



Solving continuous location–districting problems with Voronoi diagrams

Antonio G.N. Novaes^{a,*}, J.E. Souza de Cursi^b, Arinei C.L. da Silva^c, João C. Souza^a

^aFederal University of Santa Catarina, CP 476, Florianópolis, SC 88010-970, Brazil

^bINSA—Rouen, Av de l'Université BP8, 76801, Saint Etienne du Rouvray CEDEX, France

^cFederal University of Paraná, CP 019081, Curitiba, PR 81531-990, Brazil

Abstract

Facility location problems are frequent in OR literature. In districting problems, on the other hand, the aim is to partition a territory into smaller units, called districts or zones, while an objective function is optimized and some constraints are satisfied, such as balance, contiguity, and compactness. Although many location and districting problems have been treated by assuming the region previously partitioned into a large number of elemental areas and further aggregating these units into districts with the aid of a mathematical programming model, continuous approximation, on the other hand, is based on the spatial density of demand, rather than on precise information on every elementary unit. Voronoi diagrams can be successfully used in association with continuous approximation models to solve location–districting problems, specially transportation and logistics applications. We discuss in the paper the context in which approximation algorithms can be used to solve this kind of problem.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Facility location; Transportation districting; Logistics districting; Voronoi diagrams

1. Introduction

Facility location problems are concerned with the location of one or more facilities such as to optimize a certain objective function as, for example, minimize transportation cost, provide equitable service to customers, capture the largest market share, etc. [1–3]. Facility location problems are often related to geometrical and combinatorial problems. The literature shows a large number of facility location models and applications. Locating rapid transit stations, hospitals, schools, retail outlets, warehouses, are classical examples. In a typical location problem a set of customers that are spatially distributed over a geographical area originates demand for some kind of goods or services. Customer demand must be supplied by one or more facilities. The optimization process must establish where to locate the facilities taking into account users requirements and a number of constraints that can be of behavioral, economical, operational, or geographical nature.

Location problems are called continuous when the underlying space, both for facility sites and demand points, are determined by one or more variables (depending on the dimension of the problem) that will vary continuously.

* Corresponding author. Tel./fax: +55 48 32320409.

E-mail addresses: novaes@deps.ufsc.br (A.G.N. Novaes), souza@insa-rouen.fr (J.E. Souza de Cursi), lindbeck@mat.ufpr.br (A.C.L. da Silva), jcsouza@arq.ufsc.br (J.C. Souza).

The corresponding models are mainly of geometrical nature and the optimal solution of such models is sought with linear/non-linear programming and global optimization algorithms, as opposed to discrete, combinatorial and/or integer programming. More recently, the use of computational geometry and computer graphics has opened the way for other solution techniques, including approximate heuristics [4].

The aim of districting problems, on the other hand, is to get an optimized partition of a territory into smaller units, called districts or zones, subject to some side constraints [5–7]. The constraints reflect a number of common sense criteria. One of them is to balance demand among districts. Furthermore, the resulting districts must be contiguous and geographically compact [8]. Logistics and transportation districting problems usually involve additional optimization criteria and constraints. In general, apart from the basic balance, contiguity, and compactness principles, there is not a set of general criteria that are common to all districting problems.

Districting problems are associated with a number of practical applications. Political districting, in which one is interested in drawing of electoral district boundaries has received much attention in the literature [5–8]. School districting [9] and police districting [10] are two other areas of research interest. In addition, the literature presents articles on the design of sales territory [11,12], as well as emergency, health-care, and logistics districting. Among the latter, we mention the districting approach to the planning of salt spreading operations on roads [13], the balanced allocation of customers to distribution centers [14], and the design of multi-vehicle delivery tours [15–17].

Several heuristics have been proposed in the literature for the discrete districting problem. Generally, a binary linear programming is formulated in order to assign a basic areal unit i to a “seed” district center j [6,7]. The cost function to be minimized is related to the distance from i to j (the square of the distance is often used). A binary variable x_{ij} is set equal to 1 if and only if unit i is assigned to seed j , being zero otherwise. Constraints to guarantee that every unit is included in exactly one of the selected districts are incorporated into the model. Although there is no uniformly acceptable mathematical definition of compactness, a penalty cost constraint is usually introduced that penalizes the “non-compactness” of the potential districts [6]. Contiguity is attained when it is possible to travel from any point in the district to any other point in the same district without having to go through any other district. Although one could add linear inequalities into the model to enforce contiguity precisely, this would require an exponential number of constraints and would be very demanding computationally [6].

Discrete districting models are normally large and are based on diverse operations research techniques. The mathematical representation of balance, compactness, and contiguity criteria also varies widely from case to case. Continuous approximation, on the other hand, is based on the spatial density and distribution of the demand rather than on precise information on every demand unit. It allows for simple, yet robust models that are useful when planning a new service or the expansion of an existing one [17–19]. The association of continuous approximation techniques with Voronoi diagram districting opens the way to solve a number of real-life problems. In particular, the use of ordinary and non-ordinary Voronoi diagrams to solve logistics and transportation problems has been reported in the literature [20–22]. Boots and South [12] used a multiplicatively weighted Voronoi diagram approach for modeling retail trade areas. Galvão et al. [22] defined a multiplicatively weighted Voronoi diagram model to solve an urban freight distribution problem. The utilization of non-ordinary Voronoi diagrams in logistics and transportation location/districting problems, associated with a continuous demand approach, also allows for the introduction of physical barriers into the model [23]. This is an important property because it permits to treat problems with obstacles imposed by thoroughfares, highways, rivers, reservoirs, hills, etc. (see Section 4.2).

The purpose of this paper is to develop two continuous location–districting models applied to transportation and logistics problems, combining a Voronoi diagram approach with an optimization algorithm. The contributions are threefold. First, the use of non-ordinary Voronoi diagram concepts in association with transportation and logistics districting, which bypasses some difficulties encountered in discrete districting algorithms, such as problem size and the mathematical representation of compactness and contiguity constraints. Second, the utilization of recent-developed Voronoi diagram construction methods such as the plane-sweep and the quad-tree techniques, which have been reported in the literature in relation to computer graphics, robotics, etc., but not yet used in conjunction with transportation and logistics problems. And third, the extension of the Voronoi diagram methodology to solve logistics districting problems with spatial barriers.

Section 2 presents a summary of the basic concepts and properties of Voronoi diagrams that are of interest to solve transportation and logistics districting problems. Section 3 deals with computational aspects associated with the solution of Voronoi diagrams. In Section 4, two applications involving Voronoi diagrams and locational optimization are analyzed. The first is concerned with the location of stations over a rail-transit line in order to maximize rail patronage.

The second is dedicated to analyze and discuss a Voronoi diagram application to a logistics distribution problem with geographical barriers.

2. Voronoi diagrams

It is presented in this section a summary of the concepts, types, and properties of Voronoi diagrams that will further be used to solve logistics and transportation problems. Voronoi diagrams comprise a vast subject, and the reader is referred to Aurenhammer [24] and Okabe et al. [20] for more details. Voronoi diagrams are defined in \mathbb{R}^d , with $d \geq 2$, but our applications are limited to \mathbb{R}^2 . Although Voronoi diagram generators can be points, lines, circles, or areas of diverse shapes [20], our applications deal with point generators only. In fact, logistics and transportation problems usually involve “point-like facilities” such as depots, subway stations, bus stops, etc.

