1} - q_s) h(\mathbf{x}, \boldsymbol{\xi}^s) \right\}.$$

The WC-RUFL problem is a stochastic program with only $J+1$ scenarios and thus can be solved efficiently using standard methods such as Benders' decomposition.

4. Numerical Results

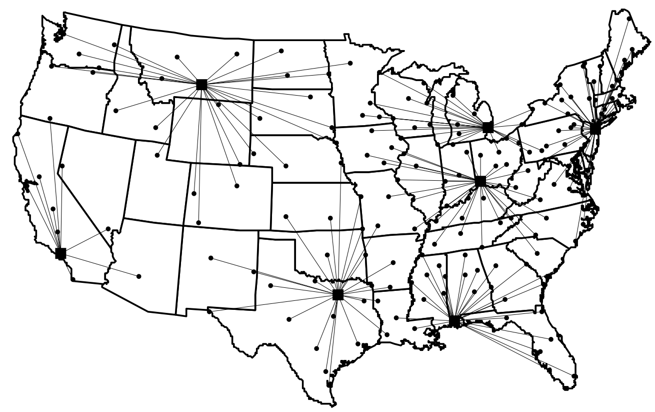
In this section we use numerical results to show the advantage of the distributionally robust model over the traditional model that assumes independent disruptions. First, we will present an example for supply chain network design and show how considering disruption correlation affects the optimal design. Then we will compare the two models in a numerical study with simulated data and draw managerial insights.

4.1. A Supply Chain Network Design Example

We apply the distributionally robust reliable facility location model to solve a supply chain network design problem under the threat of severe weather hazards. The supply chain operations data are adopted from a case study in Ballou (2004). A company that produces a line of industrial cleaning compounds is planning its distribution center (DC) network to serve its customers located throughout the 48 contiguous United States. The company has more than 70,000 individual customers that are aggregated into 191 demand points. Demand volume is estimated using historical sales data, with annual system-wide demand equal to 147 million pounds. There are 45 potential DC locations. The fixed cost of opening a DC is proportional to its unit storage cost. A linear regression model is used to estimate the unit transportation cost from each DC to each demand point as a function of the travel distance.

We assume the supply chain is under the threat of severe weather hazards. We use a data set from

Figure 1 Design I



the Storm Prediction Center of NOAA to estimate the marginal disruption probabilities of the DCs. More details of the data set are available in the online appendix. When a DC is disrupted, demand points that are assigned to it have to be served by other DCs, which results in substantial cost increases. Furthermore, when multiple DCs are disrupted, some customers may experience service interruptions. The penalty for service interruption is set to be nine times of the product value to reflect a desired service level. Given these inputs, we seek to design a reliable supply chain to minimize the total cost consisting of the fixed cost of DCs and the expected transportation/penalty cost under random disruptions.

Since only the marginal disruption probabilities are available, we are faced with two options. The first is to assume that the disruptions are independent and thus to apply the traditional RUFL model (e.g., Cui et al. 2010). The optimal design (I) is shown in Figure 1, in which the squares represent the DCs, the dots represent the demand points, and the lines represent the assignment of demand points to DCs. The second option is to consider all joint distributions with the known marginal disruption probability and apply the DR-RUFL model. The optimal design (R) is shown in Figure 2. Both designs are summarized in

