Submitted to *Transportation Science*
manuscript (Please, provide the mansucript number!)

A Continuum Approximation Approach to the Dynamic Facility Location Problem in a Growing Market

Xin Wang
Department of Civil and Environmental Engineering, wangxin1@illinois.edu,

Michael K. Lim
Department of Business Administration, mlim@illinois.edu,

Yanfeng Ouyang
Department of Civil and Environmental Engineering, yfouyang@illinois.edu,

This paper proposes a continuum approximation model framework to solve a dynamic facility location problem for a large-scale growing market. The objective is to determine the optimal facility location and deployment time that minimize the costs for facility construction and customer service. To overcome computational challenge, a continuous approximation model is developed to find the optimal facility density in the spatiotemporal continuum. Then we propose a tube model to discretize the resulted continuous facility density function into a set of time-varying facility location trajectories. To enforce consistency of facility location over time, an iterative regulation procedure based on a penalty method is applied. We present convergence properties of the proposed iterative regulation procedure and further derive conditions under which the proposed continuous approximation approach and the tube model provide tight approximation error bounds. A series of numerical experiments, including comparison with discrete model counterparts, are conducted to illustrate the performance (accuracy and convergence) of the proposed modeling framework. Our results show that the proposed method solves the dynamic facility location problem effectively to reasonable accuracy.

1. Introduction

Facility location decisions are critical to many problems in business practice, where a decision maker (e.g., a government or a firm) needs to find spatial distribution of a set of facilities (e.g., fire stations or local branches) in order to minimize the total cost (e.g. facility and transportation cost) for satisfying various demands. Facility location decisions typically imply long-term commitment of significant resources. Once the decisions are made (facilities are opened), they are very difficult to reverse. Despite the long term impacts associated with facility location decisions, due to computational challenges, a great deal of research focuses on *static* facility location problems, where parameters are constant over time. In reality, however, any parameters of the system (e.g., customer demand, operating cost) may vary over time and facility additions can occur at different times. That is, not only *where* but also *when* to build a facility becomes a critical decision.

In this paper, we focus on a *dynamic facility location problem*, in which a central planner aims to determine the best location and timing of facility deployment that minimizes the cumulative logistics cost during a planning horizon. In particular, we consider a growing market where demand increases (whereas the unit operating cost may decrease) such that new facilities are added over time. The goal of this paper is to provide an effective solution method that is suitable for large-scale instances of such dynamic facility location problems.

Suppose a set of facilities, indexed by $i = 1, 2, \cdots$, are opened in a continuous compact service area[1] $\Omega \subseteq \mathbb{R}^2$ over a finite planning horizon $\Psi = [0, T]$. Facility i will be opened at location $x_i \in \Omega$ at time $\tau_i \in \Psi$, covering a set of customers in its service region $S_i(t) \subseteq \Omega$ for all $t \in \Psi$ (or $S_i(t) = \emptyset$ for all $t < \tau_i$). Each customer is assumed to be served by a facility, hence $\{S_i(t)\}_{\forall i}$ forms a partition of Ω at any time t. We consider that the total system cost consists of three parts: one-time fixed facility investments, operating costs over time, and transportation costs over time. Suppose near location x at time t, the fixed investment is $h(x,t)$, and the prorated customer service cost including facility operating cost and transportation cost per unit area per unit time is an integrable function over space and time, $z(x, x_i, S_i(t), t)$, where $x \in S_i(t)$ and facility i is the corresponding serving facility. Considering the time value of money, both cost functions are converted into the present value. In addition, we assume opened facilities cannot be closed or relocated due to high penalties associated with such actions (Current et al. 1997). For convenience, we define $\mathbf{S}_i := \{S_i(t)\}_{\forall t}$. Then, the system design decisions with respect to location, service region, and timing are represented by $\{x_i\}_{\forall i}$, $\{\mathbf{S}_i\}_{\forall i}$, and $\{\tau_i\}_{\forall i}$, respectively. Therefore, we can formulate a generic form of a dynamic facility location problem as follows.

Dynamic Facility Location Problem (DFLP):

$$\min_{\{x_i\}_{\forall i}, \{\mathbf{S}_i\}_{\forall i}, \{\tau_i\}_{\forall i}} P = \int_{\Psi} \int_{\Omega} z(x, x_i, S_i(t), t)\, dx dt + \sum_{\forall i} h(x_i, \tau_i) \tag{1}$$

$$\text{s.t. } S_i(t) = \emptyset, \forall i, t \in [0, \tau_i), \tag{2}$$

$$S_{i_1}(t) \cap S_{i_2}(t) = \emptyset, \forall i_1 \neq i_2, t, \tag{3}$$

$$\bigcup_{\forall i} S_i(t) = \Omega, \forall t, \tag{4}$$

$$\tau_i \in \Psi,\ x_i \in \Omega, \forall i. \tag{5}$$

The objective (1) minimizes the total customer service cost over the entire spatiotemporal continuum $\Omega \times \Psi$, and the sum of fixed facility investment costs. Constraints (2) enforce that a service

[1] In fact, the service area can contain multiple separated continuous compact subareas (Ouyang and Daganzo 2006).

region should be an empty set before the construction of its corresponding facility. Constraints (3)–(4) stipulate that the set of service regions of the opened facilities always form an non-overlapping partition of Ω (e.g., customers are always served by the nearest facility where ties can be broken arbitrarily).

Following the pioneering work of Ballou (1968), there has been a stream of research literature on the DFLP. See Arabani and Farahani (2012) for a review. Wesolowsky (1973) extended the static single facility location model to one with a sequence of location decisions in multiple periods. Schilling (1980) developed a heuristic method based on multi-objective analysis for dynamic planning of public-sector facilities. Erlenkotter (1981) compared the performance of seven heuristic methods for the DFLP with a growing demand. Roy and Erlenkotter (1982) and Frantzeskakis and Watson-Gandy (1989) introduced branch and bound approaches based on dual ascent and state-relaxation, respectively. Daskin et al. (1992) discussed the impact of time horizon length on DFLP and found an optimal forecast time horizon. Hormozi and Khumawala (1996) proposed a dynamic programming method to solve the DFLP with time dependent weights. Current et al. (1997) analyzed a dynamic p-Median location problem that minimizes the expected value of opportunity loss (or the maximum regret). Despite various differences among these proposed solution approaches, they typically involve dynamic programming or its variants (often with heuristics) and hence suffer from the "curse of dimensionality", i.e., obtaining the solution becomes extremely challenging as the size of the problem instance increases.

One approach that showed promise in overcoming the computational difficulties associated with large-scale facility location problems is continuous approximation (CA). The underlying idea of CA is to approximate discrete facility locations and their service regions by using continuous facility service area functions; hence the problem is to determine the optimal service area size for each facility. This approach avoids the combinatorial nature of facility location decisions and typically yields an analytically tractable solution, often in closed-forms (e.g., Newell 1973, Langevin et al. 1996, Daganzo 2005). The CA approach has been applied extensively to various sorts of facility location problems as shown in the review paper by Geoffrion et al. (1995). In addition, there has been more diverse applications of CA in the recent literature. Shen and Qi (2007) implemented CA to solve a nonlinear sub-problem from a Lagrangian relaxation framework for supply chain network design. Cui et al. (2010), Li and Ouyang (2010), and Lim et al. (2013) used CA to investigate facility location problems under the impacts of random facility disruptions. Dasci and Laporte (2005) and Wang and Ouyang (2013) used continuous models to optimize bi-level facility location design in the presence of competition. Wang et al. (2013) further developed CA models for Stackelberg games that address complex interactions and competitions among manufacturers, farmers, markets, and the government in a biofuel supply chain.