Voronoi diagrams have been used extensively in a variety of disciplines, including Astronomy, Physics, Meteorology, Urban Planning, and Engineering. Presently, Voronoi diagrams are extensively used in computational geometry, computer graphics, robotics, pattern recognition, games, etc. [24,25]. The basic concept of Voronoi diagrams is quite simple: given a finite set of distinct and isolated points in a continuous space, we associate all locations in that space with the closest—in the sense of a given distance—member of the point set [20]. In mathematical terms, let m be an integer such that $2 \leq m < \infty$, $P \equiv \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_m\}$ be a finite set of m distinct points in the two-dimensional Cartesian space \mathbb{R}^2 and $\mu \equiv \{\mu_1, \mu_2, \dots, \mu_m\}$ be a family of continuous functions: $\mu_i : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$, for $i \in \{1, \dots, m\}$. P is called a *generator set*, and, for each index $i \in \{1, \dots, m\}$, \mathbf{P}_i is a *generator point*. The symbol μ_i is usually referred as the *metric associated to \mathbf{P}_i or the distance to \mathbf{P}_i* , although—in general—it may take negative values. Let the region $V(\mathbf{P}_i)$ be the set of locations $\mathbf{X} \in \mathbb{R}^2$ such that

$$V(\mathbf{P}_i) = \{\mathbf{X} \in \mathbb{R}^2 \mid \mu_i(\mathbf{X}, \mathbf{P}_i) \leq \mu_j(\mathbf{X}, \mathbf{P}_j), \quad j = 1, \dots, m\}. \quad (1)$$

Region $V(\mathbf{P}_i)$ is called the *Voronoi region of \mathbf{P}_i associated to the family μ* . The *dominance region of \mathbf{P}_i over \mathbf{P}_j associated to the family μ* is given by

$$Dom(\mathbf{P}_i, \mathbf{P}_j) = \{\mathbf{X} \in \mathbb{R}^2 \mid \mu_i(\mathbf{X}, \mathbf{P}_i) \leq \mu_j(\mathbf{X}, \mathbf{P}_j)\}. \quad (2)$$

And the *separator between \mathbf{P}_i and \mathbf{P}_j* is

$$sep(\mathbf{P}_i, \mathbf{P}_j) = \{\mathbf{X} \in \mathbb{R}^2 \mid \mu_i(\mathbf{X}, \mathbf{P}_i) = \mu_j(\mathbf{X}, \mathbf{P}_j)\}. \quad (3)$$

We have

$$sep(\mathbf{P}_i, \mathbf{P}_j) = Dom(\mathbf{P}_i, \mathbf{P}_j) \cap Dom(\mathbf{P}_j, \mathbf{P}_i), \quad (4)$$

$$V(\mathbf{P}_i) = \bigcap_{j=1}^m Dom(\mathbf{P}_i, \mathbf{P}_j) \quad \text{and} \quad \mathbb{R}^2 = \bigcup_{i=1}^m V(\mathbf{P}_i). \quad (5)$$

Thus, the reunion of all the Voronoi regions covers \mathbb{R}^2 and defines a partition of the space $V = \{V(\mathbf{P}_1), \dots, V(\mathbf{P}_m)\}$, which is called the *Voronoi diagram generated by P and μ* .

A classical example of Voronoi diagrams is represented by $\mu_i(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|$, where $\|\bullet\|$ is the Euclidean norm of \mathbb{R}^2 . In this case, the dominance region is called the *ordinary Voronoi polygon* associated with \mathbf{P}_i , and the partition V is the *planar ordinary Voronoi diagram* generated by P . The edges of Voronoi polygons in \mathbb{R}^2 are line segments.

Another classical example concerns the p -center problem, where we seek a set of facilities which minimizes the maximum distance from a user to its nearest facility [21]. Associated with this type of problem is the farthest-point Voronoi diagram, which corresponds to $\mu_i(\mathbf{X}, \mathbf{Y}) = -\|\mathbf{X} - \mathbf{Y}\|$.

If we are not interested in the whole space but only in a subregion $R \subset \mathbb{R}^2$, analogous definitions may be introduced by replacing \mathbb{R}^2 by R . For instance, when the planar Voronoi diagram is constrained to a bounded region, it becomes a bounded planar Voronoi diagram, as shown in Fig. 1a.

There are situations where the Euclidean distance does not represent well the attracting process. For instance, suppose that the six generator points exhibited in Fig. 1a are retail stores selling the same kind of product. Assume further that, in addition to distance, the attraction of such stores depends on a set of features, leading to the weighting coefficients

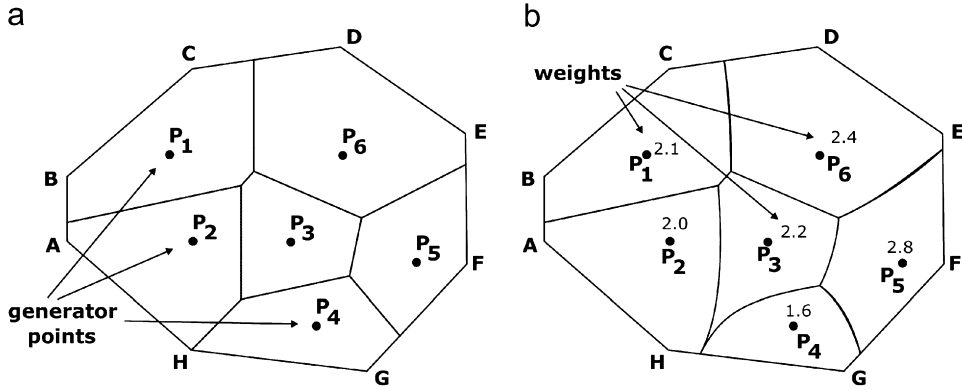


Fig. 1. Ordinary and multiplicatively weighted Voronoi diagrams: (a) ordinary Voronoi diagram; (b) multiplicatively weighted Voronoi diagram.

shown in Fig. 1b. In order to take these elements into account, several kinds of *weighted planar Voronoi diagrams* have been introduced. These diagrams use a family of weights $w = (w_1, w_2, \dots, w_m)$, such that the dominance region (2) increases with the weight w_i . For instance, *multiplicatively weighted planar Voronoi diagrams* correspond to

$$\mu_i(\mathbf{X}, \mathbf{P}_i) = \frac{1}{w_i} \|\mathbf{X} - \mathbf{P}_i\|, \quad (6)$$

where w is a family of strictly positive weights. In the case with only two generator points, the locus of the points \mathbf{X} satisfying (6) is the Apollonius circle [20], except if $w_1 = w_2$, when the bisector becomes a straight line. Fig. 1b shows an example of MW-Voronoi diagram for the weights there indicated. In general, a MW-Voronoi region is a non-empty set and need not be convex, or connected; and it may have holes [20].

Analogously, the *additively weighted Voronoi diagram* is represented by

$$\mu_i(\mathbf{X}, \mathbf{P}_i) = \|\mathbf{X} - \mathbf{P}_i\| - w_i. \quad (7)$$

Here, the sign of w_i is not restricted. Taking two points \mathbf{P}_i and \mathbf{P}_j , the bisector associated with the additively weighted Voronoi diagram is either a branch of a hyperbolic curve with foci \mathbf{P}_i and \mathbf{P}_j , or a straight line segment [20].

The combination between additive and multiplicative weights leads to the *compoundly weighted Voronoi diagram*, which is associated to

$$\mu_i(\mathbf{X}, \mathbf{P}_i) = \frac{1}{w_{i1}} \|\mathbf{X} - \mathbf{P}_i\| - w_{i2}. \quad (8)$$

We recall that the sign of w_{i2} is not restricted. In this case, the boundary of the dominance region is a fourth-order polynomial function, and its shape is fairly complex [20].

The *power Voronoi diagram* corresponds to

$$\mu_i(\mathbf{X}, \mathbf{P}_i) = \|\mathbf{X} - \mathbf{P}_i\|^2 - w_i. \quad (9)$$

In this case, only positive values of w_i are usually used. The line segment connecting \mathbf{P}_i and \mathbf{P}_j is

$$[\mathbf{P}_i, \mathbf{P}_j] = \{\mathbf{X} \in \mathbb{R}^2 : \mathbf{X} = \theta \mathbf{P}_i + (1 - \theta) \mathbf{P}_j, 0 \leq \theta \leq 1\}. \quad (10)$$

The bisector is a straight line perpendicular to the line segment $\mathbf{P}_j - \mathbf{P}_i$ passing through the point \mathbf{X}_{ij}^* given by [20]

$$\mathbf{X}_{ij}^* = \frac{\|\mathbf{P}_j\|^2 - \|\mathbf{P}_i\|^2 + w_i - w_j}{2\|\mathbf{P}_j - \mathbf{P}_i\|^2} (\mathbf{P}_j - \mathbf{P}_i) \quad (11)$$

and is well-behaving and regular. An important property of power Voronoi diagrams, useful in applications, is that the resulting Voronoi polygons are always convex. Power Voronoi diagrams are specially useful to solve districting problems with barriers, as in the case described in Section 4.2.

