Nearest Neighbor Distance in Three-Dimensional Space

Masashi Miyagawa

Department of Regional Social Management, University of Yamanashi, 4-4-37 Takeda, Kofu, Yamanashi 400-8510, Japan
E-mail address: mmiyagawa@yamanashi.ac.jp

(Received August 4, 2017; Accepted March 16, 2018)

This paper deals with the nearest neighbor distance in three-dimensional space. The distribution of the nearest neighbor distance is derived for grid and random point patterns. The distance is measured as the Euclidean and rectilinear distances. An application of the nearest neighbor distance can be found in facility location problems. The nearest neighbor distance represents the service level of facility location. The distribution shows how the distance to the nearest facility is distributed in a study region, and is useful for facility location problems in three-dimensional space. The distribution of the kth nearest neighbor distance is also derived for the random pattern.
Key words: Point Pattern, Grid Pattern, Random Pattern, Euclidean Distance, Rectilinear Distance

1. Introduction

The nearest neighbor distance—the distance from a random point to its nearest point—has been used in point pattern analysis and location analysis (Clark and Evans, 1954; Cressie, 1993; Illian *et al.*, 2008; Koshizuka and Ohsawa, 1983; Miyagawa, 2008, 2009). The nearest neighbor distance describes the characteristic of point patterns or the service level of facility location.

The distribution of the nearest neighbor distance has been derived for regular and random point patterns. The distribution of the Euclidean distance was derived by Persson (1964), Holgate (1965), and Miyagawa (2009) for the square, triangular, and hexagonal lattice patterns, and Clark and Evans (1954) for the random pattern. The distribution of the rectilinear distance was derived by Miyagawa (2008) for the square and diamond lattice patterns and Larson and Odoni (1981) for the random pattern. The nearest neighbor distance has been extended to three-dimension and the distribution of the Euclidean distance was derived by Mathai (1999) and Haenggi (2005) for the random pattern. The Euclidean distance for the regular pattern and the rectilinear distance for the regular and random patterns in three-dimensional space have not been derived previously.

In this paper, we derive the distribution of the nearest neighbor distance in three-dimensional space. To obtain analytical expressions for the distribution, we focus on grid and random patterns of points, as shown in Fig. 1, and the distance is measured as the Euclidean and rectilinear distances. The grid and random patterns are assumed to be unbounded and continue infinitely. The analytical expressions provide a clear understanding of basic properties of the nearest neighbor distance. In addition, the theoretical results of these two extreme patterns will give an insight into empirical analysis of actual patterns.

An application of the nearest neighbor distance in three-dimensional space can be found in facility location problems in three-dimensional space. Love (1969) formulated a problem of finding the optimal location of facilities in three-dimensional space. Cooper (1973) presented an n-dimensional extension of the generalized Weber problem. Radó (1988) addressed a multi-facility location problem in n-dimensional space. Kon (2001, 2007) studied a multi-criteria location problem in three-dimensional rectilinear space. O'Kelly (2009) extended the minimax hub location problem to three-dimension. Schöbel and Scholz (2010) proposed an algorithm for the three-dimensional Fermat-Weber problem. Thill *et al.* (2011) obtained the optimal location of AEDs in a five-story building. The distribution of the nearest neighbor distance that shows how the distance is distributed in a study region will supply building blocks for facility location models in three-dimensional space.

The remainder of this paper is organized as follows. The next section derives the distribution of the Euclidean nearest neighbor distance for the grid and random patterns in three-dimensional space. The following section derives the distribution of the rectilinear nearest neighbor distance. The final section presents concluding remarks.

2. Euclidean Distance

The Euclidean distance is the most important as a criterion of proximity. Let R be the Euclidean distance from a randomly selected location in three-dimensional space to the nearest point. We call R the nearest neighbor distance. In this section, we derive the distribution of the Euclidean nearest neighbor distance for the grid and random patterns.