Figure 2 Design R

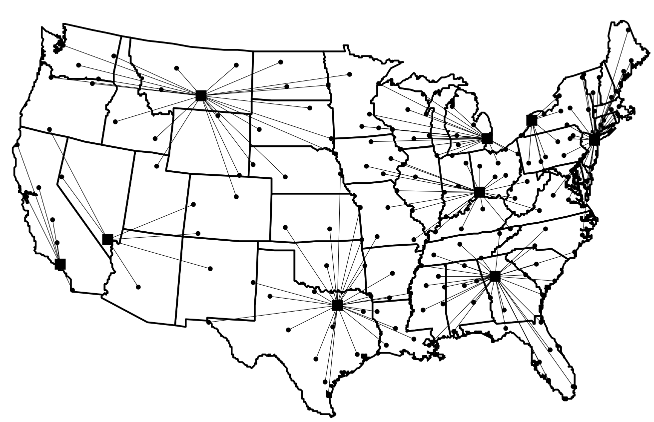


Table 2 Optimal Location Designs of the Two Models

Design I			Design R		
Location	Disruption probability	Demand allocation, %	Location	Disruption probability	Demand allocation, %
Arlington, TX	0.12	9.62	—	—	9.61
Billings, MT	0.02	7.19	—	—	7.12
Covington, KY	0.09	17.88	—	—	14.34
Detroit, MI	0.06	22.33	—	—	18.58
Long Beach, CA	0.02	10.95	—	—	9.15
New York, NY	0.05	23.04	—	—	21.44
Mobile, AL	0.08	8.99	Atlanta, GA	0.10	12.49
			Buffalo, NY	0.02	5.18
			Las Vegas, NV	0.01	2.10

Note. For design R, "—" means the location or probability is the same as in design I.

Table 2. We can see that the two designs are very similar. They share six common DCs, whereas the DC at Mobile, AL, in design I, is relocated to Atlanta, GA, in design R. Also, two additional DCs are opened at Buffalo, NY, and Las Vegas, NV, in design R. When there is no disruption, or when disruptions are independent, these changes result in very small differences in the total cost. As shown in Table 3, when there is no disruption, or when the disruptions are independent, design I performs slightly better than design R. Implementing design R will increase the cost by 1.48% or 1.37%, respectively. However, under the worst-case distribution, design R performs much better than design I. Implementing design I will lead to a cost increase of over 25%.

To understand why design I can result in such a significant cost increase under correlated disruptions, consider the two DCs in New York, NY, and Detroit, MI. In design I, these two DCs handle more than 45% of the sales. For many of the demand points assigned to them, one of these two DCs serves as the primary backup for the other. However, since their marginal disruption probabilities are very close (0.05 and 0.06, respectively), the correlation between their disruptions can be as high as 90%. When both are disrupted, the closest backup DC is in Covington, KY. However, this DC has a higher disruption probability of 0.09, and thus cannot provide effective backup under the worst-case distribution. As a result, when both the New York DC and the Detroit DC are disrupted, a large amount of goods must be shipped from DCs that are more than a thousand miles away, which incurs high additional costs. In design R, a more reliable DC is added at Buffalo, NY. Although it only handles 5% of the sales under normal conditions, it can provide effective backup in case of disruptions. Another example can be seen from the DC at Long Beach, CA. This DC serves the California market, which accounts for more than 9% of the nationwide sales. However, it is geographically isolated from all the other DCs. As a result, disruptions to this DC will result in significant cost increases. In practice, people have also observed that the isolation of the California market results in high supply chain disruption risks (Bussey 2012). Design R improves supply chain reliability of the California market by adding an additional DC in Las Vegas, NV.

The worst-case distribution is a conservative estimate of the actual distribution. However, the additional cost of design R (i.e., the increase in cost when there are no disruptions or when the disruptions are independent) is small, but the potential savings are huge. This makes design R more desirable and practical to implement. As we mentioned in §3, there are other reasons why one should consider the worst-case distribution rather than the independent distribution. In the next subsection we show this in a numerical study using simulated data.

4.2. Numerical Study

We will compare the robust model and the traditional model that assumes independent disruptions in a more comprehensive numerical study. Instead

Table 3 Comparison of the Performance of the Two Designs

	Design I			Design R		
Actual distribution	Cost	Increase	Relative increase, %	Cost	Increase	Relative increase, %
No disruption	2,423,092	—	—	2,459,074	35,982	1.48
Independent	2,444,220	—	—	2,477,740	33,519	1.37
Worst-case	3,800,646	769,545	25.39	3,031,101	—	—

Note. For designs I and R, "—" means there is no cost increase, since the solution is optimal under the actual distribution.

