Transportation Research Part B 44 (2010) 535–548

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions

Xiaopeng Li, Yanfeng Ouyang *

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 4 March 2009
Received in revised form 13 September 2009
Accepted 14 September 2009

Keywords:
Reliable facility location
Continuum approximation
Spatial correlation
Disruption

ABSTRACT

This paper studies the reliable uncapacitated fixed charge location problem (RUFL) where facilities are subject to spatially correlated disruptions that occur with location-dependent probabilities (due to reasons such as natural or man-made disasters). If a facility fails, its customers are diverted to other facilities and incur excessive transportation cost. We develop a continuum approximation (CA) model to minimize the sum of initial facility construction costs and expected customer transportation costs under normal and failure scenarios. The paper presents ways to formulate the correlation among adjacent facility disruptions, and incorporates such correlations into the CA model. Numerical experiments are conducted to illustrate how the proposed model can be used to optimize facility location design, and how the correlations influence the total system cost.

1. Introduction

The uncapacitated fixed charge location problem (UFLP) seeks optimal facility locations to minimize the one-time investment for facility constructions and the long-run transportation costs for serving spatially distributed customers. This topic has been studied extensively; see Daskin (1995) and Drezner (1995) for comprehensive surveys. Traditional models generally assume that the facilities, once built, remain operational forever. In reality, however, facility operations may be disrupted from time to time due to reasons such as natural disasters, power outages, operational accidents, labor actions or terrorist attacks. The failure of a facility will force its customers to either travel longer distances to obtain service from another facility, or give up service and incur a penalty. Either way, system operation cost increases and customer satisfaction deteriorates. The adverse effect may be further exacerbated if multiple facilities fail simultaneously.

Regarding uncertainty, earlier literature focused on facility congestion that arises from stochastic demand and attempted to enhance system availability by providing redundancy (Daskin, 1982, 1983; Ball and Lin, 1993; Revelle and Hogan, 1989; Batta et al., 1989). Recently, reliable facility location models were developed to design facility locations and customer assignment plans when probabilistic facility disruptions are possible. Snyder and Daskin (2005) studied the reliable uncapacitated fixed charge location problem, RUFL, assuming that facility disruptions occur independently with equal probability. The problem is formulated into a mixed integer program and solved with Lagrangian relaxation. Cui et al. (2009) further developed mixed integer program models to allow site-dependent disruption probabilities. Compared with UFLP, these new models have significantly improved system reliability and reduced the expected overall cost across normal and failure scenarios. These models are recently applied to deploy sensors for network traffic surveillance (Ouyang et al., 2009; Li and Ouyang, 2009).

* Corresponding author. Tel.: +1 217 333 9858; fax: +1 217 333 1924.
E-mail address: yfouyang@illinois.edu (Y. Ouyang).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2009.09.004

Most of the traditional discrete location models are NP-hard and suffer from excessive computational burden. To overcome this challenge, continuum approximation (CA) models (Newell, 1971, 1973; Daganzo, 1984a,b; Daganzo and Newell, 1986; Ouyang and Daganzo, 2006) are often developed to provide good approximate solutions to large-scale logistics problems in various contexts (Hall, 1984, 1986, 1989; Campbell, 1993a,b; Daganzo, 1999; Dasci and Verter, 2001). See Langevin et al. (1996) and Daganzo (2005) for reviews. In the context of RUFL, Cui et al. (2009) developed a CA model as an alternative for solving large-scale problems, and compared its performance with that of the discrete counterparts.

In the real world, many facility disruption cases exhibit strong spatial correlations, probably because neighboring facilities are likely to be exposed to similar hazards. Such correlations significantly influence the facility failure pattern over space and hence the system operation. For example, under positive correlations (e.g., due to natural disasters, power grid outages), neighboring facilities are more likely to fail simultaneously, and the customers will find it more costly to reach a functioning facility. In contrast, under negative correlations,[1] neighboring facilities tend to back up each other to avoid long distance travels of the customers.

However, to the authors' best knowledge, spatial correlation among facility disruptions has not been addressed in the RUFL literature. This paper aims to fill this gap by developing a reliable facility location design framework that allows correlated and site-dependent facility disruptions. Accounting for such correlations in the discrete location modeling framework generally requires scenario-based formulation, which is computationally prohibitive due to the exploding number of possible scenarios. We hence build our model upon the continuum approximation approach to estimate and design the complex system. The structure of the spatial correlation is modeled in a variety of ways to provide flexibility in addressing real-world scenarios. Numerical experiments are conducted to illustrate applications of the model. Insights are also drawn through comparisons between the optimal solutions under various spatial correlation patterns and those under independent failures. The impact of disruption correlation on the system cost is found to be significant when both failure probabilities and penalty costs for unserved customer demand are high.

The remainder of the paper is organized as follows. Section 2 introduces the notation and problem definition. Section 3 presents the formulation and solution techniques for the CA model. Section 4 presents multiple ways to model spatial correlation under different application contexts. Section 5 applies the CA model to numerical examples and draw insights into the impact of correlations. Section 6 presents concluding remarks and suggests future research.

2. Model formulation

In a two-dimensional space $\mathscr{S} \subseteq \mathbb{R}^2$, the customer demand per unit area is denoted by $\lambda(x),\ \forall x \in \mathscr{S}$. A facility can be built at any location $x \in \mathscr{S}$ with a fixed opening cost $f(x)$. The decision variables are the number of facilities, N, and their locations $\mathbf{x} := \{x_1, x_2, \ldots, x_N\} \subseteq \mathscr{S}$. Suppose that the transportation cost for facility j to serve a unit demand at x is $\alpha_t \|x - x_j\|$, where α_t is a constant factor and $\|x - x_j\|$ is the Euclidean distance. We further assume that a customer at x, if served, shall only be served by a facility within a distance $D(x)$; if not served, the customer will incur a penalty cost $\alpha_p D(x)$, where $\alpha_p \geqslant \alpha_t$. Snyder and Daskin (2005) attributed the penalty cost to lost-sales or emergency-purchases.

We assume that the customers have complete information on facility disruptions[2] and choose facilities for service accordingly. This is different from Cui et al. (2009) where each customer is preassigned to a sequence of prioritized facilities regardless of the failure scenario. We also assume, for simplicity, that the failure scenario does not change during the time that customers are traveling. At any time, customers at x will either visit a functioning facility within distance $D(x)$ if one is available, or bear the penalty cost $\alpha_p D(x)$. The optimal strategy has the following simple property.

Proposition 1. *Given a facility failure scenario, each customer should always visit the closest functioning facility within distance $D(x)$.*

Proof. If a customer visits an operational facility other than the closest one, redirecting this customer to the closest operational facility will always strictly reduce the transportation cost. Thus the original solution cannot be optimal. This completes the proof. □

Given facility location design $\mathbf{x}$, let $\overline{P}(x|\mathbf{x})$ denote the probability for the customer at x not to be served (which occurs if all facilities within distance $D(x)$ from x have failed), and let $P(x, x_j|\mathbf{x})$ denote the probability for this customer to be served by facility j (which occurs if facility j is functioning, $\|x - x_j\| \leqslant D(x)$, and all facilities closer to x have failed). The values of these probabilities should always satisfy

$$\overline{P}(x|\mathbf{x}) + \sum_{j=1}^{N} P(x, x_j|\mathbf{x}) = 1, \quad \forall x \in \mathscr{S}, \tag{1}$$

because any customer either receives service or incurs the penalty.

[1] In reality, negative failure correlation is rare but possible. For example, the facilities may compete against each other for limited critical resources (such as material supply or maintenance service), such that the failure of one facility helps other facilities to survive. In the context of terrorist attacks, the failure of one facility may raise alert and help prevent other facilities from being attacked.

[2] This assumption is reasonable given the rapid advancement of modern information technologies (such as Internet-and PDA-enabled service applications). It may not be totally realistic, however, if information availability is limited in certain situations (e.g., catastrophic disaster).

The objective is to minimize the expected overall cost with respect to $\mathbf{x}$, as follows:

$$\min_{\mathbf{x}} \sum_{j=1}^{N} f(x_j) + \alpha_p \int_{x\in\mathscr{S}} \lambda(x)D(x)\overline{P}(x|\mathbf{x})dx + \alpha_t \int_{x\in\mathscr{S}} \sum_{j=1}^{N} \lambda(x)\|x-x_j\|P(x,x_j|\mathbf{x})dx. \tag{2}$$

The three terms in (2) respectively represent the fixed facility opening costs, the expected penalty costs for unserved demands and the expected transportation costs for served demands.