2.1 Grid pattern

Suppose that points are regularly distributed on a square grid with spacing a. The density of points (the number of points per unit volume) is then expressed as $\rho = 1/a^3$. Let $F(r)$ be the cumulative distribution function of R. $F(r)$, which is the probability that $R \leq r$, is given by

$$F(r) = \frac{V(r)}{V}, \tag{1}$$

doi:10.5047/forma.2018.002

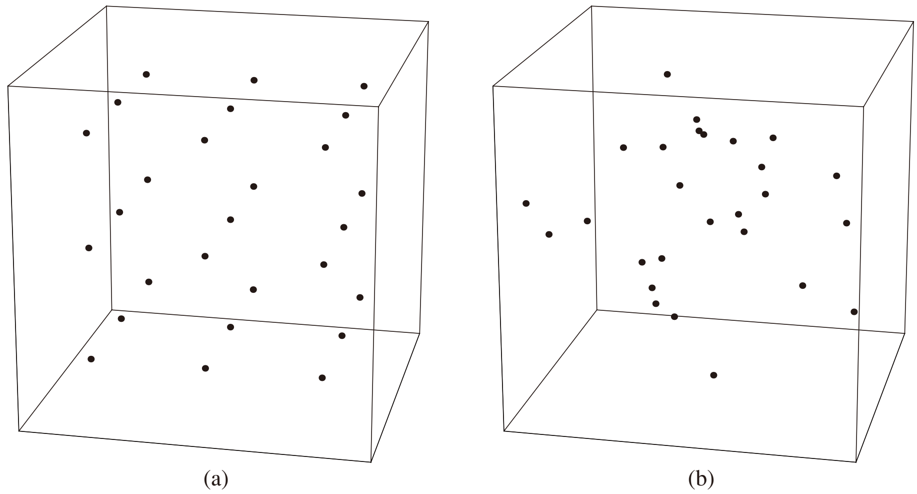

Fig. 1. Point patterns in three-dimensional space: (a) Grid; (b) Random.

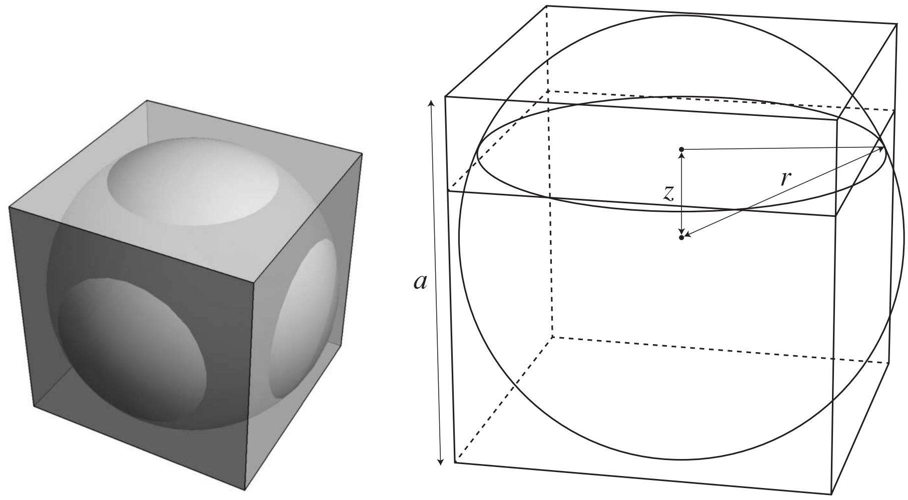

Fig. 2. Region such that $R \leq r$.

where V and $V(r)$ are the volume of the study region and the volume of the region such that $R \leq r$ in the study region, respectively. Since we assume that the grid pattern continues infinitely, the study region can be confined to the region where a point is the nearest. The study region is then given by the cube centered at the point with side length a. The region such that $R \leq r$ is expressed as the ball centered at the point with radius r. $F(r)$ is thus the ratio of the volume of the ball in the cube to the volume of the cube, as shown in Fig. 2. The volume of the cube is $V = a^3$. The volume of the ball in the cube is

$$V(r) = \begin{cases} \frac{4}{3}\pi r^3, & 0 < r \leq \frac{a}{2}, \\ 2\int_0^{a/2} S_1(z)\,\mathrm{d}z, & \frac{a}{2} < r \leq \frac{a}{\sqrt{2}}, \\ 2\int_0^{a/2} S_2(z)\,\mathrm{d}z, & \frac{a}{\sqrt{2}} < r \leq \frac{\sqrt{3}}{2}a, \end{cases} \tag{2}$$

where $S_1(z)$ and $S_2(z)$ are the cross sectional areas of the ball in the cube expressed as