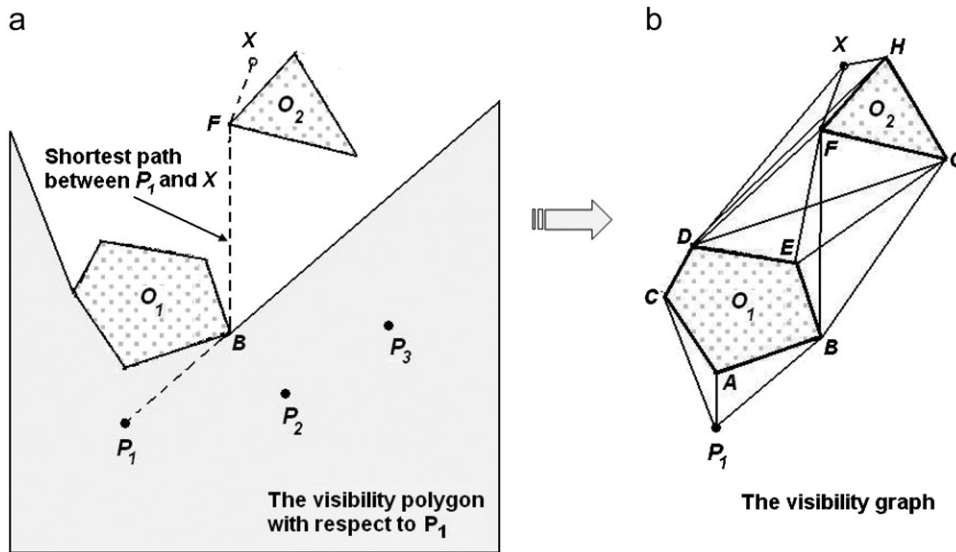


Fig. 2. The shortest-path visibility graph.

When solving Voronoi diagram problems with obstacles, the Euclidean norm is not acceptable. If an obstacle lies on the line linking an origin and a destination, it is not possible to traverse it straight. Instead, a detour around the obstacle must be taken. Following [20], consider a generator set P and a set of c closed regions $O = \{O_1, \dots, O_c\}$ ($1 \leq c < \infty$). The set O represents a set of obstacles that are not traversable. These obstacles are assumed not to overlap each other and points of P are not allowed to lie within the obstacles (Fig. 2a). Furthermore, each obstacle is assumed to be connected and with no holes. For computational convenience it is assumed that O_i ($i = 1, \dots, c$) is a polygon, but it is not assumed that O_i is necessarily convex. Line segments are also accepted as obstacles.

The visibility-shortest-path distance between a generic point X and a generator point P_i , expressed as $d_{SP}(X, P_i)$, is obtained considering all possible continuous paths connecting X and P_i that do not traverse obstacles. A *visibility polygon* with respect to P_i , and denoted by $Vis(P_i)$, is the set of points that are visible from P_i . Mathematically, $Vis(P_i)$ is defined by

$$Vis(P_i) = \left\{ X \in \mathbb{R}^2 : [X, P_i] \cap \left(\bigcup_{j=1}^c O_j \right) = \emptyset \right\}. \quad (12)$$

An example is shown in Fig. 2a, where the visibility polygon with respect to the generator point P_1 is indicated by the hatched area. To compute the visibility-shortest-path distance between a point X and a generator point P_i one uses the correspondent *visibility graph*, which is formed by all possible paths connecting X and P_i (Fig. 2b). On the visibility graph one solves the classical shortest-path problem with the aid of an appropriate algorithm such as, for example, the well-known Dijkstra method. For the example of Fig. 2b, the shortest-path between P_1 and X is $P_1 \rightarrow B \rightarrow F \rightarrow X$.

When applying ordinary Voronoi diagrams to logistics and transportation problems, the computation is really simple but, apart from some specific cases, the resulting model framework is not realistic enough. This leads to the employment of non-ordinary Voronoi diagrams, such as multiplicatively weighted, additively weighted, compound, power, and other Voronoi diagram types. The utilization of multiplicatively weighted, additively weighted and compoundly weighted Voronoi diagrams in logistics districting problems, for example, has some advantages [22]. One of them is that the fitting process leads to more equalized load factors among the districts, meaning the vehicles assigned to the zones will show more balanced utilization levels [22]. This happens because those Voronoi diagrams have more degrees of freedom when searching for the district contours, as compared to the traditional, theoretical, wedge-shaped partitioning scheme [17,22]. But, in some cases the Voronoi region is not convex, or connected, and it may have holes [20], and thus special care must be taken when developing computational iterative models to solve this type of problem. Otherwise the computing process may not converge or it may produce non-realistic results. In order to cope with this kind of problem

Galvão et al. [22], for example, departed from a previous solution obtained via a traditional method. With this approach drastic transitions in the iterative routine were avoided, thus maintaining a smooth and convergent computational process.

The locational problems analyzed in this text are of the weighted and power Voronoi-diagram types, but other types of Voronoi diagram problems reported in the literature can equally be solved with the computational approach described in this text [20,21,26]. One of them is the Voronoi fitting problem [20,26], in which one is interested in determining the divergences between cells of an existing tessellation and their corresponding Voronoi districts. This approach can be used, for example, to identify the areas in which customers cannot be served by the nearest facility, say a school or a hospital. Another case is the locational optimization of hierarchical facilities [20,26] in which facilities (as hospitals, for example) are ranked according to the level of services they are allowed to provide.

3. Computational aspects associated with generalized Voronoi diagrams

Computational geometry has attracted enormous research interest in the last few years, covering diverse areas of computer science, such as computer graphics, computer-aided design, robotics, pattern recognition, etc. In particular, readers familiar with the literature on these subjects will have noticed an increasing interest in geometrical constructs represented by Voronoi diagrams [24]. The usual approaches for constructing generalized Voronoi-diagrams are combinatorial methods, incremental techniques, divide-and-conquer methods, and approximation algorithms [20,21,24,25,27,28]. In this section it is briefly described the utilization of two approximation algorithms associated with the construction of generalized Voronoi diagrams, namely the *plane-sweep technique* [27,24] and the *quadtrees technique* [28], which were used in our applications to logistics and transportation problems.

3.1. The plane-sweep technique

Such a technique, due to Fortune [27], stands out by its conceptual and computational simplicity. Let P denote a generator set of m distinct generator points in \mathbb{R}^2 and let V be the corresponding planar Voronoi diagram. A continuous deformation of V is performed in which the Euclidean distance function with respect to the generators points is interpreted in the following way. With each generator point $\mathbf{P}_i \in P$, a cone

$$K(\mathbf{P}_i) = \{(\mathbf{X}, z) \in \mathbb{R}^2 \times \mathbb{R} : \mathbf{X} \in \mathbb{R}^2, z = \|\mathbf{X} - \mathbf{P}_i\|\} \quad (13)$$

is associated [24]. $K(\mathbf{P}_i)$ is an upwardly directed cone in \mathbb{R}^3 with vertical axis of revolution, with apex \mathbf{P}_i , and with an interior angle of $\pi/2$. Cones may be viewed as bivariate functions on \mathbb{R}^2 . A lower envelope of $\bigcup_{i=1}^m K(\mathbf{P}_i)$ is defined as the point wise minimum of these functions or, equivalently, as the surface composed of that portion of each cone that lies below all other cones. From the definition of $K(\mathbf{P}_i)$ it is evident that the vertical projection of this lower envelope—i.e., its projection onto \mathbb{R}^2 —is V .