Following the ideas in Cui et al. (2009), (2) can be transformed by partitioning $\mathscr{S}$ into service areas. From the perspective of a generic facility j, every customer in $\mathscr{S}$ can be assigned a service rank $r \in \{0,1,2,\ldots\}$ if facility j is the $(r+1)^{\text{th}}$ nearest facility to this customer. We define the *rank-r service area* of facility j, $\mathscr{A}_{j,r}$, as the subset of customers who are assigned a rank r by facility j. Obviously, the definition of $\{\mathscr{A}_{j,r},\ \forall j,r\}$ are purely based on the facility locations $\mathbf{x}$. For any j, $\{\mathscr{A}_{j,r},\ \forall r\}$ forms a non-overlapping partition of $\mathscr{S}$ when boundaries are ignored, i.e.,

$$\bigcup_{r=0}^{\infty} \mathscr{A}_{j,r} = S \quad \text{and} \quad \mathscr{A}_{j,r} \bigcap \mathscr{A}_{j,r'} = \emptyset,\ \forall r \neq r'.$$

With this, (2) can be rewritten as follows:

$$\min_{\mathbf{x}} \sum_{j=1}^{N} f(x_j) + \alpha_p \int_{x\in\mathscr{S}} \lambda(x)D(x)\overline{P}(x|\mathbf{x})dx + \alpha_t \sum_{j=1}^{N} \sum_{r} \int_{x\in\mathscr{A}_{j,r}} \lambda(x)\|x-x_j\|P(x,x_j|\mathbf{x})dx. \tag{3}$$

For notation convenience, from now on we will use $\overline{P}(x)$ and $P(x,x_j)$ to represent $\overline{P}(x|\mathbf{x})$ and $P(x,x_j|\mathbf{x})$ respectively.

3. Continuum approximation framework

This section presents a continuum approximation approach to the RUFL problem. Section 3.1 first discusses the optimal solution to an idealized case where the problem is IHI; i.e., $\mathscr{S}$ is an *infinite* and *homogeneous* plane and the facilities fail *independently*. Building on the results for IHI, Section 3.2 discusses how to incorporate correlated disruptions into the framework, and Section 3.3 further develops the continuum approximation (CA) model for the general problem where $\mathscr{S}$ is finite and heterogeneous.

3.1. Building block: the IHI problem

In an IHI problem, $\mathscr{S}$ is an infinite and homogeneous plane (i.e., $\mathscr{S} = \mathbb{R}^2$), all relevant parameters are constant everywhere (i.e., $D(x) = D,\ f(x) = f,\ \lambda(x) = \lambda,\ \forall x \in \mathscr{S}$), and every facility fails independently with an equal probability $q(x) = q$. Some properties of the optimal solution to IHI have been discussed in the literature. Toth (1959) has proven that for $q = 0$, the total cost is minimized when the *initial service areas* $\{\mathscr{A}_{j,0},\ \forall j\}$ form a regular hexagonal tessellation of $\mathscr{S}$ and each facility is located at the centroid of a hexagon. Cui et al. (2009) showed that the optimal partition for $q > 0$ should also follow the same regular hexagonal tessellation pattern.

The regular hexagonal tessellation pattern and the homogeneity of $\mathscr{S}$ imply that the only decision variable for the optimal design of IHI is the size of the hexagonal initial service area, which we denote by A. Fig. 1 illustrates how the service areas for an arbitrary facility j would partition $\mathscr{S}$. Proposition 2 below shows that all these service areas have the same size.

Proposition 2. *For an IHI problem,* $|\mathscr{A}_{j,r}| = A,\ \forall j, r.$

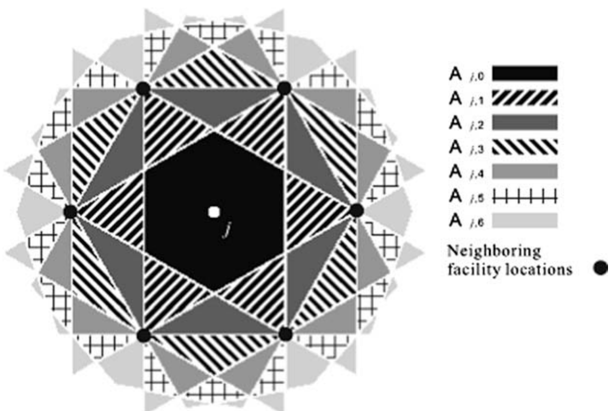

Fig. 1. Service area partition $\{\mathscr{A}_{j,r}, \forall r\}$ for the IHI problem (adapted from Cui et al. (2009)).

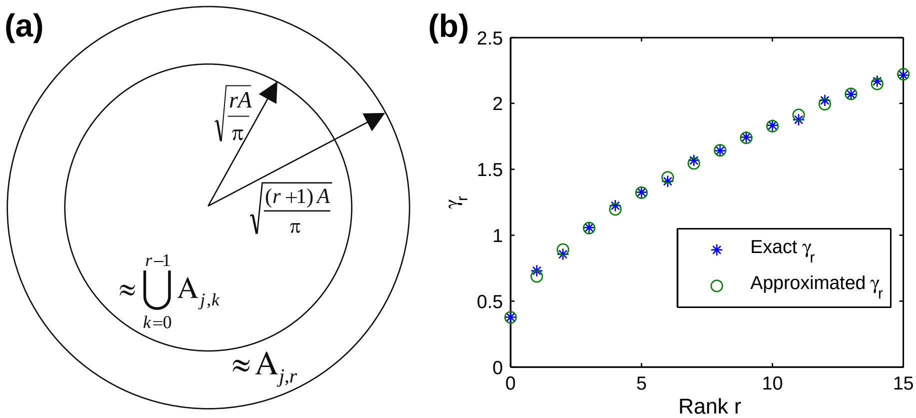

Fig. 2. (a) Approximation of $\mathscr{A}_{j,r}$ by a ring and (b) exact and approximated γ_r.

Proof. See Appendix A. □

The scalability of hexagons on the infinite plane implies that the average travel distance from the customers in $\mathscr{A}_{j,r}$ to facility j is proportional to $A^{1/2}$ and does not depend on j. We denote this average distance by $\gamma_r A^{1/2}$, where constant scalar γ_r can be calculated exactly for all r.[3] To obtain a closed-form approximation, we observe from Fig. 1 that $\mathscr{A}_{j,r}$ asymptotically converges to a ring as r increases. From Proposition 2, this ring's outer and inner circle radii are approximately $\sqrt{(r+1)A/\pi}$ and $\sqrt{rA/\pi}$ respectively; see Fig. 2a. By simple geometry, the average distance from the customers in this ring to the facility is $\frac{2}{3\sqrt{\pi}}[(r+1)^{3/2} - r^{3/2}] \cdot A^{1/2}$. Hence, for large r,

$$\gamma_r \approx \frac{2}{3\sqrt{\pi}}\left[(r+1)^{3/2} - r^{3/2}\right]. \tag{4}$$

Fig. 2b plots both the approximation (4) and the exact γ_r values. The approximation error is no more than 2% except for $r = 1, 2$, and it vanishes as r increases.

Optimizing objective function (3) for infinite and homogeneous $\mathscr{S}$ is equivalent to minimizing the expected total cost per unit area, which includes the unit-area facility opening cost C_f, the unit-area expected penalty cost C_p, and the unit-area expected transportation cost C_t. Obviously,

$$C_f = f/A. \tag{5}$$

The rest of this section provides closed-form approximations for C_p and C_t.

Note that $N_D(x)$, the number of facilities that a customer at x can visit within distance D, varies slightly with x. Hence, $\overline{P}(x) = q^{N_D(x)}$ varies with x as well. Let $\theta \in \mathbb{R}_+$ be the average value of $N_D(x)$ across $x \in \mathscr{S}$, and $\overline{P}$ the average value of $\overline{P}(x)$. The customer demand in $\mathscr{S}$ that each facility can potentially reach is $\lambda\pi D^2$, while asymptotically, each facility corresponds to λA customer demand. Hence $\lambda\pi D^2 = \lambda A \cdot \theta$, which yields $\theta = \pi D^2/A$.

If $\theta \leqslant 1$ ($\pi D^2 \leqslant A$), the situation is shown in Fig. 3. Only those customers within distance D from a facility will receive service with probability $(1-q)$; they incur penalty with probability q. All other customers incur penalty with probability 1.[4] Simple geometry yields $\overline{P}$ as follows:

$$\overline{P} = [\pi D^2 \cdot q + (A - \pi D^2) \cdot 1]/A = 1 - (1-q)\theta.$$

More generally, for $\theta > 1$ ($\pi D^2 > A$), customers lie in service areas of different ranks, as shown in Fig. 1. Exact calculation of $\overline{P}$ is tedious. However, since $N_D(x)$ obviously does not vary significantly across $\mathscr{S}$, $\overline{P}$ can be approximated by $\overline{P} \approx q^\theta$. Hence, we have

$$\overline{P} \approx \begin{cases} q^\theta, & \pi\theta^2 > 1, \\ 1 - (1-q)\theta, & \text{otherwise}, \end{cases} \tag{6}$$

and

$$C_p = \alpha_p \lambda D \overline{P}. \tag{7}$$

---

[3] See how the exact values are computed numerically in Appendix B. These exact values are unitless constants that could be directly used in the CA framework.