Table 4 Levels for Important Factors

Factor	Low	Medium	High
α	0.1	0.2	0.3
θ	200	400	800
ω	20,000	40,000	80,000

of using the severe weather hazard data from the NOAA, we use the simulated disruption probabilities in Cui et al. (2010). Let α be the probability that a disastrous event occurs at a certain source. The disaster then propagates and causes disruptions to facilities at different distances from the source. The marginal disruption probability decreases exponentially with the distance. Let D_j be the distance of location j from the source. The marginal disruption probability of location j is given by $q_j = \alpha \exp(-D_j/\theta)$. A larger θ means the disruption propagation effect is stronger. The source disaster probability α, the disruption propagation factor θ, along with the service interruption penalty, denoted by ω, are the key factors that significantly affect the cost and the optimal design. For each factor, we consider three levels as shown in Table 4, which gives us 27 different combinations. For demand, fixed cost, and transportation cost, we use the same data set as in Snyder and Daskin (2005) and Cui et al. (2010), which is a 49-node data set adopted from Daskin (1995). More details of the data set and additional results using a larger data set in Daskin (1995) are available in the online appendix. The robust model is solved using an accelerated Benders' decomposition algorithm. The traditional model is solved using the search-and-cut (SnC) algorithm in Aboolian et al. (2013), which to our knowledge is the state-of-the-art algorithm for the traditional RUFL model. Both algorithms are implemented and tested using ILOG CPLEX 12.4 with MATLAB R2009b on an Intel Core i7-930 2.80 GHz quad core processor running 64-bit Windows 7. The SnC algorithm uses four levels of backup and a neighborhood size of three (for details, please refer to Aboolian et al. 2013), and terminates with a 0.1% optimality gap or a maximum run time of 7,200 seconds, whichever occurs first.

Table 5 summarizes the solutions under different choices of α, ω, and θ. The subscript R represents the robust model, and the subscript I represents the traditional model with independent disruptions; n is the number of open facilities in the optimal solution; z is the optimal expected cost; and Δz is the regret, i.e., the increase in cost when the optimal solution under one disruption distribution is erroneously used for the other disruption distribution. For example, Δz_R is the regret if the optimal solution under independent disruptions is used when the disruptions are actually worst-case correlated; $\%\Delta z$ is the percentage relative regret, i.e., $\%\Delta z = 100 \times \Delta z/z$. CPU is the computation time, and GAP is the optimality gap when the algorithm terminates.

From Table 5 we have several observations. First, comparing columns n_R and n_I, we see that the number of open facilities in the robust solution is greater than or equal to that of the independent solution for all instances. This shows that more facilities are required to mitigate correlated disruptions. Second, from columns Δz_R and $\%\Delta z_R$, we see that failing to consider disruption correlation could lead to significant loss, with an average regret of 187,000 or 11.98%. For some instances, the relative regret is more than 20%. On the other hand, from columns $\%\Delta z_I$ and $\%\Delta z_I$, we see that although assuming the worst-case correlation is conservative, it does not lead to a significant cost increase even when disruptions are independent, with an average regret of 25,000 or 2.76%. For all instances, relative regret is less than 8%. Finally, for computational performance, comparing columns CPU_R and CPU_I, and columns GAP_R and GAP_I, we see that the robust model requires much less computational effort than the traditional model. This gives the robust model a great advantage for solving large-scale problems.

From Table 5 we also see that the performance of the solutions is affected significantly by the parameters α, ω, and θ, which measure the magnitude of the disruption risk. In Figure 3 we show the impact of these factors on the regret of the two models. Consider the source disaster probability, α, for example. The x-axis shows different values for α. For each given α, we consider different combinations of the other two factors, i.e., ω and θ, and calculate the average regret. We see that the regret of the independent model increases dramatically as α increases. On the other hand, the regret of the robust model only increases mildly. Similar results are observed for the service interruption penalty, ω, and disruption propagation effect, θ. In supply chain risk management, previous research shows that firms tend to underestimate disruption risks (Tang 2006). Our results show that the independent model will result in much higher loss than expected because of underestimation of disruption risks, whereas the performance of the robust model is less sensitive to such underestimation.