The CA solution to facility location models is normally in the form of a continuous facility density. To translate the CA solution into a set of discrete locations, Ouyang and Daganzo (2006) proposed a Disk Model, which approximates the service region of a facility by a circular disk. Each disk center indicates a specific facility location and the disk size is determined by the optimal facility density function at that location. The algorithm searches for a spatial distribution of overlapping disks that resembles a reasonable set of facility locations and service regions. We will soon discuss this model in more detail.

Although CA has been widely used to solve static facility location problems, to the best of our knowledge, very little work has been done to handle the dynamic version of these problems. The main idea behind CA is to localize the spatial decisions; i.e., the overall problem is decomposed into point-wise optimization problems, each depending only on local parameters in the neighborhood. Based on this idea, Campbell (1990) extended CA to design transportation terminals in a dynamic setting, but assuming that the terminal locations are "mobile" and can be moved at each time instant independently. However, in reality, locations of open facilities typically remain unchanged in the remainder of the planning horizon. We refer to this as the *location consistency constraint*. Due to this constraint, a facility deployment decision imposes a persisting impact on the remaining planning horizon and cannot be determined over local parameters in the temporal neighborhood. In addition, even if the solution to the DFLP is obtained using CA, a translating method is still needed to effectively convert the CA solution into discrete facility locations over time and space. The aforementioned disk model (Ouyang and Daganzo 2006) is designed for static problems and hence cannot be applied directly to the dynamic ones.

This paper aims to fill these gaps. We propose a CA based modeling framework constructed as follows. First, we formulate a continuous model for the DFLP by augmenting the time dimension, while relaxing the location consistency constraints. To translate the CA output into a set of discrete facility locations, we extend the disk model (for one static time period) to a *tube model* (for multiple time periods). Then, the location consistency constraints are enforced through a nonlinear optimization model with penalty terms. Lastly, we propose an iterative *tube regulation algorithm* to solve the penalty-based optimization problem. We analyze the accuracy and convergence of our modeling framework and conduct numerical experiments to verify its performance. The model and the solution procedure we propose are very generic and flexible; thus, it can be extended to variants of the DFLP (e.g., incorporating existing facilities at the beginning of the horizon).

The remainder of the paper is organized as follows. In §2, we present the CA model framework for the DFLP; §2.1 first reformulates the DFLP into a continuous version, and then derives the optimal solution with approximate facility building time; §2.2 extends the disk model into a tube model and shows how the CA solution is translated into discrete facility locations; §2.3 describes the tube regulation algorithm that enforces location consistency over time and discusses its convergence. In §3, we discuss the approximation accuracy of the proposed algorithm. In §4, we show a series of numerical examples to verify model performance and draw managerial insights. Finally, §5 concludes the paper and discusses future research.

2. Model Framework

2.1. Continuous approximation

We first present a continuous reformulation of the DFLP where the location consistency constraint is relaxed. Suppose Ω is sufficiently large and both z and h are slow varying with respect to space and time[2]. Then the optimal spatial service region should be as "round" as possible (Daganzo and Newell 1986) and its area size can be approximated by a slow varying smooth function $A(x,t) \approx |S_i(t)|$ for $x \in S_i(t)$. Furthermore, the customer service cost $z(x, x_i, S_i(t), t)$ can be approximated by a slow varying function $z_C(x, A(x,t), t)$ and the facility fixed cost can be prorated over time and space as $h(x,t)/(A(x,t)T)$, resulting in a local cost function $p_C(x, A(x,t), t) = z_C(x, A(x,t), t) + h(x,t)[A(x,t)T]^{-1}$ per unit area per unit time. Hence, the DFLP can be reformulated as follows.

$$\min_{A(x,t)} P_C = \int_\Psi \int_\Omega p_C(x, A(x,t), t)\,dx\,dt$$
$$\text{s.t. } A(x,t) \geq 0, \forall x \in \Omega, t \in \Psi.$$

To solve the above problem, we use the solution to a homogeneous system where all parameters are constant over time and space as a building block toward a general heterogeneous system. In a homogeneous system, the fixed cost $h(x,t) \equiv h$ is constant and the customer service cost $z_C(x, A(x,t), t) \equiv z_C(A(x,t))$ is only a function of the service region size. Under this assumption, the DFLP is reduced to a simple static facility location problem, and it is straightforward to see that all facilities should be built uniformly in the beginning of the planning horizon, i.e., $A(x,t) \equiv A$. Hence, the optimization problem for such a homogeneous system reduces to

$$\min_{A \geq 0} p_C(A) = z_C(A) + h(AT)^{-1}. \tag{6}$$

Further, in most cases, the optimal service region size, A^*, satisfies the first order condition from (6); i.e.,

$$\frac{d}{dA}[z_C(A)] - h\left(A^2T\right)^{-1} = 0.$$

For a general heterogeneous system, the optimal service region size function near location $x \in \Omega$ and time $t \in \Psi$, $A^*(x,t)$, is approximated by the solution to a homogeneous system that takes local parameters at (x,t) as the input. That is,

$$\frac{\partial}{\partial A}\left[z_C(x, A^*(x,t), t)\right] - T^{-1}h(x,t)\left[A^*(x,t)\right]^{-2} = 0, \forall x, t. \tag{7}$$

[2] In a sufficiently large service area, its boundary has negligible impact on the solution (Cui et al. 2010). Slow varying z implies the customer demands are also slow varying.

The number of opened facilities at time t for the above problem can then be approximated by

$$n(t) = \left\lfloor \int_{\Omega} \left[A^*(x,t)\right]^{-1} dx \right\rceil,$$

where operator $\lfloor x \rceil$ yields the nearest integer to x. In a growing market, $n(t)$ is a monotone increasing step-function over time t, hence it contains $n(T) - n(0)$ discontinuous jumps in the interval $(0, T)$. Without loss of generality, we assume that the facilities are sorted based on their opening time; i.e., $\tau_{i+1} \geq \tau_i$ for all i. The optimal facility opening time can be approximated by

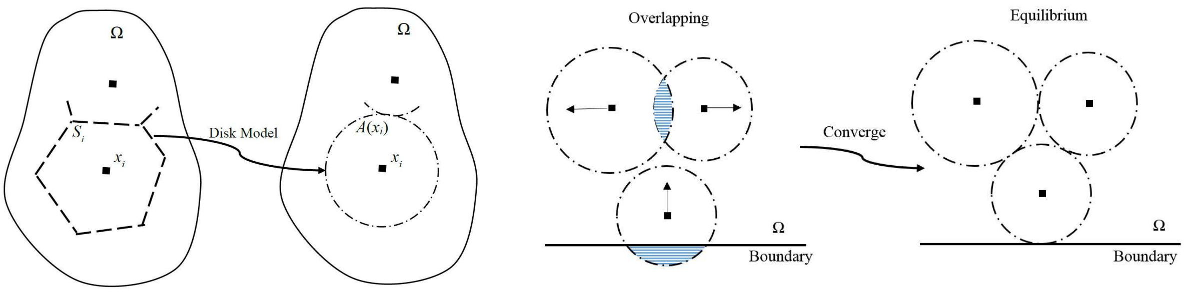
$$\tau_i \approx \inf_{t \geq 0} \{n(t) = i\}. \tag{8}$$

For notational convenience, we let $n_0 = n(0)$ and $n_T = n(T)$ respectively denote the numbers of facilities opened at the beginning and the end of the planning horizon. Hence, facilities $i = 1, \cdots, n_0$ are all opened at time 0; i.e., $\tau_i = 0$, for all $1 \leq i \leq n_0$. We further let $\tau_{n_T+1} := T$, and divide the entire planning horizon Ψ into a set of $(n_T - n_0 + 1)$ levels (i.e., time intervals), $\{\Psi_j : [\tau_{n_0+j-1}, \tau_{n_0+j}], j \in J\}$, where $J = \{1, 2, \cdots, n_T - n_0 + 1\}$. We denote $l_j = |\Psi_j|$ to be the length of level j. The number of facilities remains constant in each level. In the next section, we introduce a tube model that translates the CA solution into discrete facility locations.