[4] If θ is very close to 1, there may be a very small fraction of customers near the hexagon boundaries with $N_D(x) > 1$. This exception is numerically negligible.

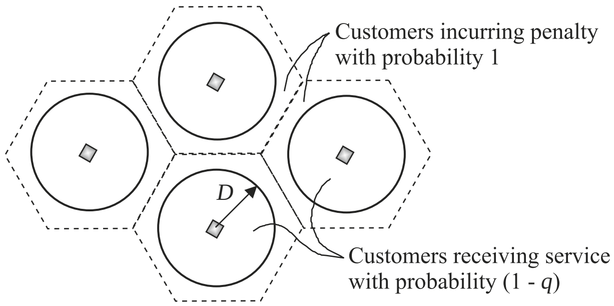

Fig. 3. Customer partition when $\theta \leqslant 1$.

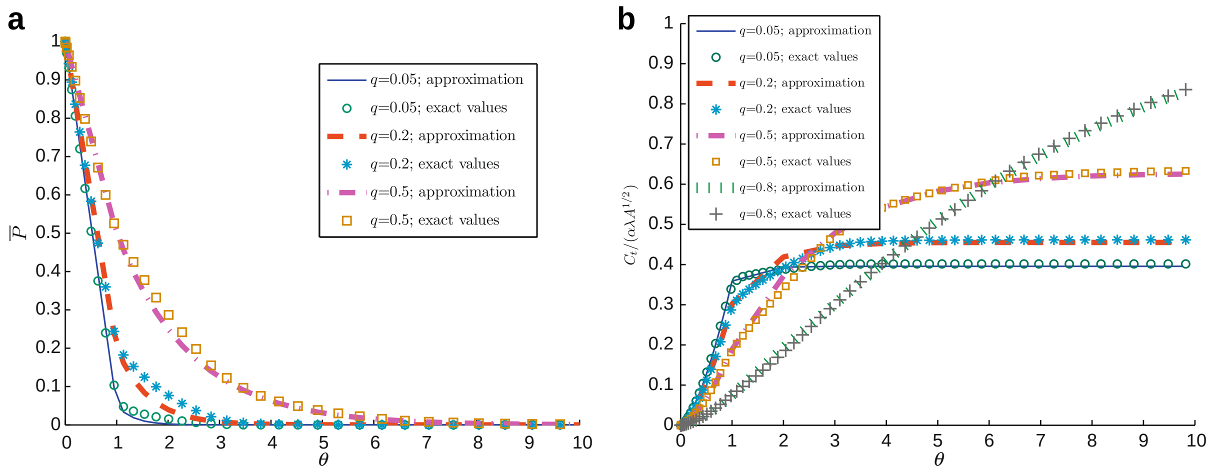

Fig. 4. Exact and approximated values of (a) $\overline{P}$ and (b) C_t.

Fig. 4a shows that Eq. (6) accurately predicts the exact value of $\overline{P}$.[5] The prediction error is almost 0 for $\theta > 3$ (for most realistic cases) and $\theta \leqslant 1$, and no more than 0.04 for θ around 2.

All customers in $\mathscr{A}_{j,r}$ receive service from facility j with equal probability, which we denote by P_r. Due to the independence of facility failures in IHI,

$$P_r = P(x, x_j) = (1-q)q^r, \quad \forall x \in \mathscr{A}_{j,r}. \tag{8}$$

In case $\theta > 1$ and $D \to \infty$, Proposition 2 implies that C_t can be directly calculated as follows:

$$C_t = \alpha_t \lambda A^{1/2} \sum_{r=0}^{\infty} P_r \gamma_r. \tag{9}$$

For finite D but $\theta > 1$, since θ may not be an integer, interpolation of (9) yields

$$C_t \approx \alpha_t \lambda A^{1/2} \left[\sum_{r=0}^{\lfloor\theta\rfloor-1} P_r \gamma_r + \bar{\theta} P_{\lfloor\theta\rfloor} \gamma_{\lfloor\theta\rfloor} \right], \tag{10}$$

where $\lfloor\cdot\rfloor$ is the floor operation and $\bar{\theta} = \theta - \lfloor\theta\rfloor$. For $\theta < 1$ (see Fig. 3), almost all served customers of facility j lie in the circle within $\mathscr{A}_{j_0}$ and their average distance from the facility is $\frac{2}{3}D$. When facility j does not fail (with probability $(1-q)$), these customers collectively incur service cost $\frac{2}{3}\alpha_t \lambda \pi D^3$. Hence, the expected service cost per unit area can be averaged across $\mathscr{A}_{j_0}$ as follows:

$$C_t = \frac{2}{3}\alpha_t \lambda \pi D^3 (1-q)/A = \frac{2}{3}\alpha_t \lambda A^{1/2} P_0 \sqrt{\theta^3/\pi}. \tag{11}$$

Eqs. (10) and (11) can be expressed as

[5] See Appendix B for details on how these exact values are computed.

$$C_t = \alpha_t \lambda A^{1/2} U(\theta, \mathbf{P}), \tag{12}$$

where $\mathbf{P} := \{P_r : \forall r\}$ and

$$U(\theta, \mathbf{P}) \approx \begin{cases} \sum_{r=0}^{\lfloor\theta\rfloor-1} P_r \gamma_r + \bar{\theta} P_{\lfloor\theta\rfloor} \gamma_{\lfloor\theta\rfloor}, & \text{if } \theta \geqslant 1, \\ \frac{2}{3} P_0 \sqrt{\theta^3/\pi}, & \text{otherwise}. \end{cases} \tag{13}$$

Obviously, the term $A^{1/2}U(\theta, \mathbf{P})$ represents the expected travel distance for a customer to reach a functioning facility. Fig. 4 compares the approximation formula (13) with the exact values.[6] Again, the error is almost 0% for $\theta > 3$ or $\theta \leqslant 1$, while the maximum percentage error is about 6%.

From (5), (7) and (12), the total cost per unit area for the IHI problem is

$$C := C_f + C_p + C_t = f/A + \alpha_p \lambda D \overline{P} + \alpha_t \lambda A^{1/2} U(\theta, \mathbf{P}). \tag{14}$$

In general, the optimal solution A^* does not have a simple analytical form because $\overline{P}$ and $U(\theta, \mathbf{P})$ are both functions of A. Section 3.3 introduces a simple bisecting method to find A^* efficiently.

3.2. Penalty and service probabilities under correlated disruptions

In the previous section, Eqs. (6) and (8) hold only when the facilities fail independently. Using these formulas will cause significant errors if facility disruptions are actually correlated, as indicated in the following proposition.

Proposition 3. *For any facility location design, the existence of positive (or negative) facility failure correlation increases (or decreases) the expected transportation and penalty cost per unit demand.*

Proof. See Appendix C. □

Hence, accurate estimation of the total cost mandates that the penalty probability $\overline{P}$ and the service probability P_r accommodate failure correlations. In the rest of this subsection, we provide a general formulation framework for $\overline{P}(x)$ and $P_r(x)$ based on conditional probabilities.

In the literature, conditional probabilities have been used to model general correlations of symmetric binary events.[7] Based on the Pascal's triangle, probabilities of symmetric disruptions can be represented as the product of facility failure probabilities conditional on the number of neighboring disruptions.

On the infinite homogeneous plane, we let $\{q_l, l = 0, 1, 2 \ldots\}$ denote the conditional failure probability of a facility given that: (i) this facility is the $(l+1)^{th}$ closest facility to a certain customer and (ii) all l closer facilities to this customer have failed (regardless of all other facilities on the plane). As such, q_0 represents the unconditional individual failure probability.[8] Generally, if q_l increases with l, the failure correlation is positive. For example, $q_l = 1, \ \forall l \geqslant 1$ yields the case of perfect correlation (i.e., facilities either all survive or all fail).[9] On the other hand, facility failure is negatively correlated if q_l decreases with l.