So far we have compared the robust model and the independent model under two extreme cases, i.e., either when disruptions are independent or when they are worst-case correlated. As we mentioned in §1, supply chain disruptions caused by natural disasters or severe weather conditions are typically positively correlated. However, the correlation will be lower than the worst case. Thus, the actual disruption distribution will be an intermediate case between the two extreme cases. We compare the performance

Table 5 Selected Results for the 49-Node Data Set

α	ω	θ	n_R	z_R	Δz_R	$\%\Delta z_R$	CPU_R	GAP_R	n_I	z_I	Δz_I	$\%\Delta z_I$	CPU_I	GAP_I
0.1	20,000	200	6	8.66	0.00	0.04	2.64	0.00	5	8.63	0.00	0.00	26.08	0.02
0.2	20,000	200	6	8.74	0.00	0.04	1.98	0.00	5	8.68	0.00	0.02	170.57	0.00
0.3	20,000	200	6	8.82	0.00	0.06	2.81	0.00	5	8.74	0.00	−0.04	200.14	0.07
0.1	40,000	200	6	8.66	0.00	0.04	2.80	0.00	5	8.63	0.00	0.00	30.32	0.02
0.2	40,000	200	6	8.75	0.00	0.04	2.54	0.00	5	8.68	0.00	0.02	194.73	0.00
0.3	40,000	200	6	8.83	0.00	0.06	2.36	0.00	5	8.74	0.00	−0.04	223.67	0.07
0.1	80,000	200	6	8.67	0.00	0.04	2.80	0.00	5	8.63	0.00	0.00	33.42	0.02
0.2	80,000	200	6	8.76	0.00	0.04	2.84	0.00	5	8.68	0.00	0.02	217.05	0.00
0.3	80,000	200	6	8.86	0.00	0.06	2.47	0.00	5	8.74	0.00	−0.04	251.07	0.07
0.1	20,000	400	6	9.39	0.02	0.20	2.14	0.00	6	8.74	−0.01	−0.08	348.97	0.10
0.2	20,000	400	7	10.13	0.09	0.93	4.11	0.00	6	8.90	0.23	2.60	1,772.63	0.09
0.3	20,000	400	7	10.80	0.24	2.24	2.84	0.00	6	9.05	0.24	2.70	4,905.02	0.07
0.1	40,000	400	7	9.76	0.10	1.05	3.20	0.00	6	8.74	0.22	2.57	396.79	0.10
0.2	40,000	400	7	10.70	0.43	3.98	2.57	0.00	6	8.90	0.23	2.60	2,013.53	0.09
0.3	40,000	400	7	11.62	0.77	6.64	3.56	0.00	6	9.05	0.39	4.35	5,551.85	0.07
0.1	80,000	400	7	10.33	0.43	4.21	2.75	0.00	6	8.74	0.22	2.57	442.22	0.10
0.2	80,000	400	7	11.75	1.18	10.07	3.69	0.00	6	8.90	0.38	4.28	2,244.51	0.09
0.3	80,000	400	7	13.15	1.95	14.84	3.88	0.00	6	9.05	0.39	4.35	6,184.70	0.07
0.1	20,000	800	7	13.36	0.73	5.47	2.89	0.00	6	8.94	0.35	3.89	1,225.04	0.00
0.2	20,000	800	7	17.75	1.80	10.16	2.12	0.00	6	9.30	0.33	3.50	7,237.74	1.13
0.3	20,000	800	8	21.94	3.11	14.16	2.94	0.00	6	9.68	0.78	8.09	7,234.39	10.36
0.1	40,000	800	7	16.90	1.90	11.27	3.05	0.00	6	8.94	0.35	3.89	1,394.15	0.00
0.2	40,000	800	7	24.85	4.15	16.72	2.87	0.00	6	9.30	0.33	3.50	7,219.95	1.43
0.3	40,000	800	8	32.57	6.63	20.35	2.76	0.00	6	9.68	0.78	8.09	7,273.99	6.63
0.1	80,000	800	7	23.99	4.25	17.73	2.61	0.00	6	8.94	0.35	3.89	1,552.36	0.00
0.2	80,000	800	7	39.03	8.85	22.68	2.88	0.00	6	9.30	0.33	3.50	7,304.21	1.64
0.3	80,000	800	8	53.85	13.68	25.40	2.73	0.00	6	9.68	0.78	8.09	7,209.54	7.02
Average			6.74	15.58	1.87	11.98	2.85	0.00	5.67	8.96	0.25	2.76	2,698.47	1.08