2.2. Tube model

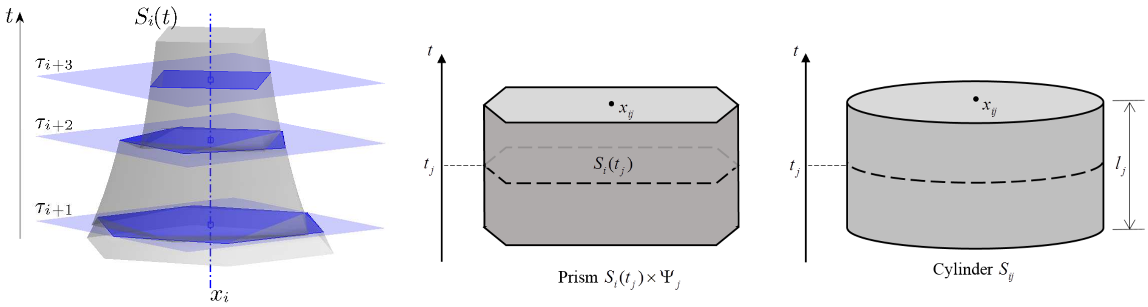
In this subsection, we extend the disk model (Ouyang and Daganzo 2006) to a tube model by incorporating a time dimension. The disk model approximates the service region of a facility to a circular disk in a static problem. As illustrated in Figure 1(a), S_i is the service region of facility i, located at x_i. This is then approximated by a disk with size $A(x_i) \approx |S_i|$, also centered at x_i. Given the output of the CA model, the optimal service region size function $A^*(x), \forall x \in \Omega$, the disk model first generates a set of disks as an initial approximation, while the number of disks is approximated by $\int_{\Omega} [A^*(x)]^{-1} dx$. At the beginning, the algorithm randomly selects the center of each disk, x_i, and computes its size as $A^*(x_i)$. Then repulsive forces are imposed on the disks that overlap with another disk or with the boundary of Ω. With the imposed forces, the disks reposition themselves to reduce the overlaps. As the disks move, the repulsive forces and disk sizes are updated according to the new disk locations. This iterative process repeats until an equilibrium distribution of non-overlapping disks is found (see Figure 1(b)). Then a weighted Voronoi tessellation is used to obtain facility service areas based on the disk centers.

We now incorporate the time dimension into the disk model. For the DFLP, the optimal service region of a facility i, $S_i(t)$, is often time varying or even discontinuous, as shown in Figure 2(a). However, under slow varying parameters, the service regions will not vary significantly over time unless a new facility is opened in the neighborhood area. In other words, we can approximate



(a) (b)

Figure 1 (a) Approximating service regions by disks; (b) Disk movements under repulsive forces.



(a) (b) (c)

Figure 2 (a) An example of time varying $\mathbf{S}_i$; (b) Approximation of $\{S_i(t)\}_{t\in\Psi_j}$ by $S_i(t_j)\times\Psi_j$; (c) Cylinder approximation of $\{S_i(t)\}_{t\in\Psi_j}$.

the service region of an open facility within a level j by a prism, as shown in Figure 2(b), i.e., $S_i(t)\approx S_i(t_j), \forall t\in\Psi_j$ where $t_j=\frac{1}{2}\left(\tau_{n_0+j}+\tau_{n_0+j-1}\right)$ is the median of Ψ_j.

Under this condition, the facility fixed cost and the service cost per unit area near x in level j near x can be respectively approximated as follows:

$$h\left(x,\tau_{n_0+j-1}\right)\approx h\left(x,t_j\right), \tag{9}$$

$$\int_{\tau_{n_0+j-1}}^{\tau_{n_0+j}} z\left(x,x_i,S_i(t),t\right)dt\approx l_j z\left(x,x_{ij},S_i(t_j),t_j\right). \tag{10}$$

Therefore, given $\{t_j\}_{j\in J}$ and $\{l_j\}_{j\in J}$, the original DFLP can be approximated by the following level-based DFLP.

Level-Based DFLP:

$$\min_{\{x_{ij}\}_{\forall i,j},\{S_i(t_j)\}_{\forall i,j}} \sum_{j\in J}\left\{l_j\int_{\Omega} z\left(x,x_{ij},S_i(t_j),t_j\right)+h(x_{n_0+j-1,j},t_j)\right\}dx+\sum_{i=1}^{n_0-1}h(x_{i1},t_1) \tag{11}$$

$$\text{s.t. } x_{ij}=\bar{x}_i, \forall i,j, \tag{12}$$

$$S_i(t_j)=\emptyset, \forall i, j<i-n_0+1, \tag{13}$$

$$S_{i_1}(t_j)\cap S_{i_2}(t_j)=\emptyset, \forall i_1\neq i_2, j, \tag{14}$$

$$\bigcup_{\forall i} S_i(t_j) = \Omega, \forall j, \tag{15}$$

$$x_{ij} \in \Omega, \forall i, j. \tag{16}$$

The level-based DFLP focuses on the variations of service regions when a new facility is open. The objective function (11) approximates (1) by summing the approximated costs (9)–(10) across all levels. The terms in the first summation capture the total service cost and facility fixed costs for all the facilities opened within each level, while the second summation captures the facility fixed costs for all those opened at time 0. Constraints (12) enforce location consistency of facilities across all levels. Constraints (13)–(15) are reformulations of Constraints (2) to (4) in each level.

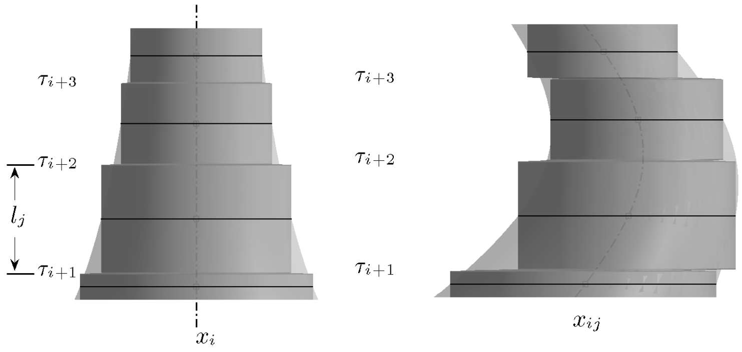
Note that $S_i(t_j)$ can be approximated by a circular disk (Ouyang and Daganzo 2006). We extend this method to approximate $\{S_i(t)\}_{t\in\Psi_j}$ by a cylinder S_{ij} as shown in Figure 2(c), where the height of the cylinder is l_j, the cross-section of the cylinder is a circle with size $|S_i(t_j)|$, and its center is located at x_{ij}. When $i > n_0 + j - 1$, facility i has not been opened at time t_j $(S_i(t_j) = \emptyset)$ and the cylinder degenerates to a line.