When $\{q_l, \ \forall l\}$ is known (e.g., from historical data), from the perspective of a customer, the probability for all m nearest facilities to fail simultaneously equals $\prod_{l=0}^{m-1} q_l$. For general $\theta \in \mathbb{R}_+, \ \overline{P}$ can be approximated by interpolating the probabilities for $\lfloor\theta\rfloor$ and $\lfloor\theta\rfloor + 1$ simultaneous failures, as follows:

$$\overline{P} \approx \begin{cases} (1-\bar{\theta}) \prod_{l=0}^{\lfloor\theta\rfloor-1} q_l + \bar{\theta} \prod_{l=0}^{\lfloor\theta\rfloor} q_l, & \theta > 1, \\ 1 - (1-q_0)\theta, & \text{otherwise}. \end{cases} \tag{15}$$

It is easy to observe that (6) bounds (15) from below/above under positive/negative correlations, indicating over-/under-estimation of penalty probability when correlation is ignored. Probability P_r equals the probability that all r nearest facilities to a customer fail while the $(r+1)^{\text{th}}$ facility survives; i.e., (8) shall be replaced by

$$P_r \approx (1 - q_r) \prod_{l=0}^{r-1} q_l. \tag{16}$$

In certain cases, the conditional probabilities may be dependent of A especially when the correlation magnitude is sensitive to the distance among facilities. The modeling framework described above remains applicable by simply specifying the appropriate $\{q_l(A), \ \forall l\}$.

---

[6] See Appendix B for details.

[7] Interested readers are referred to Bakkaloglu et al. (2002) and Tang and Iyer (1992) for reviews on this topic.

[8] If failure correlation is ignored, $\overline{P}$ and $P_r, \ \forall r$, shall be computed from (6) and (8) with probability $q = q_0$.

[9] Such an extreme case is usually induced by a shared failure source that causes simultaneous disruptions of all facilities. Examples may include outage in a power grid due to failure of the power plant.

3.3. CA model for heterogeneous space

We assume that in a finite heterogeneous space $\mathscr{S} \subset \mathbb{R}^2$, parameters $f(x),\ \lambda(x),\ D(x)$ vary slowly over $x \in \mathscr{S}$. Instead of looking for $\mathbf{x}$ directly, we propose to look for a continuous function, $A(x) \in \mathbb{R}_+,\ x \in \mathscr{S}$, that approximates the initial service area size of a facility near x; i.e., $A(x) \approx |\mathscr{A}_{j,0}|$ if $x \in \mathscr{A}_{j,0}$. We assume that $\mathscr{S}$ is far larger than $A(x)$; i.e., $|\mathscr{S}| \gg A(x),\ \forall x \in \mathscr{S}$. When all parameters are *approximately constant* over a region comparable to the size of several initial service areas, $\overline{P}(x),\ P_r(x),\ \theta(x)$ and $A(x)$ should also be approximately constant on that scale.[10]

We apply the cost formulation (14) to the neighborhood of x (i.e., imagining that this neighborhood is part of an infinite and homogeneous plane), while using the values of $f(x),\ \lambda(x),\ D(x)$ as the parameter input. Incorporating (14) into (3) yields the following:

$$\min_{A(x)} \int_{x \in \mathscr{S}} C(x, A(x)) dx, \tag{17}$$

where the total cost per unit area near x is

$$C(x, A(x)) := f(x)/A(x) + \alpha_p \lambda(x) D(x) \overline{P}(x) + \alpha_t \lambda(x) A^{1/2}(x) U(\theta(x), \mathbf{P}), \forall x \in \mathscr{S}. \tag{18}$$

Since the inverse of $A(x)$ represents the facility density at x, the number of facilities is

$$N \approx \int_{x \in \mathscr{S}} [A(x)]^{-1} dx. \tag{19}$$

For any x, (18) has only one scalar decision variable $A(x)$. We shall note that the values of $\overline{P}(x)$ and $U(\theta(x), \mathbf{P})$ depend on $\theta(x)$ and hence on $A(x)$, although $\mathbf{P}$ is independent of $A(x)$ as suggested by (8) and (16). Intuitively, the second and third terms in (18) should be increasing with $A(x)$, while the first term is decreasing with $A(x)$. Hence, the function $C(x, A(x))$ is likely to have a "V" shape with regard to $A(x)$. The optimal solution $A^*(x)$ can be obtained from a simple bisecting search.

The estimated optimal cost per unit area $C(x, A^*(x))$ and the optimal facility density function $[A^*(x)]^{-1}$ can be integrated across $\mathscr{S}$ to yield the total system cost C^* and the optimal number of facility N^*, respectively. Function $A^*(x),\ x \in \mathscr{S}$ and N^* can be used in the disk model (Ouyang and Daganzo, 2006) to design the optimal discrete facility locations. The disk model exerts repulsive forces to N^* disks that each represents a facility and its initial service area, and iteratively adjusts positions and sizes of these disks to achieve optimal layout. Interested readers are referred to Ouyang and Daganzo (2006) and Ouyang (2007) for more implementation details. These references have also shown that the total cost of the discrete design obtained from the disk model is very close to that estimated by (17).

4. Alternative correlation structures

Facility failure correlations can be modeled in a variety of ways. This section discusses two special cases. Section 4.1 simplifies the formulation with beta-binomial distributions when the correlation is always positive. Section 4.2 shows how to decompose $\overline{P}(x)$ and $P_r(x)$ into scenario-based probabilities in case that the facility failure mechanisms are known.

4.1. Positively correlated beta-binomial facility failure

The modeling approach in Section 3.2 requires a whole set of conditional failure probabilities $\{q_l,\ \forall l\}$ to be specified (most likely from historical data). This may be tedious in certain practical situations. As an alternative, the beta-binomial distribution has been used in various fields (such as computer science (Bakkaloglu et al., 2002; Goyal and Nicola, 1990) and biometrics (Griffiths, 1973)) to model positive failure correlations. The beta-binomial distribution, which we denote by $B_{n,a,b}$ with $a, b > 0$, only has three parameters. It is defined as the distribution for the number of failures in n symmetric success/failure experiments, while each experiment has a random failure probability p whose probability density function is

$$\frac{p^{a-1}(1-p)^{b-1}}{\int_0^1 p^{a-1}(1-p)^{b-1} dp}, \quad p \in [0, 1].$$

Accordingly, the probability that m out of n experiments fail is

$$B_{n,a,b}(m) := \binom{n}{m} \frac{[(a+m-1)(a+m-2)\cdots a][(b+n-m-1)(b+n-m-2)\cdots b]}{(a+b+n-1)(a+b+n-2)\cdots(a+b)}. \tag{20}$$

Eq. (20) can be equivalently structured in terms of the general conditional probabilities

$$q_l = \frac{B_{l+1,a,b}}{B_{l,a,b}} = \frac{a+l}{a+b+l},\ \forall l. \tag{21}$$

[10] Interested readers are referred to Sections 4.2, 4.4 and Appendix B in Cui et al. (2009) for discussions on the applicability and accuracy of the continuum approximation method.

$B_{n,a,b}$ has mean $n\frac{a}{a+b}$ and variance $n\frac{ab}{(a+b)^2}\frac{1+n/(a+b)}{1+1/(a+b)}$. Compared with the regular binomial distribution with n experiments and independent failure probability $\frac{a}{a+b}$, $B_{n,a,b}$ has the same mean but a larger variance; the amplification factor $\frac{1+n/(a+b)}{1+1/(a+b)}$ captures positive correlation among facility failure.[11] The positive correlation can be also seen from (21), where q_l obviously increases over l.

For RUFL, we assume that the probability for all n nearest facilities to a customer at $x \in \mathscr{S}$ to fail is given by $B_{n,a(x),b(x)}(n)$ with varying parameters $a(x)$, $b(x)$. Through interpolation (similar to (15)), $\overline{P}(x)$ can be represented as follows:

$$\overline{P}(x) \approx \begin{cases} (1-\bar{\theta})B_{\lfloor\theta(x)\rfloor,a(x),b(x)}(\lfloor\theta(x)\rfloor) + \bar{\theta}B_{\lfloor\theta(x)\rfloor+1,a(x),b(x)}(\lfloor\theta(x)\rfloor+1), & \theta(x) > 1, \\ 1-[1-B_{1,a(x),b(x)}(1)]\theta(x), & \text{otherwise}. \end{cases} \tag{22}$$

We also assume that the probability for all n nearest facility to fail but the $(n+1)^{\text{th}}$ facility to survive is given by $B_{n+1,a(x),b(x)}(n)$, and the service probability $P_r(x)$ can be approximated by

$$P_r(x) \approx B_{r+1,a(x),b(x)}(r). \tag{23}$$

Again, if the facilities fail independently, probabilities $\overline{P}(x)$ and $P_r(x)$ could be obtained from (6) and (8) respectively, with probability $q(x) = B_{1,a(x),b(x)}(1) = \frac{a(x)}{a(x)+b(x)}$.