Notes. R: robust model; *I*: independent model; *n*: number of facilities; *z*: expected cost ($\times 10^5$); Δz: regret ($\times 10^5$); $\%\Delta z$: relative regret (%); CPU: computation time (s); GAP: optimality gap at termination (%).

of the two models under such intermediate cases. Assume the disruption correlation between location j and k is given by $\beta\rho^*_{jk}$, where ρ^*_{jk} is the worst-case correlation given in (4) and $\beta \in [0, 1]$ is a parameter that controls the degree of correlation. We use simulation to evaluate the expected cost of the optimal robust solutions and independent solutions under different choices of β. Figure 4 shows the average expected cost for the different instances in Table 5. We see that even though the independent model has a slightly lower average expected cost when the disruptions are close to independent (i.e., when β is close to 0), the robust model starts to outperform the independent model when the disruptions are mildly correlated (e.g., when $\beta = 0.3$) and achieves more substantial advantages under higher correlations.

Figure 3 Impact of Important Factors on Expected Regret

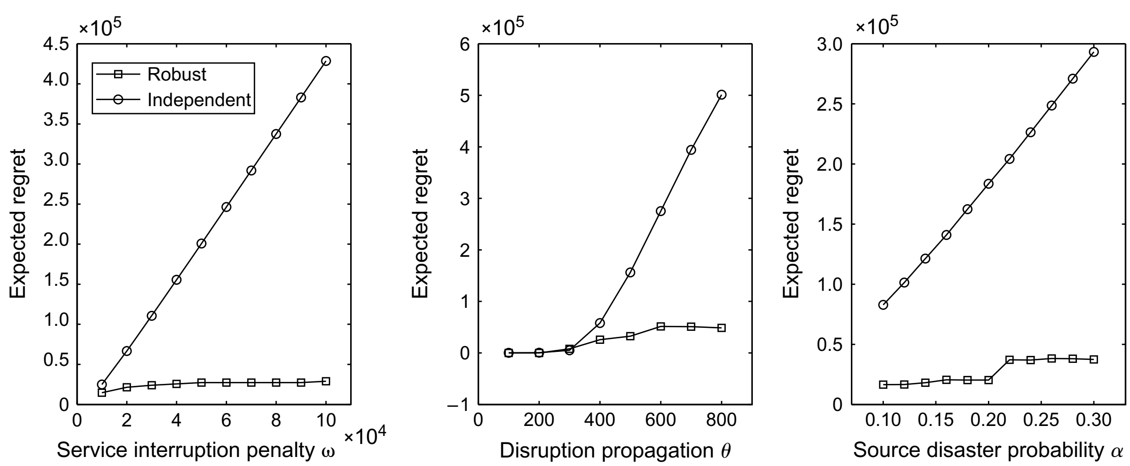
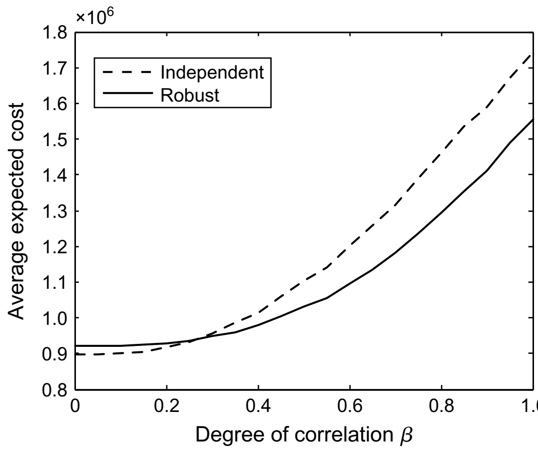