$$S_1(z) = \begin{cases} \pi(r^2 - z^2) + a\sqrt{4(r^2 - z^2) - a^2} \\ -4(r^2 - z^2)\arccos\frac{a}{2\sqrt{r^2 - z^2}}, \\ \qquad 0 < z \leq \sqrt{r^2 - \frac{a^2}{4}}, \\ \pi(r^2 - z^2), \\ \qquad \sqrt{r^2 - \frac{a^2}{4}} < z \leq \frac{a}{2}, \end{cases} \tag{3}$$

$$S_2(z) = \begin{cases} a^2, \\ \qquad 0 < z \leq \sqrt{r^2 - \frac{a^2}{2}}, \\ \pi(r^2 - z^2) + a\sqrt{4(r^2 - z^2) - a^2} \\ -4(r^2 - z^2)\arccos\frac{a}{2\sqrt{r^2 - z^2}}, \\ \qquad \sqrt{r^2 - \frac{a^2}{2}} < z \leq \frac{a}{2}, \end{cases} \tag{4}$$

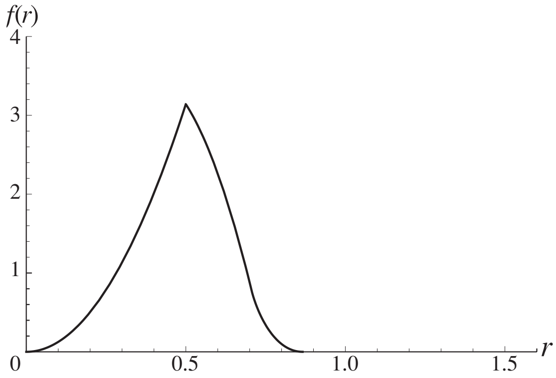

Fig. 3. Distribution of the Euclidean nearest neighbor distance for the grid pattern.

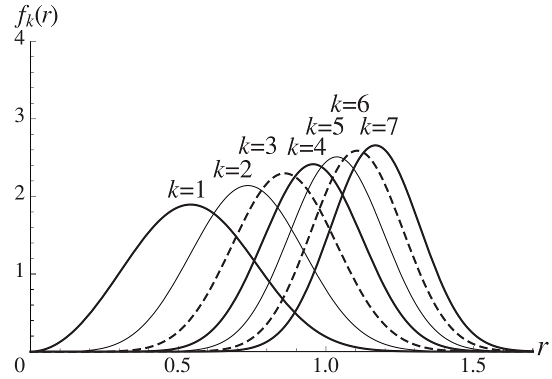

Fig. 4. Distribution of the Euclidean kth nearest neighbor distance for the random pattern.

respectively. Substituting V, $V(r)$, and $a = \rho^{1/3}$ into Eq. (1) and differentiating with respect to r yield the probability density function of R, which we call the distribution of the nearest neighbor distance. The distribution of the nearest neighbor distance $f(r)$ for the grid pattern is shown in Fig. 3, where the density of points is $\rho = 1$. From $f(r)$, we have the average nearest neighbor distance

$$E(R) = \frac{6\sqrt{3} - \pi + 2\mathrm{arcosh}(1351)}{48\rho^{1/3}} \approx \frac{0.480}{\rho^{1/3}}. \quad (5)$$

Note that the average distance is inversely proportional to the cube root of the density.

2.2 Random pattern

Suppose that points are randomly distributed with density ρ. If the distance from a location to a point is less than or equal to r, the ball centered at the location with radius r contains the point. The cumulative distribution function $F(r)$ is thus the probability that the ball with radius r contains at least one point. The probability that a region of volume V contains exactly x points, denoted by $P(x, V)$, is given by the Poisson distribution

$$P(x, V) = \frac{(\rho V)^x}{x!} \exp(-\rho V). \quad (6)$$

Since the volume of the ball is $4\pi r^3/3$, we have

$$\begin{aligned} F(r) &= 1 - P\left(0, \frac{4}{3}\pi r^3\right) \\ &= 1 - \exp\left(-\frac{4}{3}\rho\pi r^3\right). \end{aligned} \quad (7)$$

Differentiating $F(r)$ with respect to r yields the distribution of the nearest neighbor distance

$$f(r) = 4\rho\pi r^2 \exp\left(-\frac{4}{3}\rho\pi r^3\right). \quad (8)$$

The average nearest neighbor distance is

$$E(R) = \left(\frac{3}{4\rho\pi}\right)^{1/3} \Gamma\left(\frac{4}{3}\right) \approx \frac{0.554}{\rho^{1/3}}. \quad (9)$$

As expected, the average distance for the random pattern is greater than that for the grid pattern.