4.2. Correlation induced from shared hazard exposure

Sometimes the sources and causal mechanisms of facility disruptions are well understood. In such cases, the disruption probabilities can be conditioned on a set of mutually exclusive hazard occurrence states, $\mathscr{H}$. Each state $h \in \mathscr{H}$ corresponds to a possible scenario of hazard occurrence (e.g., earthquake, hurricane). Suppose state h happens with a probability Q_h ($\sum_{h\in\mathscr{H}} Q_h = 1$ if "no-disaster" is considered one of the states). Conditional on each state h, each facility near x fails independently with probability $\chi_h(x)$. It should be noted that although the facilities fail independently within each hazard occurrence state, the overall facility disruptions (due to all hazards) can be correlated.

In each state h, the penalty probability for a customer at x can be approximated by (6). Based on conditional expectation, the overall penalty probability $\overline{P}(x)$ across all possible hazard occurrence state is

$$\overline{P}(x) \approx \begin{cases} \sum_h Q_h[\chi_h(x)]^{\theta(x)}, & \theta(x) > 1, \\ \sum_h Q_h\{1-[1-\chi_h(x)]\theta(x)\}, & \text{otherwise}. \end{cases} \tag{24}$$

Similarly, given h, the service probability at rank r for a customer at x can be approximated from (8). The expected value across all states yields $P_r(x)$ as follows:

$$P_r(x) \approx \sum_h Q_h[1-\chi_h(x)]\chi_h^r(x). \tag{25}$$

Note that $\overline{P}(x)$ and $P_r(x)$ can be expressed equivalently in the form of (15) and (16), respectively, by setting

$$q_l = \frac{\sum_h Q_h \chi_h(x)^{l+1}}{\sum_h Q_h \chi_h(x)^l}, \quad \forall l.$$

Now we briefly discuss how the penalty and service probabilities will be erroneous if correlations are ignored. Note that the single facility failure probability $q(x) = \sum_h Q_h \chi_h(x)$. $\overline{P}(x)$ and $P_r(x)$ could be calculated from (6) and (8) as follows:

$$\overline{P}(x) \approx \begin{cases} \left[\sum_h Q_h\chi_h(x)\right]^{\theta(x)}, & \theta(x) > 1, \\ 1-[1-\sum_h Q_h\chi_h(x)]\theta(x), & \text{otherwise}, \end{cases} \qquad P_r(x) \approx \left[1-\sum_h Q_h\chi_h(x)\right]\left[\sum_h Q_h\chi_h(x)\right]^r.$$

Note that much of the difference in the corresponding probability formulas (with or without correlations) comes from the fact that

$$\mu_r(x) := \sum_h Q_h\chi_h^r(x) - \left[\sum_h Q_h\chi_h(x)\right]^r \geqslant 0, \quad \forall r > 2, \tag{26}$$

due to the Jensen's Inequality. Note that $\mu_r(x)$ becomes even larger as r increases.

[11] A larger value of $\frac{1}{a+b}$ corresponds to a greater variance and hence more significant correlation (Bakkaloglu et al., 2002). For example, when $\frac{1}{a+b} \to 0$, facilities fail almost independently; when $\frac{1}{a+b} \to \infty$, facility failure is almost perfectly correlated.

5. Numerical examples

This section presents four numerical examples to illustrate how the CA model can be applied to problems with correlated facility disruptions. Each example uses an aforementioned failure correlation structure. The space $\mathscr{S}$ is a $[0,1]\times[0,1]$ unit square. Customer demand is distributed with density function $\lambda(x)=\bar{\lambda}[1+\tau_\lambda\cos(\omega\|x\|)]$, and the facility opening cost at x is $f(x)=\bar{f}[1+\tau_f\cos(\omega\|x\|)]$, where $\tau_\lambda\in[-1,1]$ and $\tau_f\in[-1,1]$ control the heterogeneity of $\lambda(x)$ and $f(x)$ over $\mathscr{S}$, respectively. The scalar ω is selected to normalize the average customer density and the average facility cost (e.g., $\int_{\mathscr{S}}\lambda(x)dx=\bar{\lambda}$ and $\int_{\mathscr{S}}f(x)dx=\bar{f}$). The travel cost factor $\alpha_t=1$.

The estimated optimal total cost C^* and the estimated optimal facility number N^* are computed from (17) and (19) respectively. For comparison, we let $A_I(x)$, C_I and N_I respectively denote the optimal $A(x)$, total cost, and the number of facility when correlation is erroneously ignored. These three values may be relevant to strategic resource allocation and budget planning. Let C_{IC} denote the actual cost under correlation while solution $A_I(x)$ is implemented. The percentage difference $\varepsilon_I=\frac{C_I-C^*}{C^*}$ indicates the error in estimated system cost caused by ignoring correlations, while $\varepsilon_{IC}=\frac{C_{IC}-C^*}{C^*}$ indicates the actual cost difference after implementing the "wrong" design.

5.1. Correlation specified by conditional probabilities

Following the framework presented in Section 3.2, we set the conditional probabilities to be

$$q_1(x)=q_0(x)+\Delta q(x),$$
$$q_l(x)=\min\left\{q_{l-1}(x)+\frac{q_{l-1}(x)-q_{l-2}(x)}{2},\frac{q_{l-1}(x)+1}{2}\right\},\quad\forall l=2,3,\ldots,\tag{27}$$

Here, positive/negative $\Delta q(x)$ yields positive/negative correlations; e.g., perfect correlation can be specified by setting $\Delta q(x)=1-q_0(x)$. For demonstration purposes, we simply assume $q_0(x)=q_0,\ \Delta q(x)=\Delta q,\ \forall x$.

Substituting Eq. (27) into (15) and (16) yields the correct penalty and service probabilities, while the erroneous counterparts can be computed from (6) to (8). Table 1 illustrates the results for a range of problem instances with $\bar{f}=1,\ \bar{\lambda}=500,\ \omega=11.73,\tau_\lambda,\ \tau_f\in\{0,1\},\ q_0\in\{0.05,0.2\},\Delta q\in\{-q_0/2,(1-q_0)/2,1-q_0\},\alpha_p\in\{1,10\}$, and $D\in\{0.1,0.2\}$.

It can be observed that $N^*,\ N_I,\ C^*,\ C_I$ and C_{IC} all increase with q_0 in almost all cases, indicating that facilities should be deployed closer to each other (as back-ups) under higher failure probabilities, and as a result the total system cost increases. The same trend is observed as α_p increases; this is intuitive because higher α_p implies higher penalty cost, which would motivate a denser facility deployment. As D increases (i.e., reducing the likelihood for customers to incur penalty), the optimal numbers of facilities, N^* and N_I, both decrease; however, the values of $C^*,\ C_I$ and C_{IC} may still increase because a larger D also implies a proportionally larger penalty value.

The optimal total cost C^* is obviously influenced by the correlation, sometimes dramatically (when q_0 and α_p are large); positive correlation generally leads to higher total cost. On the contrary, the optimal number of facilities N^* decreases under positive correlation in most of the cases, probably because positively correlated failures weaken the benefit of having more facilities as backups.

The error $|\varepsilon_I|$ always increases with $|\Delta q|$, which means that assuming independent disruptions yields a poor cost estimation when correlations are actually present. The error is large in cases of high failure probability q_0 and large penalty factor α_p. As expected, all ε_I values are negative for positive Δq (leading to underestimation of disruption risks) and nonnegative for negative Δq. This is consistent with the discussions in Section 3. On the other hand, the actual cost error ε_{IC} is always nonnegative. This is not surprising because $A_I(x)$ is suboptimal to the cost-minimization problem. For most of cases, $|\varepsilon_{IC}|$ is not large. This is probably because the objective function is quite flat near the optimal solution (similar to many other facility location problems). Nevertheless, $|\varepsilon_{IC}|$ is large for large α_p and $\theta^*\approx 1$, as the solutions under these scenarios impose a large penalty risk to the customers.

When $\lambda(x)$ is heterogeneous (i.e., $\tau_\lambda=1$), N^* and C^* are lower than those in the corresponding homogeneous cases (i.e., $\tau_\lambda=0$). This suggests that uneven distribution of customers generally reduces the optimal total cost. In addition, heterogeneous $\lambda(x)$ seems to slightly inflate $|\epsilon_I|$ under positive correlations and reduce it under negative correlations. However, when the facility opening cost $f(x)$ varies in proportion to $\lambda(x)$ (i.e., $\tau_\lambda=\tau_f=1$), which may happen due to higher land prices in areas with high population density, the results are almost the same as those with homogeneous $\lambda(x)$ and $f(x)$ (i.e., $\tau_\lambda=\tau_f=0$).