Then a collection of cylinders corresponding to facility i across all levels $j \in J$, $\{S_{ij}\}_{j\in J}$, forms a *tube* that approximates $\mathbf{S}_i$. That is, each time level is represented by a cylinder and a set of cylinders form a tube. The location consistency constraints require $x_{ij} = x_i, \forall j \in J$. We refer to such a tube satisfying the location consistency constraints as a *right* tube, as shown in Figure 3(a). Otherwise, we say a tube is *oblique* if there exists $j, j' \in J$ such that $x_{ij} \neq x_{ij'}$, as shown in Figure 3(b). The center of a general tube $\{S_{ij}\}_{j\in J}$ is defined as the weighted average of all its non-degenerated cylinder centers; i.e.,

$$\bar{x}_i := \left(\sum_{j=\max\{i-n_0+1,1\}}^{n_T-n_0+1} l_j x_{ij}\right)\left(\sum_{j=\max\{i-n_0+1,1\}}^{n_T-n_0+1} l_j\right)^{-1}. \tag{17}$$

Hence, by approximating each prism by a cylinder, the optimal solution to the level-based DFLP can be represented by a set of non-overlapping right tubes.

The above tube approximation inspires us to convert the CA solution (to the original DFLP, $A(x,t)$) into the solution to the level-based DFLP (which is approximated by a set cylinders that are piled into right tubes); i.e., $A(x,t) \approx S_i(t_j), \forall x \in S_i(t), t \in \Psi_j$. The optimal CA solution $A^*(x,t)$, obtained from (7), can therefore be discretized into a set of non-overlapping right tubes. In doing so, we propose a two-step method: (i) first relax the location consistency requirement (over time) and convert $A^*(x,t)$ to a set of (oblique) tubes; (ii) enforce the tubes into right ones via penalty-based regulation method. The first step is rather straightforward. We implement the disk model for each level $j \in J$ to convert $A^*(x,t_j)$ into a set of cylinders $\{S_{ij}\}_{i\in I}$, each with a given height l_j. Since the disk model is implemented independently on each level, it is likely that $x_{j_1 i} \neq x_{j_2 i}$ for



(a) (b)

Figure 3 (a) A right tube; (b) An oblique tube.

$j_1 \neq j_2$ and $\{S_{ij}\}_{j\in J}$ collectively form an oblique tube (see Figure 3(b)). The second step involves an iterative tube regulation algorithm to adjust the cylinder centers in each tube. Once the tubes converge to a set of non-overlapping right ones, we have obtained an solution to the level-based DFLP. More detail on tube regulation will be introduced in the following section.

2.3. Tube regulation

Note that any set of oblique tubes approximates a feasible solution to the level-based DFLP, where the location consistency constraints (12) are relaxed. We now allow the locations of facilities to change over time and enforce Constraints (12) by imposing a penalty term in the objective. Hence, the level-based DFLP is extended to the following problem.

Tube Regulation Problem (TRP):

$$\begin{aligned}\min_{\{x_{ij}\}_{\forall i,j},\{S_i(t_j)\}_{\forall i,j}} &\sum_{j\in J}\left\{l_j\int_\Omega z\left(x,x_{ij},S_i(t_j),t_j\right)+h(x_{n_0+j-1,j},t_j)\right\}dx+\sum_{i=1}^{n_0-1}h(x_{i1},t_1)\\ &+\sum_{\forall i}\sum_{j=\max\{i-n_0+1,1\}}^{n_T-n_0+1}\phi\left(\lambda_{ij}||x_{ij}-\bar{x}_i||\right)\\ \text{s.t. } &(13)-(16),\end{aligned} \tag{18}$$

where

$$\phi(r) := \begin{cases}0, & r\le 1,\\ \infty, & r>1,\end{cases}$$

is a step-wise penalty function and $\lambda_{ij} > 0$ is the penalty coefficient bounding the location "deviation" of cylinder S_{ij}. When there exists a facility location x_{ij} such that $||x_{ij} - \bar{x}_i|| > 1/\lambda_{ij}$, the objective function goes to infinity. Hence, the feasible region of x_{ij} in the TRP is given by $\{x : ||x - \bar{x}_i|| \le 1/\lambda_{ij}\}$. Given $\{\lambda_{ij}\}_{\forall i,j}$, the optimal solution to the TRP, $\{x_{ij}(\lambda_{ij})\}_{\forall i,j}$ and $\{S_i(t_j, \lambda_{ij})\}_{\forall i,j}$, determine a set of oblique tubes. For all i, j, there exists a minimum penalty coefficient, i.e.,

$\bar{\lambda}_{ij} = ||x_{ij} - \bar{x}_i||^{-1}$, with which the optimal solution to the TRP corresponds to the aforementioned oblique tubes converted from $A^*(x,t)$ via the disk model (as described in Section 2.2). Hence, we set the initial value of $\{\lambda_{ij}\}_{\forall i,j}$ to its corresponding minimum penalty coefficient, $\{\bar{\lambda}_{ij}\}_{\forall i,j}$.

To solve TRP, we increase the penalty coefficients $\lambda_{ij}, \forall i,j$, in each iteration k; e.g., making it proportional to $\sqrt{k}$. As λ_{ij} increases over the iterations, the feasible region for x_{ij} shrinks, which forces the cylinders to align toward the center. Thus, the corresponding tubes become "less" oblique. This however may result in a greater degree of overlaps among the cylinders. This procedure reflects the trade-off between the system cost and the facility location consistency requirement.

We define a repulsive force between any two overlapping cylinders, and let its magnitude represent the loss of the objective value due to such overlap. To illustrate the idea, we use the simpler notation from the homogeneous continuous model (6), and assume the objective $p_C(A)$ is twice differentiable (due to the slow varying assumption). When the cylinders are with the optimal size A^*, the first order derivative of p_C is zero, i.e., $\frac{d}{dA}p_C(A)\big|_{A=A^*} = 0$. Suppose the size of the cross-sectional overlapping area between two adjacent cylinders is $\Delta_A \ll A$, from Taylor series expansion of p_C at A^*, the change of p_C due to the change in Δ_A can be approximated as

$$\Delta p_C = p_C(A^* + \Delta_A) - p_C(A^*) \approx \frac{\Delta_A^2}{2} \cdot \frac{d^2 p_C(A)}{dA^2}\bigg|_{A=A^*}.$$

Then, we let the repulsive force $\mathbf{f}(\Delta_A)$ be a function of the overlapping area Δ_A (Peng et al. 2013), whose magnitude equal to the gradient of Δp_C with respect to Δ_A; i.e.,

$$|\mathbf{f}(\Delta_A)| = \frac{d(\Delta p_C)}{d(\Delta_A)} = \Delta_A \cdot \frac{d^2 p_C(A)}{dA^2}\bigg|_{A=A^*}. \tag{19}$$

The direction of the repulsive forces $\mathbf{f}(\Delta_A)$ is along the line connecting the centers of overlapping cylinders.

At every step of the iteration, for every level j, all repulsive forces are calculated based on the current locations and the sizes of the cylinders. Figure 4(a) illustrates the generation of repulsive forces among overlapping cylinders. The total force on each tube $\{S_{ij}\}_{j\in J}$ is computed as the weighted vector sum of all forces across all non-degenerated cylinders, with weights equal to the cylinder heights. The formula would have a similar form to that of the tube center formula (17), except that location is now replaced by a vector force.