Figure 4 **Average Expected Cost Under Different Degrees of Disruption Correlation**



One common criticism of robust optimization is that it focuses on the worst case; thus, its solution can be overly conservative and too expensive to implement. In supply chain risk management, managers are usually unwilling to make a large investment, unless it can be justified by a cost/benefit analysis (Tang 2006). To address this issue, we consider a weighted-average objective function $\phi^\gamma(\mathbf{x}) = \gamma\phi^1(\mathbf{x}) + (1-\gamma)\phi^0(\mathbf{x})$, where ϕ^1 is the objective of the DR-RUFL problem and ϕ^0 is the total cost when there is no disruption. We refer to ϕ^0 as the normal operating cost and the weight $\gamma \in [0, 1]$ as the conservativeness factor. Let $\mathbf{x}^\gamma = \arg\min\{\phi^\gamma(\mathbf{x})\}$, i.e., the optimal solution with conservativeness factor γ; $\mathbf{x}^0$ minimizes the normal operating cost, i.e., it is the most cost-effective but also the least reliable location design. Applying a more reliable design, $\mathbf{x}^\gamma$, with $\gamma > 0$, has two competing effects. On the one hand, it reduces the expected cost under disruptions by $\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^\gamma)$, which is the benefit of the reliable design. On the other hand, it increases the normal operating cost by $\phi^0(\mathbf{x}^\gamma) - \phi^0(\mathbf{x}^0)$, which can be considered the cost of the reliable design. Figure 5 compares the benefit and cost under different choices of γ. We see that a large benefit can be achieved with a relatively low cost. The most reliable but also the most conservative design, $\mathbf{x}^1$, is obtained by setting the conservativeness factor $\gamma = 1$. It will result in the largest benefit $\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^1)$ and the highest cost $\phi^0(\mathbf{x}^0) - \phi^0(\mathbf{x}^1)$. Using a conservative factor $\gamma \in (0, 1)$, $(\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^\gamma))/(\phi^1(\mathbf{x}^0) - \phi^1(\mathbf{x}^1))$ is the proportion of the largest benefit that $\mathbf{x}^\gamma$ captures. Similarly, $(\phi^0(\mathbf{x}^\gamma) - \phi^0(\mathbf{x}^0))/(\phi^0(\mathbf{x}^1) - \phi^0(\mathbf{x}^0))$ is the proportion of the highest cost that $\mathbf{x}^\gamma$ incurs. Figure 6 compares these two proportions for different choices of γ. We see that over 90% of the largest benefit can be captured using a small conservative factor (e.g., $\gamma = 0.2$) while incurring only 40% of the highest cost. This shows that the DR-RUFL model is not expensive to implement. Managers can use a small conservativeness factor and still capture most of the benefit of the model.

Figure 5 **Benefit and Cost of the Robust Design with Different Conservativeness Factors**