5.2. Correlations specified by the beta-binomial distribution

We assume that all parameters remain the same as those in the previous example, excepted that the correlations are expressed via beta-binomial distribution with parameters $a(x)=a,\ b(x)=b,\ \forall x$. Eqs. (22) and (23) are used to estimate the penalty and service probabilities $\overline{P}(x)$ and $P_r(x)$.

Table 2 shows the results for a range of problem instances, where $\bar{f}=1,\ \bar{\lambda}=500,\ \omega=11.73,\tau_\lambda,\ \tau_f\in\{0,1\},\ a\in\{0.1,0.01\}$, $b\in\{19a,4a\},\ \alpha_p\in\{1,10\}$, and $D\in\{0.1,0.2\}$. Since the beta-binomial formulation is simply a special case of the general conditional probability formulation, the results are consistent with those in Section 5.1. Facility number N^* and optimal cost C^* generally

Table 1
CA cost estimation when correlation is specified by conditional probabilities ($\theta^* = \pi D^2/A^*$).

#	τ_λ	τ_f	q_0	Δq	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I (%)	ε_{IC} (%)
1	0	0	0.05	−0.025	1	0.2	2.7	21	21	64	64	64	0	0
2	0	0	0.05	−0.025	10	0.2	3	24	22	64	64	64	0	0
3	0	0	0.05	−0.025	10	0.1	1.4	44	44	88	80	88	−9	0
4	0	0	0.05	0.475	1	0.2	2.6	21	21	65	64	65	−1	0
5	0	0	0.05	0.475	10	0.2	3.2	26	22	83	64	84	−22	1
6	0	0	0.05	0.475	10	0.1	1	32	44	89	80	92	−10	4
7	0	0	0.05	0.95	1	0.2	2.5	20	21	65	64	65	−2	0
8	0	0	0.05	0.95	10	0.2	2.5	20	22	110	64	110	−42	0
9	0	0	0.05	0.95	10	0.1	1	32	44	89	80	96	−10	8
10	0	0	0.2	−0.1	1	0.2	3	24	21	70	70	70	0	0
11	0	0	0.2	−0.1	10	0.2	3.2	25	31	71	74	72	4	1
12	0	0	0.2	−0.1	10	0.1	2	64	67	101	110	103	9	2
13	0	0	0.2	0.4	1	0.2	2.6	21	21	72	70	72	−3	0
14	0	0	0.2	0.4	10	0.2	4.2	33	31	153	74	154	−52	1
15	0	0	0.2	0.4	10	0.1	2	64	67	146	110	148	−25	1
16	0	0	0.2	0.8	1	0.2	2.2	18	21	74	70	74	−5	1
17	0	0	0.2	0.8	10	0.2	2.2	18	31	254	74	258	−71	2
18	0	0	0.2	0.8	10	0.1	1	32	67	159	110	185	−31	17
19	1	0	0.2	−0.1	1	0.2	2.5	20	20	64	64	64	0	0
20	1	0	0.2	−0.1	10	0.2	3.2	25	28	67	69	68	3	1
21	1	0	0.2	−0.1	10	0.1	1.6	51	56	92	98	95	6	3
22	1	0	0.2	0.4	1	0.2	2.3	18	20	66	64	67	−4	0
23	1	0	0.2	0.4	10	0.2	3.9	31	28	147	69	147	−53	1
24	1	0	0.2	0.4	10	0.1	1.5	47	56	135	98	137	−28	1
25	1	0	0.2	0.8	1	0.2	2	16	20	69	64	70	−8	1
26	1	0	0.2	0.8	10	0.2	2.2	17	28	250	69	254	−72	2
27	1	0	0.2	0.8	10	0.1	0.9	27	56	156	98	175	−37	13
28	1	1	0.2	−0.1	1	0.2	3	24	21	70	70	70	0	0
29	1	1	0.2	−0.1	10	0.2	3.2	25	31	71	73	72	4	1
30	1	1	0.2	−0.1	10	0.1	2	64	67	101	110	102	9	2
31	1	1	0.2	0.4	1	0.2	2.6	21	21	72	70	72	−3	0
32	1	1	0.2	0.4	10	0.2	4.2	33	31	153	73	154	−52	1
33	1	1	0.2	0.4	10	0.1	2	64	67	146	110	148	−25	1
34	1	1	0.2	0.8	1	0.2	2.2	18	21	74	70	74	−5	1
35	1	1	0.2	0.8	10	0.2	2.2	18	31	254	73	258	−71	2
36	1	1	0.2	0.8	10	0.1	1	32	67	158	110	185	−31	17

increase over the failure probability $\frac{a}{a+b}$, the correlation $\frac{1}{a+b}$, and the penalty factor α_p. The estimation error $|\epsilon_I|$ is large, especially when the penalty cost and the correlation are high, though $|\epsilon_{IC}|$ is only large for a few cases. Heterogeneities in the system again help reduce the optimal number of facilities and the total cost.

5.3. Flooding hazard

Now we suppose that facility failure may be caused by a potential flooding hazard, and the flood, whenever happening, always immerses the whole $\mathscr{S}$.[12] Following the framework introduced in Section 4.2, there are $|\mathscr{H}| = 2$ exclusive hazard occurrence states; assume that state $h = 1$ represents no-disaster, which occurs with a high probability $Q_1 = 0.9$, while state $h = 2$ represents flooding disaster, which occurs with a low probability $Q_2 = 0.1$.[13] For $h = 1, 2$, the associated facility failure probability $\chi_h(x) = \chi_h$ for all $x \in \mathscr{S}$, where $\chi_1 \ll 1$ and $\chi_2 > 0$. Penalty and service probabilities $P_r(x)$ and $\overline{P}(x)$ are computed from (24) and (25) respectively.

Now that we have two hazard occurrence states, we use ε_{IC1} and ε_{IC2} to replace ε_{IC}, representing the actual total cost error under states 1 and 2, respectively. Table 3 shows the results for a range of instances where $\bar{f} = 1,\ \bar{\lambda} = 500,\ \omega = 11.73$, $\tau_\lambda,\ \tau_f \in \{0, 1\}$, $[\chi_1, \chi_2] \in \{[0, 0.5], [0, 1]\}$, $\alpha_p \in \{1, 10\}$, and $D \in \{0.1, 0.2\}$. Recall that $\mu_2 = \sum_h Q_h \chi_h^2(x) - [\sum_h Q_h \chi_h(x)]^2$ indicates the magnitude of positive correlations. We can observe that the impacts of failure probability, correlation, penalty and parameter heterogeneities on the optimal number of facilities and the total cost are similar to those seen in the previous numerical experiments.

[12] For problems where only some subareas are subject to such hazards, we can partition $\mathscr{S}$ accordingly and solve a subproblem for each subarea.
[13] This is for illustration only; in the real world Q_2 should be much smaller.

Table 2
CA cost estimations when correlation is specified by the beta-binomial distribution.

#	τ_λ	τ_f	$\frac{a}{a+b}$	$\frac{1}{a+b}$	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I (%)	ε_{IC} (%)
1	0	0	0.05	0.5	1	0.2	2.4	19	21	68	64	68	−6	0
2	0	0	0.05	0.5	10	0.2	3.6	28	22	77	64	78	−16	2
3	0	0	0.05	0.5	10	0.1	1	32	43	89	80	91	−10	3
4	0	0	0.05	5	1	0.2	2.5	20	21	66	64	66	−10	2
5	0	0	0.05	5	10	0.2	3	24	22	101	64	101	−60	2
6	0	0	0.05	5	10	0.1	1	32	43	89	80	95	−31	3
7	0	0	0.2	2	1	0.2	1.9	15	21	78	70	80	−11	0
8	0	0	0.2	2	10	0.2	5	40	30	183	74	186	−60	0
9	0	0	0.2	2	10	0.1	1.1	36	66	159	110	163	−31	7
10	0	0	0.2	20	1	0.2	2.2	17	21	74	70	75	−6	1
11	0	0	0.2	20	10	0.2	3	24	30	246	74	247	−70	1
12	0	0	0.2	20	10	0.1	1	32	66	159	110	182	−31	15
13	1	0	0.2	2	1	0.2	1.6	12	19	74	64	77	−14	4
14	1	0	0.2	2	10	0.2	4.6	36	28	175	69	179	−61	2
15	1	0	0.2	2	10	0.1	1.3	42	55	147	98	151	−34	3
16	1	0	0.2	20	1	0.2	1.9	14	19	70	64	71	−9	2
17	1	0	0.2	20	10	0.2	2.7	21	28	241	69	242	−71	1
18	1	0	0.2	20	10	0.1	0.9	27	55	156	98	172	−37	10
19	1	1	0.2	2	1	0.2	1.9	15	21	78	70	80	−10	2
20	1	1	0.2	2	10	0.2	5	39	30	183	73	186	−60	2
21	1	1	0.2	2	10	0.1	1.1	35	66	158	110	163	−31	3
22	1	1	0.2	20	1	0.2	2.2	17	21	74	70	75	−6	1
23	1	1	0.2	20	10	0.2	3	23	30	245	73	247	−70	1
24	1	1	0.2	20	10	0.1	1	31	66	158	110	182	−31	15

Table 3
CA cost estimations for flooding hazard.