Based on the repulsive forces, we shift each cylinders and tubes to find the best displacement as the solution to the TRP for each iteration k. The single movement of each cylinder or tube is controlled by a diminishing and cumulatively unbounded force-to-distance factor, β_k, satisfying $\lim_{k\to\infty} \beta_k = 0$ and $\sum_{k=1}^{\infty} \beta_k = \infty$, and a small force perturbation is added to avoid local optimum at each move. Specifically, the shift includes the following two stages: (i, cylinder movement), for each tube, shift all cylinders in the tube according to their repulsive forces (when the cylinder center goes out its feasible region, it will be projected to the nearest feasible point) until convergence, as shown in Figure 4(b), updating the repulsive forces after each movement; (ii, tube movement) shift the entire tube according to the total force until converging, i.e., all cylinders in the tube are moved identically. Such an algorithm continues for each iteration until the system converges to a set of right tubes with non-overlapping cylinders at each level, and its convergence is shown by the following proposition.

PROPOSITION 1. *(a)* ***Convergence of cylinder movements****. Consider a non-degenerate cylinder. Suppose all the other cylinders and the feasible region sizes are fixed. The cylinder center location will converge to a local minimum with respect to the total cost over iterations.*

(b) ***Convergence of tube movement****. Consider a tube. Suppose the relative locations of each cylinder in the tube, feasible region sizes, and all other tubes are fixed. The tube center location will converge to a local minimum with respect to the total cost under such forces over iterations.*

Proof. (a) Without loss of generality, let's consider a cylinder S_{11}. When all other cylinders are fixed, the total cost increment is a function with respect to the location of S_{11}. From the definition of the repulsive force (19), we know the vector sum of all repulsive forces on S_{11} is a gradient descent direction of the total cost increment (caused by the overlaps between S_{11} and its adjacent cylinders or boundaries). Hence the total cost decreases when S_{11} moves along such a direction for a sufficiently small distance. Note that each movement is projected into a convex circular feasible region. Given a diminishing force-to-distance factor, S_{11} will converge to a local minimum with respect to the total cost under such gradient projection method (Boyd and Vandenberghe 2004).

(b) The tube movement convergence is a simple extension of cylinder movement convergence. From the first term in the objective function (11), we know the total cost increment caused by a tube is the weighted sum of all those from its cylinders across all time levels. Hence, the weighted vector sum of overlapping forces is the gradient descent direction of the total cost. Given a diminishing force-to-distance factor, the tube will converge to a local minimum with respect to the total cost under such gradient search method. □

The aforementioned algorithm fixes all other cylinders or tubes while searching for a local minimum for one tube. Although this is an accurate process, solving the TRP at each iteration can be very time consuming and ineffective in practice. Therefore, we propose an accelerated algorithm by shifting all cylinders and tubes simultaneously without awaiting convergence for a particular tube. We observe that the step size of cylinder movements approaches zero as the feasible region shrinks to the center of the tube, and as such, the tube movement will become the dominating component of the recursion (which resembles solving the TRP to convergence). Obviously, the rate at which the penalty coefficients increase affects the convergence rate of cylinder movements in the accelerated algorithm. To see this, we consider a polar coordinate system built at the center of the tube with the polar axis pointing from the tube center to the cylinder center. Suppose the angle between the polar axis and the joint repulsive force on the cylinder is θ, simple calculation shows that in iteration k, the angular coordinate of the cylinder center changes by a small angle

$$\frac{c_2 \sin\theta}{k/\lambda_k + c_2 \cos\theta},$$

where c_2 is a positive constant, and λ_k is a parameter related to the value of λ_{ij} in iteration k. As long as λ_k increases at a rate that is sublinear to k, the change of the angular coordinate of the cylinder will converge to zero. In our algorithm, we choose $\lambda_k = \bar{\lambda}_{ij}\sqrt{k}$.

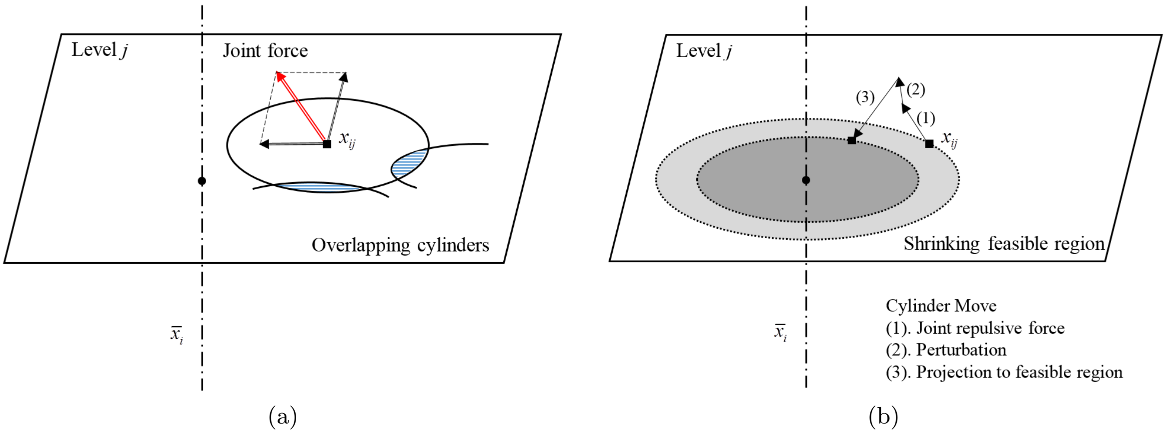


Figure 4 **Tube regulation. (a) Repulsive forces; (b) Location adjustment.**

3. Approximation Error Bounds

In this section, we evaluate the approximation error of our solution and explore upper and lower bounds of the optimal DFLP solution. The upper bound is directly given by the discrete solution from the proposed tube regulation algorithm since it is feasible, while the lower bound is normally given by the objective of the CA solution under certain conditions.

To stay focused, we consider the following specific DFLP. Near location x at time t, the customer demand is denoted by $D(x,t)$ and the facility operating cost is denoted by $C(x,t)$. Suppose customers always choose the nearest facility and the transportation cost is 1 per unit distance. Then, we can obtain the objective function of this DFLP as

$$P = \int_\Psi \sum_{\forall i} \int_{x\in S_i(t)} ||x - x_i|| D(x,t)dxdt + \int_\Psi \sum_{\forall i} C(x_i,t)dxdt + \sum_{\forall i} h(x_i,\tau_i). \tag{20}$$

Accordingly, the service cost in the CA formulation is

$$z_C\left(x, A(x,t), t\right) = \frac{2}{3\sqrt{\pi}} D(x,t) A^{1/2}(x,t) + C(x,t) A^{-1}(x,t)$$

and the objective function of CA per unit area per unit time is

$$p_C\left(x, A(x,t), t\right) = \frac{2}{3\sqrt{\pi}} D(x,t) A^{1/2}(x,t) + C(x,t) A^{-1}(x,t) + \frac{1}{T} h(x,t) A^{-1}(x,t).$$

Hence, we can calculate the first and second order derivatives of p_C with respect to A to obtain the optimal service region size function as follows:

$$A^*(x,t) = \begin{cases} \left[\frac{3\sqrt{\pi}}{D(x,t)T}\left(C(x,t)T + h(x,t)\right)\right]^{2/3}, & \text{if } \frac{\partial^2}{\partial A^2}\left[p_C(x, A(x,t), t)\right]\Big|_{A(x,t)=A^*(x,t)} > 0 \\ \infty, & \text{otherwise.} \end{cases}.$$

Further, the minimum of the CA objective can be obtained by

$$P_C^* = \int_\Psi \int_\Omega p_C\left(x, A^*(x,t), t\right) dx dt.$$

Proposition 2. *We have $P_C^* \leq P$ if the following three conditions are satisfied:*
(a) x_i is the centroid of $S_i(t)$ for all i and t;
(b) within each service region $S_i(t)$ at a given time t, the demand density $D(x,t)$ is constant, $D_i(t)$, and the operating cost $C(x,t)$ is concave with respect to x; and
(c) the fixed cost $h(x,t)$ is uniformly far smaller than the total serving cost, or decreasing over time and far larger than the total serving cost, or remains constant across all x and t.