The distribution of the distance to the kth nearest point can be similarly obtained for the random pattern. The cumulative distribution function of the kth nearest neighbor distance, denoted by $F_k(r)$, is the probability that the ball with radius r contains at least k points. Using the Poisson distribution (6), we have

$$\begin{aligned} F_k(r) &= 1 - \sum_{x=0}^{k-1} P\left(x, \frac{4}{3}\pi r^3\right) \\ &= 1 - \sum_{x=0}^{k-1} \frac{(4\rho\pi r^3/3)^x}{x!} \exp\left(-\frac{4}{3}\rho\pi r^3\right). \end{aligned} \quad (10)$$

Differentiating $F_k(r)$ with respect to r yields the distribution of the kth nearest neighbor distance

$$f_k(r) = \frac{3(4\rho\pi r^3/3)^k}{r(k-1)!} \exp\left(-\frac{4}{3}\rho\pi r^3\right). \quad (11)$$

Figure 4 shows $f_k(r)$ for $k = 1, 2, \ldots, 7$. It can be seen that the range of the nearest neighbor distance distribution for the random pattern is greater than that for the grid pattern. The average kth nearest neighbor distance is

$$E(R_k) = \frac{1}{(k-1)!} \left(\frac{3}{4\rho\pi}\right)^{1/3} \Gamma\left(\frac{1}{3} + k\right). \quad (12)$$

3. Rectilinear Distance

The rectilinear distance has frequently been used as an approximation for the travel distance in a building (Koshizuka, 1996; Savaş *et al.*, 2002; Sarkar *et al.*, 2007). The rectilinear distance is thus suited to facility location problems in a building. The rectilinear distance between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) is defined as $|x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|$. Let R be the rectilinear distance from a randomly selected location in three-dimensional space to the nearest point. In this section, we derive the distribution of the rectilinear nearest neighbor distance for the grid and random patterns.

3.1 Grid pattern

The region such that $R \leq r$ is expressed as the octahedron centered at a point with radius r. The cumulative distribution function $F(r)$ is thus the ratio of the volume of the octahedron in the cube to the volume of the cube, as shown

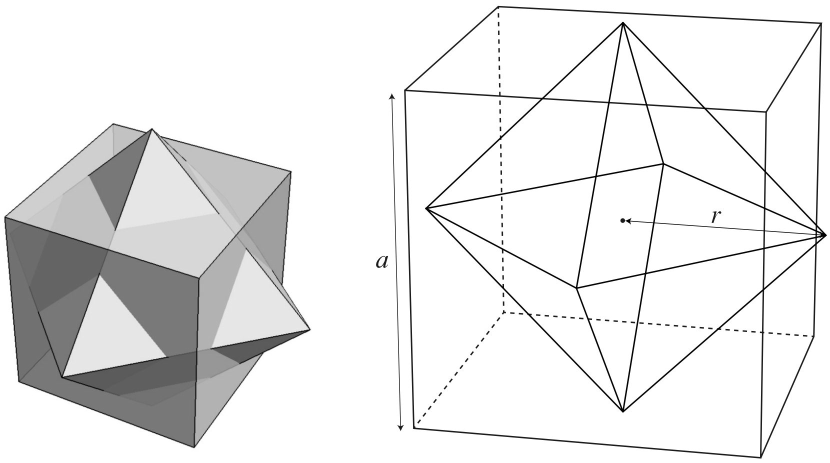

Fig. 5. Region such that $R \le r$.

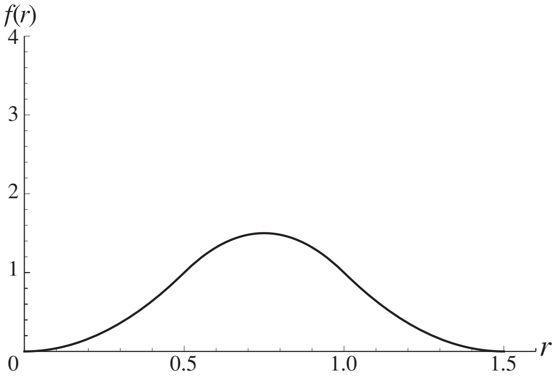

Fig. 6. Distribution of the rectilinear nearest neighbor distance for the grid pattern.