#	τ_λ	τ_f	$\sum_h Q_h \chi_h$	μ_2	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I (%)	ε_{IC} (%)	ε_{IC1} (%)	ε_{IC2} (%)
1	0	0	0.05	0.02	1	0.2	2.6	21	21	64	64	64	−1	0	0	0
2	0	0	0.05	0.02	10	0.2	4.1	33	22	73	64	78	−12	6	−5	54
3	0	0	0.05	0.02	10	0.1	1	32	44	89	80	91	−10	2	11	−14
4	0	0	0.1	0.09	1	0.2	2.4	19	21	68	66	68	−3	0	0	2
5	0	0	0.1	0.09	10	0.2	2.4	19	25	158	67	159	−58	1	1	1
6	0	0	0.1	0.09	10	0.1	1	32	51	112	90	125	−19	12	19	4
7	1	0	0.05	0.02	1	0.2	2.3	19	19	59	58	59	−1	0	0	0
8	1	0	0.05	0.02	10	0.2	3.6	29	22	67	60	69	−9	4	−3	36
9	1	0	0.05	0.02	10	0.1	1.1	36	38	83	74	84	−12	1	2	−3
10	1	0	0.1	0.09	1	0.2	2.1	17	19	63	60	63	−4	0	0	1
11	1	0	0.1	0.09	10	0.2	2.3	19	24	154	63	155	−59	1	1	0%
12	1	0	0.1	0.09	10	0.1	0.9	28	44	109	82	119	−25	9	14	3
13	1	1	0.05	0.02	1	0.2	2.6	21	21	65	64	65	−1	0	0	0
14	1	1	0.05	0.02	10	0.2	4.1	33	22	73	64	78	−12	6	−5	54
15	1	1	0.05	0.02	10	0.1	1	32	44	89	80	91	−10	2	11	−14
16	1	1	0.1	0.09	1	0.2	2.4	19	21	68	66	68	−3	0	0	2
17	1	1	0.1	0.09	10	0.2	2.4	19	25	158	67	159	−58	1	1	1
18	1	1	0.1	0.09	10	0.1	1	32	51	112	90	125	−19	12	19	4

5.4. Earthquake hazard

This section considers a heterogeneous case where earthquake hazards impose site-dependent failure probability over $\mathscr{S}$. The setting is the same as that in Section 5.3 except that hazard occurrence state $h = 2$ is induced by an earthquake source centered at $(0, 0)$, and ω is set to be 2.038 to ensure that $\lambda(x)$ is monotone (either always decreases or always increases) as we move away from the earthquake center. When an earthquake occurs, a facility at $x \in \mathscr{S}$ fails with a probability $q(x) = \exp(-\beta\|x\|)$, where β is a scalar.

Table 4 shows the results for a range of instances. The average values of $\sum_h Q_h \chi_h(x)$ and $\mu_2(x)$ are set to be comparable to those in the flooding examples. As expected, we observe consistent effects of failure probability, correlation, penalty and het-

Table 4
CA cost estimations for earthquake hazard ($\bar{q} = \int_{x\in\mathscr{S}} \sum_h Q_h \chi_h(x)dx$ and $\bar{\mu}_2 = \int_{x\in\mathscr{S}} \mu_2(x)dx$).

#	τ_λ	β	$\bar{q}$	$\bar{\mu}_2$	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I (%)	ε_{IC} (%)	ε_{IC1} (%)	ε_{IC2} (%)
1	0	1	0.05	0.02	1	0.2	2.6	21	21	64	64	64	−1	0	0	0
2	0	1	0.05	0.02	10	0.2	4.1	32	22	75	64	79	−14	6	−5	47
3	0	1	0.05	0.02	10	0.1	1	32	43	88	80	90	−9	3	10	−12
4	0	0.05	0.1	0.08	1	0.2	2.5	20	21	68	66	68	−3	0	0	1
5	0	0.05	0.1	0.08	10	0.2	3.4	27	25	148	67	149	−55	0 %	−1	1
6	0	0.05	0.1	0.08	10	0.1	1	32	51	110	89	122	−19	11	19	2
7	1	1	0.05	0.02	1	0.2	2.5	20	20	63	62	63	−1	0	0	0
8	1	1	0.05	0.02	10	0.2	4.2	33	22	74	63	80	−16	8	−7	60
9	1	1	0.05	0.02	10	0.1	1.1	36	43	91	78	92	−14	1	6	88
10	1	0.05	0.1	0.08	1	0.2	2.3	19	20	66	64	66	83	0	0	1
11	1	0.05	0.1	0.08	10	0.2	3.2	25	25	148	65	149	−56	0	0	0
12	1	0.05	0.1	0.08	10	0.1	1	32	49	111	87	120	−21	9	15	2
13	−1	1	0.05	0.02	1	0.2	2.4	19	19	61	61	61	0	0	0	0
14	−1	1	0.05	0.02	10	0.2	3.5	28	22	69	62	71	−9	3	−3	27
15	−1	1	0.05	0.02	10	0.1	1.1	36	41	83	76	84	−8	1	4	−7
16	−1	0.05	0.1	0.08	1	0.2	2.3	18	19	65	63	65	−3	0	0	1
17	−1	0.05	0.1	0.08	10	0.2	4	32	24	140	65	143	−54	2	−6	7
18	−1	0.05	0.1	0.08	10	0.1	1	31	48	109	86	118	−21	8	15	1

erogeneities. Particularly, the spatial distribution patterns of customer demand $\lambda(x)$ and facility failure probability $q(x)$ seem to jointly influence the optimal system design. For example, when customer density increases with the distance from the earthquake center (i.e., $\tau_\lambda = -1$), the optimal cost C^* drops. This desirable situation is probably due to not only the demand heterogeneity but also the concentration of demand in places with lower facility failure risks. On the contrary, when customer density decreases with the distance from the earthquake center (i.e., $\tau_\lambda = 1$), the change of C^* is not always monotone. Although the heterogeneity of $\lambda(x)$ tends to reduce the total cost, the fact that more customers live in places with higher facility failure risk tends to increase the total system cost.

6. Conclusions

This paper presents a continuum approximation approach to analyze the reliable uncapacitated facility location problem. Facilities are subject to spatially correlated disruptions that occur with location-dependent probabilities, and if a facility fails, its customers are diverted to other facilities and incur excessive transportation cost. The proposed model builds on the properties of idealized special cases and solves the problem for general heterogeneous and correlated disruptions. Particularly, spatial correlations have been formulated in a variety of ways to provide modeling flexibility. The model is tested with a series of numerical experiments where the effects of disruption correlations on the location design and cost estimation are discussed.

Future research can be conducted in a few directions. In this paper all facilities are assumed to have infinite capacity and inventory cost is not considered in the model. This should be addressed in future studies. The CA model will work well in large-scale space with slow-varying conditions. When extremely heterogeneous conditions are present, a discrete model and efficient solution algorithms will be useful to help evaluate the accuracy and robustness of the CA model. Furthermore, all case studies in this paper are based on artificially generated test cases. Future research is needed to develop effective procedures to map discrete data into the continuous setting so that the proposed modeling framework can be used to solve existing and realistic problems in the discrete setting.

Acknowledgments

This research was partially supported by the National Science Foundation through Grants CMMI-0748067, EFRI-RESIN-0835982, and by the NEXTRANS Center (USDOT Region V Regional University Transportation Center) through Project 012IY01. The valuable comments from the two anonymous reviewers are gratefully acknowledged.