Proof. We build a piece-wise constant service area function $A_s(x,t)$ satisfying $A_s(x,t) = |S_i(t)|, \forall t, x \in S_i(t)$. It is obvious that $P_C^* \leq \int_\Psi \int_\Omega p_C\left(x, A_s(x,t), t\right) dx dt$ since $A^*(x,t)$ is the minimum. We only need to show $P \geq \int_\Psi \int_\Omega p_C\left(x, A_s(x,t), t\right) dx dt$.

First, we consider the transportation cost term and show

$$\int_\Psi \sum_{\forall i} \int_{x \in S_i(t)} ||x - x_i|| D(x,t) dx dt \geq \int_\Psi \int_\Omega \frac{2}{3\sqrt{\pi}} D(x,t) A_s^{1/2}(x,t) dx dt.$$

It is sufficient to show

$$\int_{x \in S_i(t)} ||x - x_i|| D(x,t) dx \geq \int_{x \in S_i(t)} \frac{2}{3\sqrt{\pi}} D(x,t) A_s^{1/2}(x,t) dx, \ \forall i, t,$$

which is equivalent to

$$\int_{x \in S_i(t)} ||x - x_i|| dx \geq \frac{2}{3}\sqrt{\frac{|S_i(t)|}{\pi}} \tag{21}$$

due to the piece-wise constant property of $D(x,t)$ and $A_s(x,t)$. Since x_i is the centroid of $S_i(t)$, (21) holds (Ouyang and Daganzo 2006).

Then we consider the operating cost term. Since $C(x,t)$ is concave with respect to x, the Jensen inequality indicates

$$C(x_i,t) \geq \frac{1}{|S_i(t)|}\int_{x\in S_i(t)} C(x,t)dx,\ \forall i,t.$$

Hence $C(x_i,t) \geq \int_{x\in S_i(t)} C(x,t)A_s^{-1}(x,t)dx$. Integrating it over time Ψ and summing it across i leads to

$$\int_\Psi \sum_{\forall i} C(x_i,t)dxdt \geq \int_\Psi \int_\Omega C(x,t)A_s^{-1}(x,t)dxdt.$$

Finally, we investigate the fixed cost term. A trivial case is when $h(x,t)$ is significantly smaller than the total service cost. We can directly neglect the fixed cost and conclude $P \geq \int_\Psi \int_\Omega p_C\left(x, A_s(x,t), t\right) dxdt$. When $h(x,t)$ is significantly larger than the total serving cost and decreases over t, most facility should be built at the beginning and we have

$$\int_\Psi \int_{x\in S_i(t)} \frac{1}{T} h(x,t)|S_i(t)|^{-1} dxdt \leq h(x_i,\tau_i).$$

When $h(x,t) \equiv h, \forall x,t$, we have

$$\int_\Psi \int_{x\in S_i(T)} \frac{1}{T} h(x,t)|S_i(T)|^{-1} dxdt \leq h.$$

Now, we have shown each term in P is greater or equal to that in $\int_\Psi \int_\Omega p_C\left(x, A_s(x,t), t\right) dxdt$, which completes the proof. □

4. Numerical Examples

In this section, we provide numerical examples to illustrate the application of the proposed CA framework, test its computational performance, and explore the impact of heterogeneity.

4.1. Illustrative Case Study

We consider the specific type of DFLP discussed in §3, whose objective function follows (20). Suppose the entire customer area is a unit square, i.e., $\Omega = [0,1]\times[0,1]$. Further, we set $T = 10$, $D(x,t) = 500\cos(||x||)e^{1+0.04t}$, $C(x,t) = 3\sin(||x||)e^{2-0.04t}$, and $h(x,t) = 10\cos(||x||)e^{1-0.04t}$. Since $D(x,t)$ increases and $C(x,t)$, $h(x,t)$ decrease with respect to t, it is intuitive that $A^*(x,t)$ decreases with t. This implies a growing market. The optimal facility deployment times, $n(x,t)$, are obtained via continuous approximation, as shown in Figure 5(a). The solution to this numerical example suggests deploying five facilities in the beginning, $n_0 = 5$, and adding four additional facilities until the end of the planning horizon, $n_T = 9$. Specifically, we have $\tau_i = 0,\ i = 1,2,3,4,5$, and $\tau_6 = 0.56$, $\tau_7 = 3.69$, $\tau_8 = 6.37$, and $\tau_9 = 8.72$. Figures 5(b) and 5(c) illustrate the optimal service region at $t = 0$ and T, respectively.

We now illustrate how the tube model is implemented for this example. For each iteration k, we update $\lambda_{ij} = \bar{\lambda}_{ij}\sqrt{k}$ to generate a force perturbation satisfying a bivariate uniform distribution in