However, according to Aurenhammer [24], it is more convenient to introduce a different projection: let us introduce a mapping associating to each point $\mathbf{X} = (x_1, x_2) \in V(\mathbf{P}_i)$ a new point $\mathbf{X}^* = (x_1^*, x_2^*) \in \mathbb{R}^2$, given by

$$x_1^* = x_1, \quad x_2^* = x_2 + \|\mathbf{X} - \mathbf{P}_i\|. \quad (14)$$

Clearly, generator points are invariant under the transformation: $\mathbf{P}_i^* = \mathbf{P}_i$. Further, each separator $sep(\mathbf{P}_i, \mathbf{P}_j)$ is transformed into a hyperbola with bottommost point \mathbf{P}_i , if \mathbf{P}_i is below \mathbf{P}_j , and bottommost point \mathbf{P}_j , otherwise. This hyperbola degenerates to a vertical linear segment if \mathbf{P}_i and \mathbf{P}_j have the same x_2 -coordinate. Consequently, the deformed region $[V(\mathbf{P}_i)]^*$ of \mathbf{P}_i is the intersection of hyperbolically bounded half planes of \mathbb{R}^2 . It can be shown that this transformation preserves the topological properties of a Voronoi diagram [24]. The algorithmic advantage of this transformation is that the bottommost point of each region is its defining generator point.

The properties of $V^* = \{[V(\mathbf{P}_1)]^*, [V(\mathbf{P}_2)]^*, \dots, [V(\mathbf{P}_m)]^*\}$ are utilized for the construction of V with the plane-sweep technique. Generally, the technique proceeds as follows. A horizontal line L is swept across the region in which the Voronoi diagram is being constructed from below, by keeping invariant that the portion of the object below L is complete at any point of the computing process. During the plane sweeping, the cross section of L with this object has to be updated at certain critical points. We thus have to handle a one-dimensional dynamic problem instead of a two-dimensional static problem. A Voronoi diagram in the Euclidean plane with m generator points can be

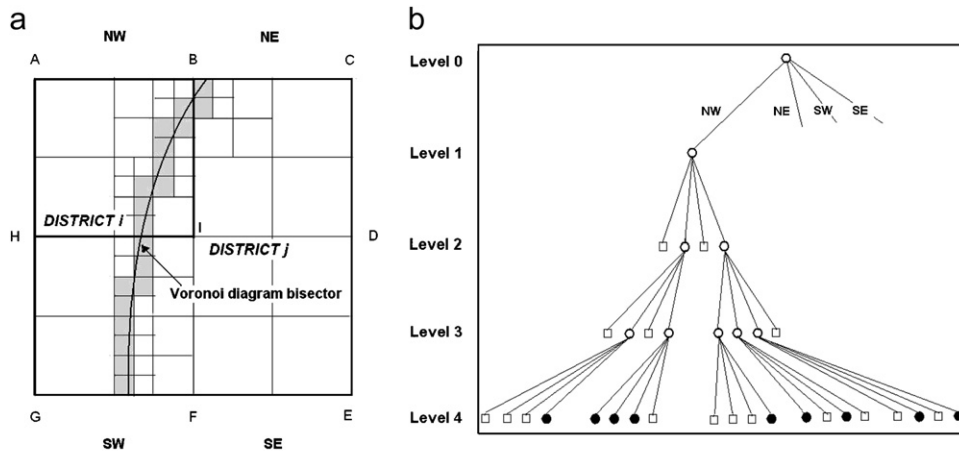


Fig. 3. Quad-tree technique applied to a districting problem.

constructed in time $O(m \log m)$ [24,29]. An interesting practical demonstration of Fortune's method is encountered in www.diku.dk/hjemmesider/studerende/duff/Fortune.

This simple construction method can be extended to other non-ordinary Voronoi diagrams. For instance, it can also be applied to the additively weighted type since the correspondence to lower envelopes extends provided $K(\mathbf{P}_i)$ is translated downward by the weight w_i , for each generator point \mathbf{P}_i . The intersection of $K(\mathbf{P}_i)$ with \mathbb{R}^2 is then a circle with center \mathbf{P}_i and radius w_i , and this situation may be interpreted as the Voronoi diagram with circles generators under Euclidean distance. More details of the plane-sweep technique are encountered in [24]. Generalization to other norms in the plane seems possible, but more general metrics have apparently not yet been investigated [24,29].

3.2. The quad-tree technique

The quad-tree technique [28] is an approach to image representation based on the successive subdivision of the image into quadrants and following a hierarchical data structure. Suppose we want to define the Voronoi diagram bisector which separates two adjacent districts, i and j , as shown in a simplified manner in Fig. 3a. Here the quadrants are squares, but could be rectangles as well. At each stage of the process each node of a quadrant is examined to determine the generator point closest to it (here the distance is measured according to a pre-defined norm). A simple function F_Q is defined as follows: (a) $F_Q = i$, if the four nodes of the quadrant are related to the same generator point i ; (b) $F_Q = 0$ if part of the nodes are related to a generator point i , and the other nodes are related to generator points different from i .

The quad-tree process (Fig. 3a) is represented by a tree of outdegree 4 in which the root represents a father and the four sons represent in order the NW, NE, SW, and SE quadrants [28]. We start with the full image formed by the square ACEGA in Fig. 3a. For quadrant NW (ABIH in Fig. 3a) one has $F_Q = 0$ since nodes A, B, and H are related to the generator point i , while node I is related to the generator point j . Quadrants NE and SW also have $F_Q = 0$, but $F_Q = j$ for quadrant SE. We assume that each quadrant data is stored as a record containing six fields. The first five fields contain the identification of the quadrant's father and of its four sons. The sixth field contains the value of F_Q for the quadrant.

The following rules are established for the quad-tree process:

- The subdivision process of a quadrant terminates when F_Q is of the $F_Q = k$ ($k \neq 0$) type. For example, quadrant SE (square DEFI) in Fig. 3a.
- The subdivision process of a quadrant will continue as long as $F_Q = 0$. The values of F_Q are determined for the four resulting quadrants and criterion (a) is applied again.
- When the quadrant subdivision reaches a predefined level of accuracy the quad-tree process stops.

Fig. 3b shows the encoding sequence of the quad-tree related to the north-western quadrant ABIHA of Fig. 3a. Blank circular nodes represent quadrants of type $F_Q = 0$. Black circular nodes indicate the final cells that form the bisector

under search. In the example, only four levels were examined, and the resulting line representation is quite coarse, requiring further subdivisions. Square nodes represent terminal quadrants of type $F_Q = k$, with $k \neq 0$.

The quad-tree technique leads to substantial reduction in computing time. First, Voronoi diagram bisector points are sparse over region R , and many quadrants will be of the $F_Q = k$ ($k \neq 0$) type at the various levels, thus shortening the subdivision process. Second, after the Voronoi-diagram construction is completed, it is necessary to compute integrals of some variables for each district (Section 3.3), and the resulting quadrants of different sizes facilitate this task.

The quad-tree technique was used to solve the transportation and logistics problems described in Section 4.2.

3.3. Continuous approximation

A number of discrete transportation and logistics problems can be converted into problems involving continuous functions, with good practical results. Continuous demand approximation models are based on the spatial density variables rather than on precise information on every servicing point [18,19]. A number of freight-distribution districting problems have been solved with this approach [17,22,23,30–33]. Before using a specific technique to solve Voronoi diagrams we represent the data in a continuous format, which will be summarized next.

Let us consider a set \mathcal{D} formed by n servicing points of the region R

$$\mathcal{D} = \{Z_i = (x_i, y_i), i = 1, \dots, n\} \subset R. \quad (15)$$

For each $(x, y) \in \mathbb{R}^2$ we define

$$\mathcal{D}^*(x, y) = \{Z_i \in \mathcal{D} | x_i \leq x \text{ and } y_i \leq y\}. \quad (16)$$

Let U be a quantity associated with the servicing points. For instance, when computing the number of points,

$$U(Z) = 1 \quad \text{if } Z \in \mathcal{D}^*(x, y), \quad U(Z) = 0 \quad \text{otherwise.} \quad (17)$$

We introduce a bi-dimensional cumulative function $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that

$$\Psi(x, y) = \sum_{Z_i \in \mathcal{D}^*(x, y)} U(Z_i). \quad (18)$$

Considering Eq. (18), Ψ represents the number of servicing points having the coordinates limited by the upper bounds (x, y) :

$$\Psi(x, y) = \text{card } \mathcal{D}^*(x, y). \quad (19)$$

In the framework of statistical description of data, a cumulative function Ψ_p is usually introduced and gives the relative frequency (probability) of $\mathcal{D}^*(x, y)$. Ψ is analogous to Ψ_p , except that it gives the absolute number of servicing points and not a frequency or a probability.