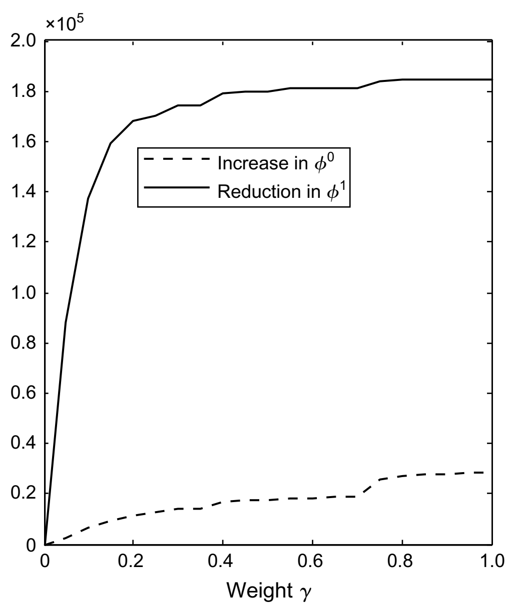
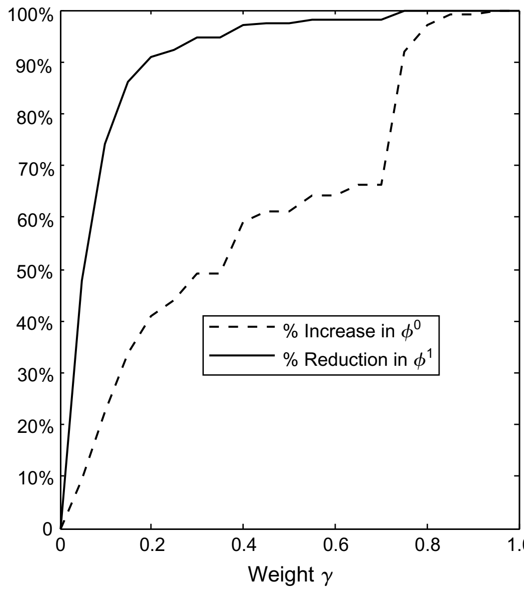


Figure 6 **Percentage of Largest Benefit and Highest Cost with Different Conservativeness Factors**



5. Conclusions and Future Work

In this paper, we present a distributionally robust optimization model to incorporate correlated disruptions in reliable facility location design. We find that this seemingly complicated problem is actually equivalent to a much simpler problem and can be solved efficiently. Our numerical results show that this model has several advantages compared to the traditional model assuming independent disruptions; thus, we believe it can serve as a promising alternative approach for reliable facility location design problems.

One limitation of our model is that it focuses on the worst-case distribution, which can be overly conservative in practice. In our future work, we plan to study a more general model in which the disruption correlation is explicitly given. Also, we focus on locating facilities for regular supply chain operations and assume that the demand is deterministic and not affected by the disruptions. When locating facilities for humanitarian operations in disaster relief, the demand will be highly uncertain and depend on the disruptions. We will incorporate uncertain demand in our future work. Our model is suitable for facility location design in a medium-to-large area (e.g., nationwide). New models need to be developed for problems in a relatively small area (e.g., a city). We also plan to consider the facility fortification problem under correlated disruptions and study the impact of correlation on the effect of fortification.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/msom.2015.0541.

Acknowledgments

The authors thank the editor-in-chief, the associate editor, and three referees for their helpful comments and suggestions. The severe weather hazard climatology data were provided by the NOAA Storm Prediction Center and the NOAA National Severe Storms Laboratory. The authors also thank Harold Brooks and Patrick Marsh for assistance in acquiring the data. This research was partially supported by the National Science Foundation [Grant CMMI 1265671] and the National Science Foundation of China [Grants 71210002, 71272058, and 71332005].

References

Aboolian R, Cui T, Shen Z-JM (2013) An efficient approach for solving reliable facility location models. *INFORMS J. Comput.* 25(4):720–729.

Agrawal S, Ding Y, Saberi A, Ye Y (2010) Correlation robust stochastic optimization. *Proc. Twenty-First Annual ACM-SIAM Sympos. Discrete Algorithms* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia), 1087–1096.

Agrawal S, Ding Y, Saberi A, Ye Y (2012) Price of correlations in stochastic optimization. *Oper. Res.* 60(1):150–162.

An Y, Zhang Y, Zeng B (2015) The reliable hub-and-spoke design problem: Models and algorithms. *Transportation Res.* B 77: 103–122.

An Y, Zeng B, Zhang Y, Zhao L (2014) Reliable p-median facility location problem: Two-stage robust models and algorithms. *Transportation Res. B* 64:54–72.