Appendix A. Proof for Proposition 2

Assume first that $\mathscr{S}$ is bounded but sufficiently large so that the boundary effect can be ignored. Each customer has one and only one facility as its r^{th} choice. This implies that the service areas of all different facilities with the same service rank form a mutually exclusive partition of $\mathscr{S}$, i.e.,

$$\bigcup_j \mathscr{A}_{j,r} = \mathscr{S} \quad \text{and} \quad \mathscr{A}_{i,r} \bigcap \mathscr{A}_{j,r} = \emptyset, \ \forall i \neq j, r. \tag{28}$$

Since almost every facility in $\mathscr{S}$ is translationally symmetric, $|\mathscr{A}_{j,r}| = |\mathscr{A}_{i,r}|$, for almost all i, j (except those near the boundary), Eq. (28) implies that $|\mathscr{A}_{j,r}| \approx |\mathscr{S}|/N$ for all service rank r and facility j, where N is the total number of facilities. When $\mathscr{S} \to \mathbb{R}^2$, the boundary effect can be totally eliminated. Thus $|\mathscr{A}_{j,r}| = |\mathscr{A}_{j,0}| = A, \ \forall j, r$. This completes the proof. □

Appendix B. Computing exact values of γ_r, P_r, C_t and $\overline{P}$ for the IHI problem

We deploy facilities such that the initial service areas form a regular hexagonal partition (each with hexagon size A) on a sufficiently large area $\mathscr{S}$ (e.g., with $|\mathscr{S}| > 100A$) centered at $(0, 0)$. Then $\mathscr{S}$ is diced into infinitesimal squares (e.g., with size <0.001A), each representing a customer neighborhood. To eliminate the influence from the boundary of $\mathscr{S}$, only those squares sufficiently far away from the boundary are considered. For any given values of q and θ, we conduct the following computations to obtain exact values of $\gamma_r, \ P_r, \ C_t$ and $\overline{P}$.

Without losing generality, we focus on facility j which is located at $x_j = (0, 0)$ and determine the service area partition $\{\mathscr{A}_{j,r}, \ \forall r\}$ as shown in Fig. 1. For any customer neighborhood at $x \in \mathscr{A}_{j,r}$, the travel distance to facility j is $\|x\|$ and the corresponding service probability is $(1-q)q^r$. For each r, the exact values of $\gamma_r, \ P_r$ are calculated by averaging $\|x\|/A^{1/2}$ and $(1-q)q^r$ respectively across all the corresponding infinitesimal squares in $\mathscr{A}_{j,r}$.

The expected total transportation cost to facility $j, C_{t,j}$, is the summation of $\|x\|(1-q)q^r$ across all customer neighborhoods that satisfy $\|x\| \leqslant D$. Due to symmetry, the value of $C_{t,j}$ is identical for all j, and hence the transportation cost per unit area is $C_t = \frac{C_{t,j}}{A}$. For every customer neighborhood at $x \in \mathscr{A}_{j,0}$, also count $N_D(x)$, the number of facilities that are within distance D. Penalty probability $\overline{P}(x) = q^{N_D(x)}$, and $\overline{P}$ is computed as the average value of $\overline{P}(x)$ across x.

Appendix C. Proof for Proposition 3

Any customer at $x \in \mathscr{S}$ may travel a distance $\delta \in [0, D(x))$ to receive service. Define $c(\delta), \ \forall \delta \leqslant D(x)$ to be the cost for one unit of demand at x; i.e.,

$$c(\delta) = \begin{cases} \alpha_t \delta, & \delta < D(x), \\ \alpha_p \delta, & \delta = D(x). \end{cases}$$

Obviously, $c(\delta)$ is an increasing function of δ since $\alpha_p \geqslant \alpha_t$. Under correlated facility failure, let $F(\delta)$ denote the probability for the customer to travel farther than distance δ. The expected cost for one unit of demand at x is

$$E[c(\delta)] = \int_{\delta=0}^{D(x)} c(\delta) d[1 - F(\delta)] + c(D(x))F(D(x)) = \int_{\delta=0}^{D(x)} F(\delta) dc(\delta) + c(D(x))F(D(x)). \tag{29}$$

If facility failure is independent, the probability for the customer to travel farther than distance δ is denoted by $F_I(\delta)$. The expected cost becomes

$$E_I[c(\delta)] = \int_{\delta=0}^{D(x)} c(\delta) d[1 - F_I(\delta)] + c(D(x))F_I(D(x)) = \int_{\delta=0}^{D(x)} F_I(\delta) dc(\delta) + c(D(x))F_I(D(x)). \tag{30}$$

Note that $F(\delta)$ and $F_I(\delta)$ are the probabilities for all facility within distance δ from x to fail. By definition, for any $\delta, \ F(\delta) \geqslant F_I(\delta)$ under positive correlations, or $F(\delta) \leqslant F_I(\delta)$ otherwise. Hence, comparison between (29) and (30) clearly shows that $E[c(\delta)] \geqslant E_I[c(\delta)]$ when the correlation is positive; the contrary is also true. This completes the proof. □

References

Bakkaloglu, M., Wylie, J.J., Wang, C., Ganger, G.R., 2002. On correlated failures in survivable storage systems. Technical Report CMU-CS-02-129, Carnegie Mellon University.

Ball, M., Lin, F., 1993. A reliability model applied to emergency service vehicle location. Operations Research 41 (1), 18–36.

Batta, R., Dolan, J., Krishnamurthy, N., 1989. The maximal expected covering location problem – revisited. Transportation Science 23 (4), 277–287.

Campbell, J.F., 1993a. Continuous and discrete demand hub location problems. Transportation Research Part B 27 (6), 473–482.

Campbell, J.F., 1993b. One-to-many distribution with transshipments: an analytic model. Transportation Science 27 (4), 330–340.

Cui, T., Ouyang, Y., Shen, Z.M., 2009. Reliable facility location under the risk of disruptions. Operations Research, in press.

Daganzo, C., 1984a. The distance traveled to visit N points with a maximum of C stops per point: a manual tour-building strategy and case study. Transportation Science 18 (4), 331–350.

Daganzo, C., 1984b. The length of tours in zones of different shapes. Transportation Research Part B 18 (2), 135–146.

Daganzo, C., 1999. On planning and design of logistics systems for uncertain environments. In: New Trends in Distribution Logistic. Springer, Berlin, Germany, pp. 3–21.

Daganzo, C., 2005. Logistics System Analysis, fourth ed. Springer, Berlin, Germany.

Daganzo, C., Newell, G., 1986. Configuration of physical distribution networks. Networks 16 (2), 113–132.

Dasci, A., Verter, V., 2001. A continuous model for production–distribution system design. European Journal of Operational Research 129 (2), 287–298.

Daskin, M., 1982. Application of an expected covering model to emergency medical service system design. Decision Science 13 (3), 416–439.

Daskin, M., 1983. A maximum expected covering location model: formulation, properties and heuristic solution. Transportation Science 17 (1), 48–70.

Daskin, M., 1995. Network and Discrete Location: Models, Algorithms, and Applications. John Wiley, New York.

Drezner, Z. (Ed.), 1995. Facility Location: A Survey of Applications and Methods. Springer, New York.

Goyal, A., Nicola, V.F., 1990. Modeling of correlated failures and community error recovery in multi-version software. IEEE Transaction on Software Engineering 16 (3), 350–359.
Griffiths, D.A., 1973. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29 (4), 637–648.
Hall, R., 1984. Travel distance through transportation terminals on a rectangular grid. Journal of the Operational Research Society 35 (12), 1067–1078.
Hall, R., 1986. Discrete models/continuous models. OMEGA – International Journal of Management Science 14 (3), 213–220.
Hall, R., 1989. Configuration of an overnight package air network. Transportation Research Part A 23 (2), 139–149.
Langevin, A., Mbaraga, P., Campbell, J., 1996. Continuous approximation models in freight distribution: an overview. Transportation Research Part B 30 (3), 163–188.
Li, X., Ouyang, Y., 2009. Reliable Sensor Deployment for Network Traffic Surveillance. Working Paper, University of Illinois at Urbana-Champaign, submitted for publication.
Newell, G., 1971. Dispatching policies for a transportation route. Transportation Science 5 (1), 91–105.
Newell, G., 1973. Scheduling, location, transportation and continuum mechanics: some simple approximations to optimization problems. SIAM Journal on Applied Mathematics 25 (3), 346–360.
Ouyang, Y., 2007. Design of vehicle routing zones for large-scale distribution systems. Transportation Research Part B 41 (10), 1079–1093.
Ouyang, Y., Daganzo, C., 2006. Discretization and validation of the continuum approximation scheme for terminal system design. Transportation Science 40 (1), 89–98.
Ouyang, Y., Li, X., Barkan, C., Kawprasert, A., Lai, Y.C., 2009. Optimal locations of railroad wayside defect detection installations. Computer-aided Civil and Infrastructure Engineering 24 (5), 309–319.
Revelle, C., Hogan, K., 1989. The maximum availability location problem. Transportation Science 23 (3), 192–200.
Snyder, L., Daskin, M., 2005. Reliability models for facility location: the expected failure cost case. Transportation Science 39 (3), 400–416.
Tang, D., Iyer, R.K., 1992. Analysis and modeling of correlated failures in multicomputer systems. IEEE Transactions on Computers 41 (5), 567–577.
Toth, L.F., 1959. Sur la representation dune population infinie par un nombre fini delements. Acta Mathematica Hungarica 10 (3–4), 299–304.