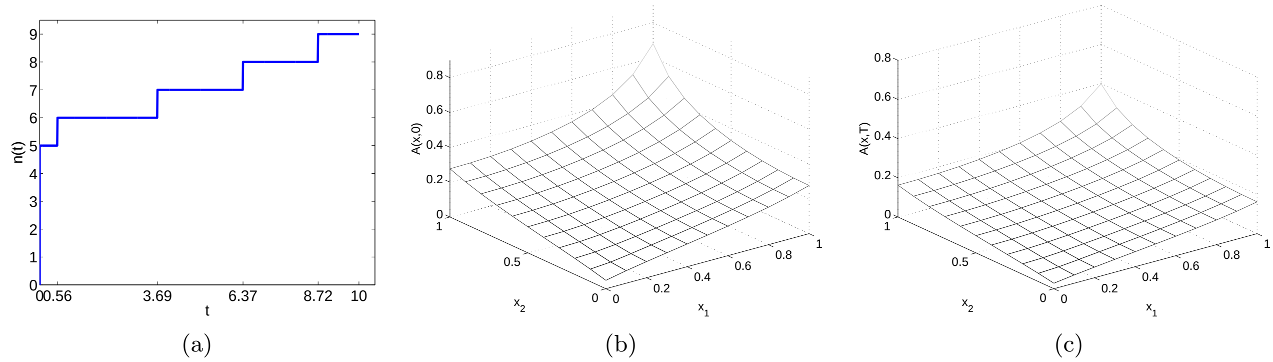


Figure 5 **(a) Result of** $n(t)$**; (b)** $A^*(x,0)$**; (c)** $A^*(x,T)$**.**

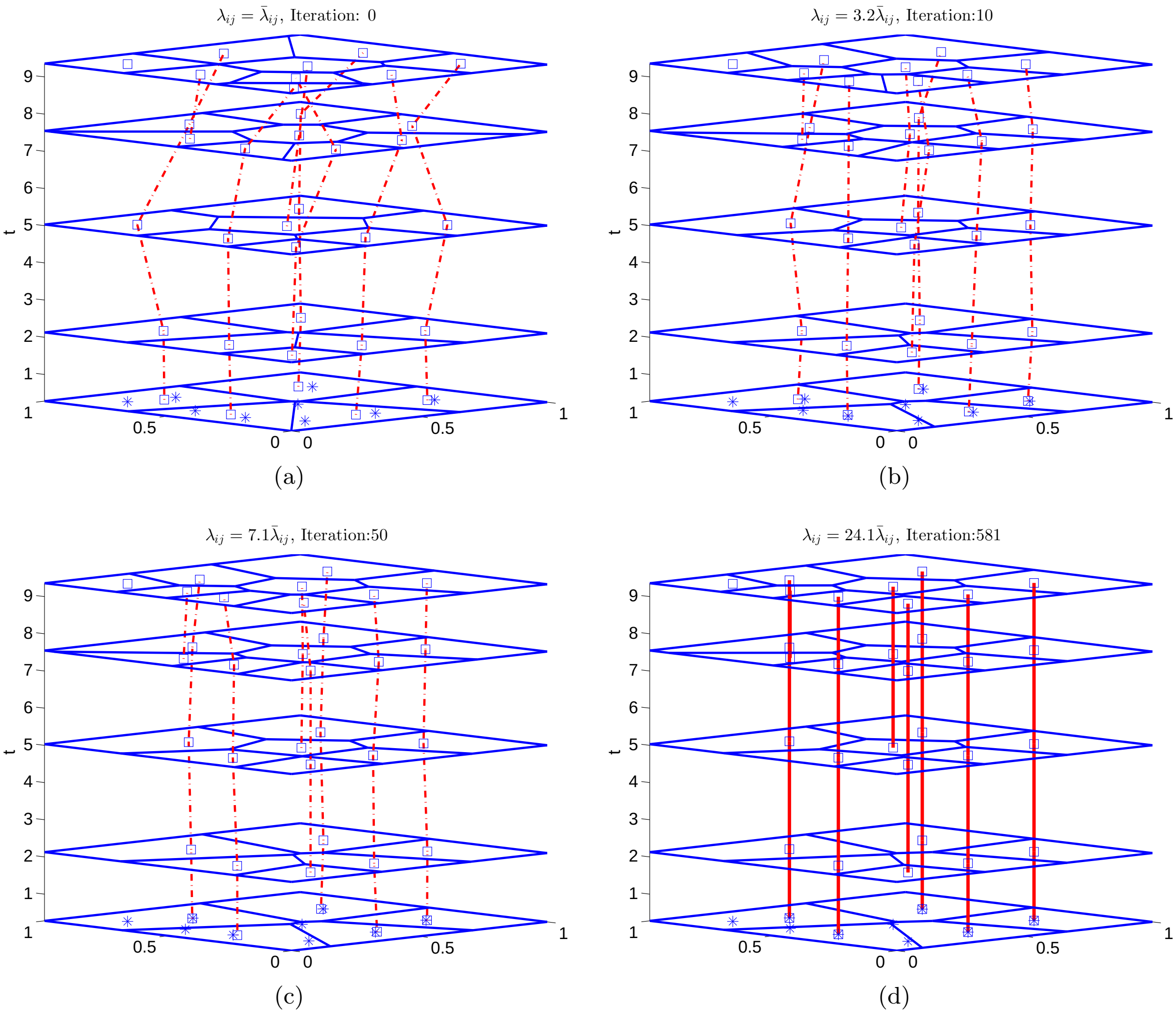


Figure 6 **Tube model solution: (a) Initial oblique tubes generated by the disk model; (b) Solution after 10 iterations, where** $\lambda_{ij} = 3.2\bar{\lambda}_{ij}$**; (c) Solution after 50 iterations, where** $\lambda_{ij} = 7.1\bar{\lambda}_{ij}$**; (d) Converged solution after 581 iterations.**

$[-0.05, 0.05]^2$, and set the force-to-distance factor as $\frac{0.01}{k}$. Figures 6 (a)–(d) exhibit some snapshots of the tube regulation procedure. For clear illustration, each cylinder of a tube is simply represented on a plane, and cylinders of the same tube are connected through lines. The five planes in each sub-figure represent the timing for new facility construction, where the Voronoi diagram captures the service region of each facility (i.e., we let customers choose the nearest facility). The square markers represent the facility locations x_{ij} and the star markers on the first level give the center of each tube $\bar{x}_i$, at convergence to a right tube. The converged solution shown in Figure 6(d) is reached after 581 iterations, where $\bar{x}_i$ becomes an approximation of x_i. In addition, we compute the upper and lower bounds of the problem based on the discussion in §3. For this example, the lower and upper bound of the optimal total costs are 2544.4 and 2683.5, respectively, indicating a 5.47% approximation error.

4.2. Comparison with Discrete Model

Next, we compare the performance of our proposed CA model to that of the discrete model (1)–(4) through a series of DFLP instances. The first test problem instance remains the same as that in §4.1. For the CA model, we assume that the customer demand density, facility operating cost, and facility fixed cost are $D(x,t) = 400\cos(||x||)e^{\frac{1+0.05t}{1+||x||}}$, $C(x,t) = \sin(||x||)e^{\frac{2-0.05t}{1+||x||}}$, and $h(x,t) = 0.001\cos(||x||)e^{1-0.05t}$, respectively. For the discrete model, we discretize Ω into $q \times q$ discrete spatial grid cells and use the center of each cell to represent a customer and a candidate facility location. Thus, we have q^2 overlapping customer demand points and candidate facility locations. We then discretize Ψ into 10 time periods. The total demand of each discrete point is aggregated from each cell. The remaining input parameters are the same as those in §4.1. The discrete model is solved by the CPLEX solver. We set a maximum optimality gap of 1% and a maximum computation time of 10 hours. The solution and computation times for various q values are summarized as Model I in Table 1.

Then, we consider another test problem instance and repeat the same analysis. All parameters remain the same, but the facility operating cost now is a function of the total demand served per unit time, $C(x_i,t) = 5\left(\int_{x\in S_i(t)} D(x,t)dx\right)^{1/2}$. Note that this DFLP is no longer convex, and the discrete model is solved by the GAMS-SBB solver (with the same termination criteria). The solution and computation times for various q values are summarized as Model II in Table 1.

We observe that the computation time for the discrete model increases drastically with the size of instances q^2 for both cases. Further, the solver fails to obtain a solution for the second instance when $q = 15$. In contrast, the CA solution obtains a good approximation of the optimal solution within a very short amount of time, which highlights its practical value.

Table 1 Solution comparison for continuous and discrete model.