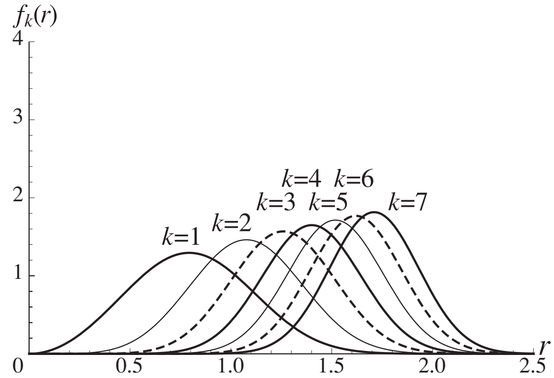

Fig. 7. Distribution of the rectilinear kth nearest neighbor distance for the random pattern.

in Fig. 5. The volume of the octahedron in the cube is

$$V(r) = \begin{cases} \frac{4}{3}r^3, & 0 < r \le \frac{a}{2}, \\ \frac{4}{3}r^3 - 4\left(r - \frac{a}{2}\right)^3, & \frac{a}{2} < r \le a, \\ a^3 - \frac{4}{3}\left(\frac{3}{2}a - r\right)^3, & a < r \le \frac{3}{2}a. \end{cases} \tag{13}$$

Substituting V, $V(r)$, and $a = \rho^{1/3}$ into Eq. (1) and differentiating with respect to r yield the distribution of the nearest neighbor distance

$$f(r) = \begin{cases} 4\rho r^2, & 0 < r \le \frac{1}{2\rho^{1/3}}, \\ -8\rho r^2 + 12\rho^{2/3}r - 3\rho^{1/3}, & \frac{1}{2\rho^{1/3}} < r \le \frac{1}{\rho^{1/3}}, \\ 4\rho r^2 - 12\rho^{2/3}r + 9\rho^{1/3}, & \frac{1}{\rho^{1/3}} < r \le \frac{3}{2\rho^{1/3}}. \end{cases}$$

$f(r)$ for the grid pattern is shown in Fig. 6. The average nearest neighbor distance is

$$E(R) = \frac{3}{4\rho^{1/3}} \approx \frac{0.750}{\rho^{1/3}}. \tag{14}$$

The average rectilinear distance for the grid pattern is greater than the average Euclidean distance for the random pattern as well as that for the grid pattern.

3.2 Random pattern

The cumulative distribution function $F(r)$ is the probability that the octahedron with radius r contains at least one point. The volume of the octahedron is $4r^3/3$. Using the Poisson distribution (6), we have

$$\begin{aligned} F(r) &= 1 - P\left(0, \frac{4}{3}r^3\right) \\ &= 1 - \exp\left(-\frac{4}{3}\rho r^3\right). \end{aligned} \tag{15}$$

Differentiating $F(r)$ with respect to r yields the distribution of the nearest neighbor distance

$$f(r) = 4\rho r^2 \exp\left(-\frac{4}{3}\rho r^3\right). \tag{16}$$

The average nearest neighbor distance is

$$E(R) = \left(\frac{3}{4\rho}\right)^{1/3} \Gamma\left(\frac{4}{3}\right) \approx \frac{0.811}{\rho^{1/3}}. \tag{17}$$

The cumulative distribution function of the kth nearest neighbor distance is the probability that the octahedron with radius r contains at least k points. Using the Poisson distribution (6), we have

$$F_k(r) = 1 - \sum_{x=0}^{k-1} P\left(x, \frac{4}{3}r^3\right) = 1 - \sum_{x=0}^{k-1} \frac{(4\rho r^3/3)^x}{x!} \exp\left(-\frac{4}{3}\rho r^3\right). \quad (18)$$

Differentiating $F_k(r)$ with respect to r yields the distribution of the kth nearest neighbor distance

$$f_k(r) = \frac{3(4\rho r^3/3)^k}{r(k-1)!} \exp\left(-\frac{4}{3}\rho r^3\right). \quad (19)$$

Figure 7 shows $f_k(r)$ for $k = 1, 2, \ldots, 7$. The average kth nearest neighbor distance is

$$E(R_k) = \frac{1}{(k-1)!}\left(\frac{3}{4\rho}\right)^{1/3} \Gamma\left(\frac{1}{3} + k\right). \quad (20)$$

The average rectilinear distance is $\pi^{1/3}(\approx 1.465)$ times as large as the average Euclidean distance (see Eq.(12)).