Ballou RH (2004) *Business Logistics/Supply Chain Management*, 5th ed. (Pearson/Prentice Hall, Upper Saddle River, NJ).

Berman O, Krass D, Menezes MBC (2007) Facility reliability issues in network p-median problems: Strategic centralization and co-location effects. *Oper. Res.* 55(2):332–350.

Berman O, Krass D, Menezes MBC (2013) Location and reliability problems on a line: Impact of objectives and correlated failures on optimal location patterns. *Omega* 41(4):766–779.

Bertsimas D, Natarajan K, Teo C-P (2004) Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds. *SIAM J. Optim.* 15(1):185–209.

Bussey J (2012) California's gas price: Is there a villain? *Wall Street Journal* (October 19), http://www.wsj.coms/articles/SB10000872396390444592704578064851501999088.

Chen Q, Li X, Ouyang Y (2011) Joint inventory-location problem under the risk of probabilistic facility disruptions. *Transportation Res. B* 45(7):991–1003.

Church RL, Scaparra MP (2007) Protecting critical assets: The r-interdiction median problem with fortification. *Geographical Anal.* 39(2):129–146.

CNN (2014) Hurricane Sandy fast facts. (November 5), http://www.cnn.com/2013/07/13/world/americas/hurricane-sandy-fast-facts/.

Cui T, Ouyang Y, Shen Z-JM (2010) Reliable facility location design under the risk of disruptions. *Oper. Res.* 58(4):998–1011.

Daskin MS (1995) *Network and Discrete Location* (John Wiley & Sons, New York).

Edmonds J (1971) Matroids and the greedy algorithm. *Math. Programming* 1(1):127–136.

Li Q, Zeng B, Savachkin A (2013a) Reliable facility location design under disruptions. *Comput. Oper. Res.* 40(4):901–909.

Li X, Ouyang Y (2010) A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. *Transportation Res. B* 44(4):535–548.

Li X, Ouyang Y (2012) Reliable traffic sensor deployment under probabilistic disruptions and generalized surveillance effectiveness measures. *Oper. Res.* 60(5):1183–1198.

Li X, Ouyang Y, Peng F (2013b) A supporting station model for reliable infrastructure location design under interdependent disruptions. *Transportation Res. E* 60:80–93.

Liberatore F, Scaparra MP, Daskin MS (2012) Hedging against disruptions with ripple effects in location analysis. *Omega* 40(1):21–30.

Lim MK, Bassamboo A, Chopra S, Daskin MS (2013) Facility location decisions with random disruptions and imperfect estimation. *Manufacturing Service Oper. Management* 15(2):239–249.

Lim MK, Daskin MS, Bassamboo A, Chopra S (2010) A facility reliability problem: Formulation, properties, and algorithm. *Naval Res. Logist.* 57(1):58–70.

Scaparra MP, Church RL (2008) An exact solution approach for the interdiction median problem with fortification. *Eur. J. Oper. Res.* 189(1):76–92.

Shen Z-JM, Zhan RL, Zhang J (2011) The reliable facility location problem: Formulations, heuristics, and approximation algorithms. *INFORMS J. Comput.* 23(3):470–482.

Simchi-Levi D, Schmidt W, Wei Y (2014) From superstorms to factory fires: Managing unpredictable supply-chain disruptions. *Harvard Bus. Rev.* 92(1–2):96–101.

Snyder LV, Daskin MS (2005) Reliability models for facility location: The expected failure cost case. *Transportation Sci.* 39(3):400–416.

Snyder LV, Atan Z, Peng P, Rong Y, Schmitt AJ, Sinsoysal B (2014) OR/MS models for supply chain disruptions: A review. Working paper, Lehigh University, Bethlehem, PA. http://papers.ssrn.com/abstract=1689882.

Tang CS (2006) Perspectives in supply chain risk management. *Internat. J. Production Econom.* 103(2):451–488.