A density of probability ψ_p may be associated to the cumulative probability function Ψ_p : for a given subset $A \subset \mathbb{R}^2$, the value of Ψ_p associated to A is given by

$$\Psi_p(A) = \int_A \psi_p(x, y) dx dy. \quad (20)$$

In an analogous way, a density ψ may be associated to Ψ . Since set $\mathcal{D}^*(x, y)$ is discrete, Ψ is discontinuous and ψ is a sum of Dirac measures: the value of Ψ associated to A is given by the sum of the values of U for the servicing points of A :

$$\Psi(A) = \int_A \psi(x, y) dx dy = \sum_{Z_i \in A} U(Z_i). \quad (21)$$

In the sequel, we shall introduce a regular approximation $\widehat{\Psi}$ of the function Ψ having suitable mathematical properties of differentiability. Namely, the density $\widehat{\psi}$ associated to $\widehat{\Psi}$, will be a continuous function such that

$$\widehat{\psi}(x, y) = \partial^2 \widehat{\Psi}(x, y) / \partial x \partial y. \quad (22)$$

We approximate

$$\Psi(A) \approx \widehat{\Psi}(A) = \int_A \widehat{\psi}(x, y) dx dy. \quad (23)$$

In the application analyzed in this paper, the variable $\psi(x, y)$ may represent the existence of a point at (x, y) , the demand level at point (x, y) , or other attribute of interest. Methods for the construction of the regular approximations $\widehat{\Psi}$ and $\widehat{\psi}$ may be found in the literature. In this work, the approximation is attained with a bi-quadratic spline [34] combined with a finite element discretization [35] of the region R . Our choice is guided by our final objective: the optimization procedures imply repeated evaluations of the quantities defined by Eq. (23). An useful situation adopted in this paper is to consider quadrilateral finite elements with four nodes (Q_1 finite elements), where each element E is a rectangle R_{ABCD} with vertices $A = (x_{\min}, y_{\min})$, $B = (x_{\min}, y_{\max})$; $C = (x_{\max}, y_{\max})$; and $D = (x_{\max}, y_{\min})$, the nodes being the vertices. For this case

$$\widehat{\Psi}(E) = \widehat{\Psi}(C) - \widehat{\Psi}(B) - \widehat{\Psi}(D) + \widehat{\Psi}(A). \quad (24)$$

Due to the simplicity of Eq. (24), the utilization of Q_1 finite elements saves computational time, and thus it is used in our calculations, but actually the method may be implemented with any kind of finite element mesh, by using the appropriate weights in order to evaluate $\widehat{\Psi}(E)$. For more details, the reader is referred to [22].

4. Applications to transportation and logistics problems

Voronoi diagrams are useful for spatial analysis and locational optimization of diverse transportation problems. In this section we describe and discuss two cases: (1) the problem of locating transit boarding points on an urban region, and (2) an urban freight distribution problem with geographical barriers.

4.1. Defining transit station locations and respective patronage areas with Voronoi diagrams

Rapid transit projects are normally expensive, but they bring large benefits to the users and to the local community in general. Two important aspects regarding rapid transit projects are the optimal location of boarding points or stations and the public patronage attracted to each boarding point considering alternative choices of displacement, subway, bus, walk, or others [36–38]. In this section we present a specific case in which commuters travel from their origin to a CBD terminal, taking either a bus, the subway, or walking to their destinations. The problem is to seek the number of subway stations and their best locations in order to optimize a specific objective function. In our case, the objective is to maximize the subway patronage level. Other objective functions can be chosen as, for example, minimization of average commuter travel time [20, Section 9.2.7], minimization of vehicle fleet size for a given demand level [38], etc.

4.1.1. Problem setting

Let us assume an urban region R (Fig. 4) that is presently served by a bus route converging to a CBD terminal. Apart from the CBD terminal, there are seven bus stops (numbered $1, \dots, 7$) in the served region. A new subway line is under analysis. It will follow a circumference path with radius $r = 3.85$ km, as shown in Fig. 4. With $\Gamma = \{\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_m\}$ representing all the boarding points (bus and subway alike), plus the CBD terminal, the following lexicographic representation is adopted here: (a) $i = 0$ represents the CBD terminal (which attracts people walking directly to it); (b) $i = 1, 2, \dots, b$ representing the bus stops; (c) $i = b + 1, b + 2, \dots, b + s$, where s is the number of subway stations, with $1 + b + s = m$. Thus, the set Γ of generating points splits in three disjoint subsets: $\Gamma = \Gamma_0 \cup \Gamma_b \cup \Gamma_s$, where $\Gamma_0 = \{\mathbf{P}_0\}$, $\Gamma_b = \{\mathbf{P}_1, \dots, \mathbf{P}_b\}$, $\Gamma_s = \{\mathbf{P}_{b+1}, \dots, \mathbf{P}_{b+s}\}$.

When adding a new station along a rail transit line, two contrary effects will influence the possibility of capturing new commuters or loosing existing ones [36,37]. First, the train running time will be increased due to the additional train acceleration and deceleration times, plus the additional stopping time at the new station. For some commuters this additional time may change their modal choice (from subway to bus or to walking). Second, this situation may attract to the rail transit new users located nearby the new station. Thus, changing the number of rail transit stations and their locations, and consequently balancing those contrary effects, one searches for the optimal configuration represented by the maximum subway commuter patronage [36–38].

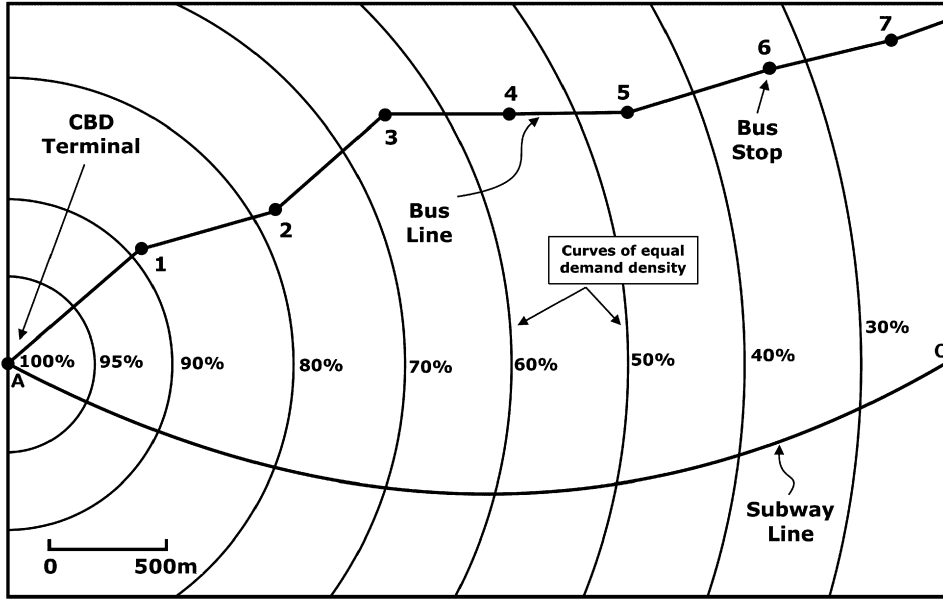


Fig. 4. An urban region served by a bus line and a new subway line.