Model		q	n_0	n_T	Total cost	Time (s)
I	Discrete	8	15	29	837	19
		10	17	25	874	178
		12	16	21	880	914
		15	16	20	890	16661
	Continuous	∞	15	23	886	112
II	Discrete	8	2	17	5758	316
		10	2	12	6037	2744
		12	1	6	6477	5264
		15	no solution			36000
	Continuous	∞	6	10	6536	72

4.3. Sensitivity to Heterogeneity

Finally, we study how spatial and temporal heterogeneity in data input influences the optimal solution. We consider the same problem instance as in §4.1 and generate spatial and temporal heterogeneity with respect to three key parameters: D, C, and h. We set $D(x,t) = 150(1+\rho_D\cos(\frac{\pi}{\sqrt{2}}||x||)e^{1+0.05(1+\rho_t)t}$, $C(x,t) = (1+\rho_C\sin(\frac{\pi}{\sqrt{2}}||x||)e^{2-0.05(1+\rho_t)t}$, and $h(x,t) = 10(1+\rho_h\cos(\frac{\pi}{\sqrt{2}}||x||)e^{1-0.05(1+\rho_t)t}$, where ρ_D, ρ_C, and ρ_h capture *spatial heterogeneity* and ρ_t captures *temporal heterogeneity*. Note that the system becomes spatially homogeneous when $\rho_D = \rho_C = \rho_h = 0$. To illustrate the impact of heterogeneity, we solve a series of DFLPs, each with one heterogeneity parameter increasing from 0 to 0.5, while others are fixed at zero. The relative difference in the total costs with respect to each type of heterogeneity is shown in Figure 7. Since the solution to DFLP from our CA approach is not necessarily unique, we solve each instance five times and plot the average. The results show that the total cost decreases as spatial heterogeneity increases, and

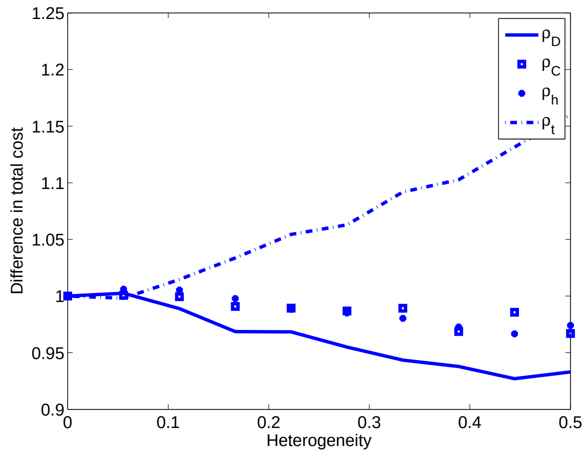


Figure 7 Sensitivity of spatial and temporal heterogeneities.

the heterogeneity of demand density has the most significant impact. In contrast, we find that the total cost increases as temporal heterogeneity increases. As demand density varies faster and the facility operating/opening cost decreases faster over time, it becomes more challenging to maintain low cost.

5. Conclusion

This paper proposes a continuum approximation framework for solving the dynamic facility location problem. We first extend the static continuous facility location problem by augmenting the time dimension. While the conversion yields a solution without maintaining consistent facility locations over time, it provides an analytically tractable solution. We then develop a tube model, an extension of the disk model, to convert the CA solution into discrete facility locations. Facility location consistency is enforced through an iterative tube regulation procedure, i.e., via a penalty method, where the converged solution yields an approximate solution to the DFLP. Through a series of numerical examples, we show the impact of spatial and temporal heterogeneity and also compare the performance of the CA model with that of the discrete counterpart.

As a final note, we point out that the proposed model can be applied when a set of facilities may have already been built at the beginning of the planning horizon. For such cases, we can simply fix the values of such location variables. During the tube regulation procedure, we do not update the centers of the corresponding tubes. This extension provides great practical value in deploying facilities. For example, in case of high degree of system uncertainty and/or long planning horizon, our model can be implemented in a rolling-horizon. We can redefine (modify) the DFLP in each prediction cycle to apply the proposed algorithm and implement the solution in the nearest future. The set of facilities opened up to that point then becomes the fixed initial condition for the new problem, and this can be handled using the procedure discussed above.

This research can be extended in several ways. First, it would be interesting to address the possibility that an open facility could be closed, i.e., the market does not have to maintain growth everywhere. Another possible research direction is to consider the scenario under which an open facility could be disrupted due to natural or human induced hazards. We leave these promising research directions for future work.

References

Arabani, A. B., Farahani, R. Z., 2012. Facility location dynamics: An overview of classifications and applications. Comput. Ind. Eng. 62 (1), 408–420.

Ballou, R. H., 1968. Dynamic warehouse location analysis. J. Marketing Res. 5 (3), 271–276.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, Cambridge, UK.

Campbell, J. F., 1990. Locating transportation terminals to serve an expanding demand. Transportation Res. B 24 (3), 173–192.

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

Current, J., Ratick, S., ReVelle, C., 1997. Dynamic facility location when the total number of facilities is uncertain: A decision analysis approach. Eur. J. Oper. Res. 110 (3), 597–609.

Daganzo, C. F., 2005. Logistics Systems Analysis, 4th Edition. Springer-Verlag, Heidelberg, Germany.

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

Dasci, A., Laporte, G., 2005. A continuous model for multistore competitive location. Oper. Res. 53 (2), 263–280.

Daskin, M. S., Hopp, W. J., Medina, B., 1992. Forecast horizons and dynamic facility location. Ann. Oper. Res. 40 (1), 125–151.

Erlenkotter, D., 1981. A comparative study of approaches to dynamic location problems. Eur. J. Oper. Res. 6 (2), 133–143.

Frantzeskakis, M., Watson-Gandy, C. D. T., 1989. The use of state space relaxation for the dynamic facility location problem. Ann. Oper. Res. 18 (1), 187–211.

Geoffrion, A. M., Morris, J. G., Webster, S. C., 1995. Distribution system design. In: Drezner, Z. (Ed.), Facility Location: A Survey of Applications and Methods. Springer-Verlag, Berlin, pp. 185–202.

Hormozi, A. M., Khumawala, B. M., 1996. An improved algorithm for solving a multi-period facility location problem. IIE Trans. 28 (2), 105–114.

Langevin, A., Mbaraga, P., Campbell, J. F., 1996. Continuous approximation models in freight distribution: An overview. Transportation Res. B 30 (3), 163–188.

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.

Lim, M. K., Achal, B., Chopra, S., Daskin, M. S., 2013. Facility location decisions with random disruptions and imperfect estimation. Manufacturing Service Oper. Management 15 (2), 239–249.

Newell, G. F., 1973. Scheduling, location, transportation, and continuum mechanics; some simple approximations to optimization problems. SIAM J. Appl. Math. 25 (3), 346–360.

Ouyang, Y., Daganzo, C. F., 2006. Discretization and validation of the continuum approximation scheme for terminal system design. Transportation Sci. 40 (1), 89–98.

Peng, F., Ouyang, Y., Wang, X., 2013. Continuum approximation of discrete spatial data for optimal facility location design. In: Proceedings of the 92nd Annual Meeting of the Transportation Research Board (CDROM).

Roy, T. J. V., Erlenkotter, D., 1982. A dual-based procedure for dynamic facility location. Management Sci. 28 (10), 1091–1105.

Schilling, D. A., 1980. Dynamic location modeling for public-sector facilities: A multicriteria approach. Decision Sci. 11 (4), 714–724.

Shen, Z.-J. M., Qi, L., 2007. Incorporating inventory and routing costs in strategic location models. Eur. J. Oper. Res. 179 (2), 372–389.

Wang, X., Lim, M. K., Ouyang, Y., 2013. Food, energy, and environment trilemma: sustainable farmland use and biofuel industry development. Working paper, University of Illinois at Urbana–Champaign.

Wang, X., Ouyang, Y., 2013. A continuum approximation approach to competitive facility location design under facility disruption risks. Transportation Res. B 50, 90–103.

Wesolowsky, G. O., 1973. Dynamic facility location. Management Sci. 19 (11), 1241–1248.