4. Conclusions

This paper has extended the nearest neighbor distance to three-dimension. The distribution of the nearest neighbor distance has been derived for the grid and random patterns in three-dimensional space. The analytical expressions for the distribution are useful for facility location problems in three-dimensional space as follows. First, they give an estimate for the service level of facility location and can be used to evaluate actual patterns. Second, they demonstrate how the density of facilities affects the nearest neighbor distance. This relationship helps planners to determine the number of facilities required to achieve a certain level of service. Finally, they provide all the information about the nearest neighbor distance. The average and standard deviation of the distance are obtained from the distribution. The average can be used as a criterion of efficiency, whereas the standard deviation can be used as a criterion of equity.

The distance to the kth nearest point is important particularly for location problems with closing of facilities (Miyagawa, 2009). Examining the kth nearest distance for regular patterns in three-dimensional space is an interesting topic for future research.

Acknowledgments. I am grateful to anonymous reviewers for their helpful comments and suggestions.

References

Clark, P. and Evans, F. (1954) Distance to nearest neighbor as a measure of spatial relationships in populations, *Ecology*, **35**, 85–90.

Cooper, L. (1973) N-dimensional location models: An application to cluster analysis, *Journal of Regional Science*, **13**, 41–54.

Cressie, N. (1993) *Statistics for Spatial Data*, John Wiley & Sons, New York.

Haenggi, M. (2005) On distances in uniformly random networks, *IEEE Transactions on Information Theory*, **51**, 3584–3586.

Holgate, P. (1965) The distance from a random point to the nearest point of a closely packed lattice, *Biometrika*, **52**, 261–263.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008) *Statistical Analysis and Modeling of Spatial Point Patterns*, John Wiley & Sons, Chichester.

Kon, M. (2001) Efficient solutions of multicriteria location problems with rectilinear norm in $\mathbb{R}^3$, *Scientiae Mathematicae Japonicae*, **54**, 289–299.

Kon, M. (2007) Quasiefficient solutions of multicriteria location problems with rectilinear norm in $\mathbb{R}^3$, *Journal of the Operations Research Society of Japan*, **50**, 264–275.

Koshizuka, T. (1996) Comparison between low and high buildings with respect to travel distance, *Journal of the City Planning Institute of Japan*, **31**, 31–36 (in Japanese).

Koshizuka, T. and Ohsawa, Y. (1983) Probability density function of the distance in urban facility planning, *Journal of the City Planning Institute of Japan*, **18**, 25–30 (in Japanese).

Larson, R. and Odoni, A. (1981) *Urban Operations Research*, Prentice Hall, Englewood Cliffs, NJ.

Love, R. (1969) Locating facilities in three-dimensional space by convex programming, *Naval Research Logistics Quarterly*, **16**, 503–516.

Mathai, A. (1999) *An Introduction to Geometrical Probability: Distributional Aspects with Applications*, Gordon and Breach Science Publishers, Amsterdam.

Miyagawa, M. (2008) Analysis of facility location using ordered rectilinear distance in regular point patterns, *FORMA*, **23**, 89–95.

Miyagawa, M. (2009) Order distance in regular point patterns, *Geographical Analysis*, **41**, 252–262.

O'Kelly, M. (2009) Rectilinear minimax hub location problems, *Journal of Geographical Systems*, **11**, 227–241.

Persson, O. (1964) Distance methods: The use of distance measurements in the estimation of seedling density and open space frequency, *Studia Forestalia Suecica*, **15**, 1–68.

Radó, F. (1988) The Euclidean multifacility location problem, *Operations Research*, **36**, 485–492.

Sarkar, A., Batta, R., and Nagi, R. (2007) Placing a finite size facility with a center objective on a rectangular plane with barriers, *European Journal of Operational Research*, **179**, 1160–1176.

Savaş, S., Batta, R., and Nagi, R. (2002) Finite-size facility placement in the presence of barriers to rectilinear travel, *Operations Research*, **50**, 1018–1031.

Schöbel, A. and Scholz, D. (2010) The big cube small cube solution method for multidimensional facility location problems, *Computers & Operations Research*, **37**, 115–122.

Thill, J.-C., Dao, T., and Zhou, Y. (2011) Traveling in the three-dimensional city: applications in route planning, accessibility assessment, location analysis and beyond, *Journal of Transport Geography*, **19**, 405–421.