Let γ_a and γ_d be the acceleration and deceleration rate of the transit train, and let V_C be its posted speed limit [38]. During the acceleration and deceleration phases, the distances covered by the vehicle are, respectively,

$$D^{(a)} = \frac{V_C^2}{2\gamma_a} \quad \text{and} \quad D^{(d)} = \frac{V_C^2}{2\gamma_d}. \quad (25)$$

If $D_{i,i+1}$ is the rail distance between the adjacent stations i and $i+1$, the corresponding subway average vehicle speed in that link is given by

$$\bar{V}_{i,i+1}^{(s)} = \left[(D^{(a)} + D^{(d)}) \frac{V_C}{2} + (D_{i,i+1} - D^{(a)} - D^{(d)}) V_C \right] / D_{i,i+1} \quad (26)$$

if $D^{(a)} + D^{(d)} \leq D_{i,i+1}$ and by

$$\bar{V}_{i,i+1}^{(s)} = \frac{1}{2} \sqrt{\frac{2\gamma_a\gamma_d D_{i,i+1}}{\gamma_a + \gamma_d}} \quad (27)$$

otherwise [38]. For each subway configuration, represented by a number of stations and their locations over the rail line, the average train speed for each link is computed according to (26) or (27). Thus, the in-vehicle traveling time between subway station i and the CBD is given by

$$tv_i^{(s)} = \sum_{j=0}^{i-1} \frac{D_{j,j+1}}{\bar{V}_{j,j+1}^{(s)}} + \sum_{j=1}^i \bar{ST}^{(s)}, \quad (28)$$

where $\bar{ST}^{(s)}$ is the average train stopping time at a station. For the bus service, since the number of stops and their locations are assumed fixed in this problem, and since bus movement is affected by disruption of the traffic flow, we assumed a constant average speed $\bar{V}^{(b)}$, and the in-vehicle traveling time between boarding point i and the CBD is

analogously given by

$$tv_i^{(b)} = \sum_{j=0}^{i-1} \frac{D_{j,j+1}}{\bar{V}^{(b)}} + \sum_{j=1}^i \bar{ST}^{(b)}, \quad (29)$$

where $\bar{ST}^{(b)}$ is the average bus stopping time at any boarding point.

The commuter walking time from his (her) point of origin \mathbf{X} to boarding point \mathbf{P}_i is given by

$$t_H^{(i)} = \frac{k}{V_H} \|\mathbf{X} - \mathbf{P}_i\|, \quad (30)$$

where V_H is the average walking speed and k is a corrective coefficient (*route factor*) reflecting the road network impedance, since the structure of the street network determines walking distances often non-Euclidean in nature [22,37].

The expected waiting time $t_W^{(L)}$ at the boarding point i , on the other hand, is half the headway τ_L observed on route L containing i ($L = s$ for subway and $L = b$ for bus).

The total time spent by a generic commuter traveling from point \mathbf{X} to the *CBD*, through boarding point i , is given by the sum of these three elements

$$T_i = tv_i^{(L)} + t_H^{(i)} + t_W^{(L)} \quad (i = 1, \dots, m). \quad (31)$$

Thus, for every point \mathbf{X} , the model searches for the boarding point i ($i = 1, \dots, m$) that makes $T^{(i)}$ minimum. Making substitutions in (31), T_i can be expressed as

$$T_i = c \|\mathbf{X} - \mathbf{P}_i\| + w_i \quad (i = 1, 2, \dots, m), \quad (32)$$

where c is a constant. Dividing both terms of (32) by c , the resulting districting process, at each stage of the location-allocation model, can be represented by an additively-weighted Voronoi diagram associated to $P \equiv \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_m\}$.

4.1.2. Model development

In the application, travel demand is assumed to follow a decaying density function around the *CBD*. The circular curves in Fig. 4 indicate the density variation (in percentage) over the region. Assuming that the density attains its maximum at the *CBD* point, say a 100% level, the density drops to a 25% level at point 11 (Fig. 4). Furthermore we assumed $k = 1.35$ [22], $V_H = 4.4$ km/h, $\bar{V}^{(b)} = 20$ km/h along the bus route, $V_C = 60$ km/h along the subway line, vehicle stopping time constant and equal to 90 s and 30 s for the bus and the subway, respectively, headway of 5 min for busses and 1.5 min for the subway.

Moreover, commuters are not willing to walk long distances to take the public transportation, preferring to walk, take a taxi, or drive. On the other hand, surveys of public transportation riders have shown that commuters tend to walk farther to reach a *LRT* or subway station than to a bus stop [39]. In addition, when directly walking to their places of work without taking public transportation, commuters tend to walk even farther. Thus, walking distance restrictions were introduced into the model. It was assumed a maximum walking time of 20 min for commuters going to a subway station, 15 min for the ones going to a bus stop, and 30 min for the people who walk directly to the *CBD*.

Let $V = \{V(\mathbf{P}_1), \dots, V(\mathbf{P}_m)\}$ be the additively weighted Voronoi diagram generated by Γ . As previously remarked, $\bigcup_{i=1}^m V(\mathbf{P}_i)$ covers all the region R . Let $\phi(\mathbf{X})$ be the demand density function at \mathbf{X} . The aim is to maximize the subway patronage with two decision variables: (a) the number of subway stations; (b) their locations over the subway line.

Since the path of the subway line is previously designed, the points of Γ_s are restricted to a given curve (the subway line), which may be parametrically described by a real variable θ , say, an angle or an arc length, such that any point on the subway line corresponds to a value of θ lying in the interval $(\theta_{\min}, \theta_{\max})$. We introduce a parametric equation $P = P(\theta)$ describing the geometrical place of the subway line. Then, the elements of Γ_s are defined by a real vector $\Theta = (\theta_1, \dots, \theta_s)$, i.e., there exists a bijective correspondence between the feasible subsets Γ_s and a subset A of the real space R^s , for convenient s and $A: \Gamma_s = \Pi(\Theta_s)$, where Π is a bijection. We have

$$P_{b+i} = P(\theta_i), \quad \Pi(\Theta) = \{P(\theta_i) : 1 \leq i \leq s\}, \quad A = \{\Theta \in R^s | \theta_{\min} \leq \theta_i \leq \theta_{\max}\}. \quad (33)$$

Table 1
Optimization results as a function of the number of stations (case 4.2)

Number of subway stations (s)	Number of iterations ξ	Mean distance between stations (m)	Subway patronage (% of total travel demand)
1	29	2103	25.38
2	65	1402	44.04
3	205	1051	50.71
4	227	841	52.19
5	359	701	52.62 ^a
6	441	601	52.32
7	443	526	51.61

^aMaximum subway catchment due to contrary attracting effects on rail transit patronage.

Thus, the unknowns to be numerically determined are s and Θ_s . The objective function to be *maximized* is

$$G(\Theta_s) = \sum_{i=b+1}^{b+s} \int_{V_i} \phi(\mathbf{X}) d\mathbf{X}. \quad (34)$$

The numerical evaluation of $G(\Theta_s)$ involves the construction of the additively weighted Voronoi diagram defined by the generator set $\Gamma = \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_m\} = \Gamma_0 \cup \Gamma_b \cup \Gamma_s$, such that $\Gamma_s = \Pi(\Theta_s)$. Such a construction must be performed by using an appropriate algorithm [24,27,28]. In the application, we used the plane-sweep technique (Section 3.1).

Moreover, a minimal average distance between stations must be respected. If \widehat{AB} is the total extension of the subway line from the CBD terminal to station s , the following condition must be satisfied

$$\frac{\widehat{AB}}{s+1} \geq d_{\min}, \quad (35)$$

where d_{\min} is the minimum allowable mean distance between subway stations ($d_{\min} = 800$ m in our example). This condition furnishes an upper bound s_{\max} to the number of subway line stations, i.e., $s \leq s_{\max}$ and the adequate value of s can be chosen by comparison of the results for $s = 1, 2, \dots, s_{\max}$. Thus, we adopt the following method of resolution:

Step 1: Initialization: Set $s \leftarrow 1$.

Step 2: Determine the optimal solution Θ_s , $\Gamma_s = \Pi(\Theta_s)$ and the associated value $G(\Theta_s)$: Θ_s maximizes G for the given value of s and Γ_s defines the regions associated to each station. This step involves the use of a convenient optimization procedure. For instance, we may use a standard iterative optimization method which reads as:

- 2.1 Define a maximal iteration number ξ_{\max} , a minimal precision $\eta > 0$, and an initial guess $\Theta_s^{(0)}$ corresponding to a set of subway stations $\Gamma_s^{(0)} = \Pi(\Theta_s^{(0)})$ over the designed subway line. In the example, the initial set was defined dividing the arc \widehat{AC} (Fig. 4) into $s+1$ equal segments. Set the optimization iteration number $\xi \leftarrow 0$ and compute $G_s^{(0)} = G(\Theta_s^{(0)})$.
- 2.2 Determine a new point $\Theta_s^{(\xi+1)} \in A$, corresponding to $\Gamma_s^{(\xi+1)} = \Pi(\Theta_s^{(\xi+1)})$, such that $G_s^{(\xi+1)} = G(\Theta_s^{(\xi+1)})$ satisfies $\Delta_s^{(\xi)} = G_s^{(\xi+1)} - G_s^{(\xi)} \geq 0$.
- 2.3 Increase the optimization iteration number $\xi \leftarrow \xi + 1$. If $\Delta_s^{(\xi)} \leq \eta$ or $\xi \geq \xi_{\max}$ then set $\Theta_s = \Theta_s^{(\xi)}$, $G(\Theta_s) = G_s^{(\xi)}$ and go to step 3. Otherwise, go to step 2.2.

Step 3: Increase the value of s : $s \leftarrow s + 1$ and check restriction (35). If satisfied, go to step 2. Otherwise go to step 4.

Step 4: Since s is a discrete variable, the choice of the adequate number of boarding subway stations is done by direct inspection of the results, as shown in Table 1.

The optimization method to be used in step 2 has to be chosen by the user, but it must be taken into account that G is continuous but not well behaving [20], and therefore may not satisfy convexity assumptions. Thus, one must use global optimization methods. In order to illustrate the feasibility, we considered the classical Hooke–Jeeves direct search method [40]. The model ran with $\xi_{\max} = 500$ and $\eta = 0.0001$ of the interval $(\theta_{\min}, \theta_{\max})$, i.e. $\eta = 38''$. The results are shown in Table 1.

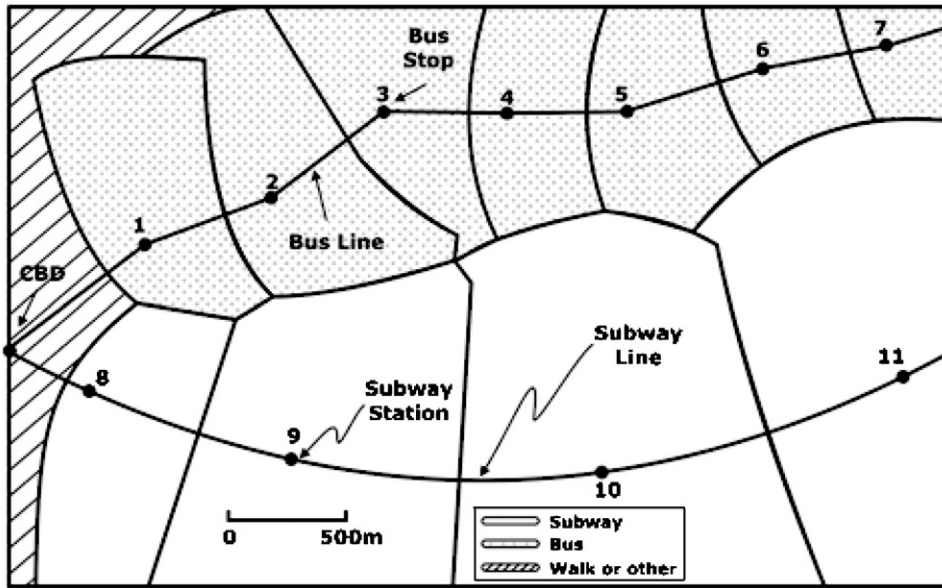


Fig. 5. Additively-weighted Voronoi diagram approach to solve a transit location–districting problem.

It can be seen from Table 1 that increasing the number of subway stations up to 5 the subway patronage increases, decaying thereafter. But, due to operational and cost reasons, the mean distance between stations is restricted to a minimum, which was assumed $d_{\min} = 800$ m in this application. Therefore, the optimum number of subway stations is $s = 4$, and the optimum locations are the ones indicated in Fig. 5. The leftmost hatched area in Fig. 5 stands for the situation in which the minimum travel time to the CBD, considering bus or subway, is greater than the direct walking time to the CBD. This situation also includes the case in which all the traveling alternatives present walking time greater than the maximum allowable limit, implying that the commuter would take another mode of transport (car, taxi, etc.).

Other optimization criteria may be adopted to get the best station locations as, for example, the minimum average travel time to reach the CBD, etc. Additionally, the passenger catchment of each boarding point can be easily computed.

4.1.3. Remarks

Remark 1. This approach may be extended to the situation where the path of the subway line is set free in the designing process. If the points P_{b+i} , $i = 1, \dots, s$ are free to take any place on the region R , then the real vector Θ is formed by the $2s$ coordinates of the points:

$$k = 2s, \quad P_{b+i} = (\theta_{2i-1}, \theta_{2i}), \quad \Pi(\Theta) = \{(\theta_{2i-1}, \theta_{2i}) : 1 \leq i \leq s\}. \quad (36)$$

For instance, when using Cartesian orthogonal coordinates, the odd and pair indexes may correspond to the horizontal and vertical coordinates, respectively: a point is characterized by its coordinates $(x_i = \theta_{2i-1}, y_i = \theta_{2i})$. If the region R is bounded by the rectangle $(x_{\min}, x_{\max}) \times (y_{\min}, y_{\max})$, then

$$A \subset \{\Theta \in \mathbb{R}^k | x_{\min} \leq \theta_{2i-1} \leq x_{\max} \text{ and } y_{\min} \leq \theta_{2i} \leq y_{\max}\}. \quad (37)$$

In these cases, an appropriate curve setting must be externally adjusted to the station points in order to represent the subway line contour, and the minimum distance between stations must be established accordingly.

Remark 2. Since the Hooke–Jeeves method [40] does not guarantee convergence to the global optimum, the results may correspond to local minima (or maxima). In order to check this point and to ensure convergence to a global

minimum (or maximum), stochastic perturbations of the basic method may be introduced as follows:

- we set $\Theta_{s,0}^{(\xi+1)} = \Theta_{s,0}^{(\xi)}$ —the value furnished by the classical Hooke–Jeeves method;
- we generate nt supplementary trial sets of points $\Theta_{s,i}^{(\xi+1)} = \Theta_{s,0}^{(\xi)} + \sigma \mathbf{Z}_i, i = 1, \dots, n$, where $\sigma > 0$ and $(\mathbf{Z}_1, \dots, \mathbf{Z}_{nt})$ is a sample of nt variates of a given random variable \mathbf{Z} .
- Then $\Theta_s^{(\xi+1)}$ corresponds to the best value among all these points: $\Theta_s^{(\xi+1)} = \text{Arg} \max_{0 \leq i \leq nt} G(\Theta_{s,i}^{(\xi+1)})$.

The convergence of the perturbed method follows from classical results (see, for instance, [41,42] for the basic theorems). Other developments may be found in [43–49]. We have tested this approach and the results obtained are very close to those given in Table 1. For instance, with $m = 5$, $\eta = 0$, $nt = 10$, $\sigma = 0.01$ and \mathbf{Z} as the restriction of a Gaussian vector of \mathbb{R}^k to the admissible set A , we got, for $\xi = 1101$, the subway relative patronage of 52.69% (instead of 52.62%), what corresponds to an improvement of only 0.07%. Thus, the use of perturbations should be regarded with caution, since it may significantly increase the computational cost: $nt \times \xi$ supplementary evaluations of the objective function must be performed.

4.2. A parcel distribution problem with geographical barriers

Most logistics distribution and collecting problems involve spatial variables associated with operational and economic elements, such as routing, vehicle capacity, vehicle costs, servicing times, etc. As pointed out by Muyltermans et al. [13], districting associated with transportation and logistics problems should be performed at the strategic and tactical level, whereas routing should be performed at the operational level. In other words, districting involves a more global view and is often related to the managerial and administrative levels, while routing processes are more detailed and linked to day-to-day operations.

In many applications, locations are characterized by Cartesian coordinates and distances are estimated by an L_2 (Euclidean) metric, corrected by a routing factor [22]. Additionally, in one-to-many distribution and collecting problems with multiple tours [33], an idealized dense ring-radial network pattern is frequently adopted as a theoretical modeling basis [16,17,31,32]. In such cases, the ideal configuration of the districts should be wedge-shaped and elongated toward the depot [31]. While interesting from a theoretical point of view, this approach is not readily applicable to real life situations. For more generic metrics, in fact, the optimal orientation of the districts and its shape are not obvious. The indefiniteness of the real local network metric makes the ideal shape of the zones unclear [45]. Furthermore, since the real transportation infrastructure usually presents a coarse network of roads with varying velocity, the ideal orientation of the districts is also unclear [50].

Some computational tests [22] have indicated that the adoption of a Voronoi diagram partitioning process changes the resulting total distribution cost only marginally when compared with the corresponding ring-radial results, and therefore the Voronoi formulation may be used as an adequate districting approximation in a variety of practical applications. In this section we describe and discuss one application of a generalized Voronoi diagram to a logistics distribution problem with barriers in order to illustrate the possibilities of the method.

4.2.1. Problem setting

The utilization of generalized Voronoi diagrams in logistics districting problems has some advantages. The fitting process, for instance, leads to more equalized load factors among the districts, meaning the vehicles assigned to the zones will show more balanced utilization levels. This happens because the generalized Voronoi diagrams have more degrees of freedom when searching for the district contours when compared to the wedge-shaped, geometrical partitioning scheme. Furthermore, as mentioned, the resulting total distribution cost is only marginally affected by such an approximation [22]. Additionally, the utilization of an appropriate Voronoi diagram approach opens the way to solve districting problems with geographical barriers imposed by thoroughfares, highways, rivers, reservoirs, parks, steep hills, etc.

The problem concerns an urban distribution service covering part of the city of São Paulo, Brazil. The objective is to define the number of districts and their boundaries to be assigned to the delivery vehicles in order to: (a) minimize total daily delivery costs; (b) balance the distribution effort among the vehicles, and (c) respect capacity constraints. Additionally, the resulting districts must be contiguous and geographically compact [6]. A homogeneous fleet of m delivery vehicles is assumed and each vehicle is allocated to a district, performing a complete cycle per working day

from a central depot, in a one-to-many distribution scheme [33]. The served urban region \mathfrak{R} is of irregular shape and has an area of approximately 670 sq km. The density of servicing points varies over \mathfrak{R} but is nearly constant and Poisson distributed over distances comparable with a district size [31]. The demand is formed by a total of 6632 servicing points, with an average of 5.76 kg of cargo delivered at each client's location, and with a mean stopping time of 3 min 49 s per visiting point.

Galvão et al. [22] developed a multiplicatively weighted Voronoi diagram model to solve this problem with no geographical barriers. That same problem is now extended to the situation in which geographical obstacles restrain districting limits. Here, the obstacles are represented by two freeways located along the margins of two water courses, the Tietê and the Pinheiros rivers [23]. Those rivers are traversed by a number of bridges, which are congested most of the time. As a practical consequence, transportation operators do not design delivery districts covering both sides of the freeways. This practical situation is handled, in the model, with the introduction of two obstacles represented by the mentioned rivers (Fig. 4). The problem was solved with a power Voronoi diagram formulation associated with the visibility-shortest-path metric (Section 2).

4.2.2. Partitioning criterion and constraints

The balancing criterion for this application is to equalize the distribution effort among vehicles. Each vehicle is assigned to one district and performs one delivery tour, or a cycle, per day. The distribution service along a generic route may be constrained by time or loading capacity. The cycle time is the sum of the line-haul time, the local travel time within the district, and the stopping times at the delivery points. The expected value of the cycle time T_i , for a generic district i , $i = 1, 2, \dots, m$ is [22]

$$T_i = \frac{2kD_i^{(L)}}{v_L} + \frac{k'\sqrt{A_i n_i}}{v_D} + \Psi_{ST}(A_i) \quad (i = 1, 2, \dots, m), \quad (38)$$

where $D_i^{(L)}$ is the expected Euclidean line-haul travel distance (one way) from the depot to the district i , v_L is the average line-haul speed, v_D is the average local speed, A_i is the district area, n_i is the number of delivery points in district i , k and k' are route factors [17,22], and $\Psi_{ST}(A_i)$ is the integral of the stopping times spent in delivering the cargo to the customers in district A_i (Section 3.3). The number of servicing points in the district is given by the integral of the density of points (δ), represented by $\Psi_\delta(A_i)$, as generally indicated in Section 3.3.

If H is the maximum working time per day, the cycle time for any district i ($i = 1, 2, \dots, m$) must satisfy the constraint

$$T_i \leq H \quad (i = 1, 2, \dots, m). \quad (39)$$

On the other hand, let $Q_i = F_i(q)$ be the integral of the quantity of cargo delivered in a generic district i . If W is the vehicle loading capacity, the following capacity constraint must be satisfied:

$$Q_i \leq W \quad (i = 1, 2, \dots, m). \quad (40)$$

In place of restrictions (39) and (40), two loading factors are used with the same objective, the first taking into account time utilization and, the second, vehicle capacity utilization

$$\varphi_i^{(T)} = \frac{T_i}{H} \leq 1 \quad \text{and} \quad \varphi_i^{(Q)} = \frac{Q_i}{W} \leq 1 \quad (i = 1, 2, \dots, m). \quad (41)$$

The load factor for district i is largest of $\varphi_i^{(T)}$ and $\varphi_i^{(Q)}$

$$\varphi_i = \max\{\varphi_i^{(T)}, \varphi_i^{(Q)}\} \quad (i = 1, 2, \dots, m). \quad (42)$$

Thus, the balancing criterion is to equalize load factors among the m districts

$$|\varphi_i - \varphi_j| \leq \varepsilon \quad (i, j = 1, \dots, m), \quad (43)$$

where $\varepsilon > 0$ is a small tolerance factor.

4.2.3. Defining the number of districts

The number m of districts is determined by trial and error. Assuming a value for m and applying the model, it will be necessary to increase m if the resulting values of φ_i ($i = 1, 2, \dots, m$) are greater than one. Conversely, if the resulting load factors are too low, the value of m should be reduced. This process continues until one gets a suitable solution respecting (43). Since the computing process takes some time, it is recommended to choose a better estimate of m to be initially used in the model.

Let Q_R be the total quantity of cargo carried per day in zone R . If W is the cargo capacity of a vehicle, a rough estimate of m is

$$m_Q = Q_R / W. \quad (44)$$

Let A_R be the area of region R and N_R the total number of delivery points in R . The average area of a district and the average number of delivery points per district are, respectively,

$$\bar{A} = A_R / m \quad \text{and} \quad \bar{n} = N_R / m. \quad (45)$$

The average vehicle cycle time can be estimated as

$$\bar{T} = \frac{2k\bar{D}^{(L)}}{v_L} + \frac{k'\sqrt{\bar{A}\bar{n}}}{v_D} + \bar{n}\bar{\tau} = H, \quad (46)$$

where $\bar{\tau}$ is the average stopping time spent in one delivery. If the region R is approximately circular and the depot is fairly centralized, the average distance $\bar{D}^{(L)}$ may be assumed to be equal to $\frac{2}{3}\sqrt{\bar{A}/\pi}$. In this application, however, the region R is somewhat elongated and the depot is located to the south. So, we have made $\bar{D}^{(L)}$ equal to the largest Euclidean distance from the depot to the delivery points in R .

Substituting (45) into (46) and simplifying, one gets

$$m_T = \frac{(k'\sqrt{A_R N_R} / v_D) + N_R \bar{\tau}}{H - 2k\bar{D}^{(L)} / v_L}, \quad (47)$$

and the largest of m_Q and m_T is taken as the initial value of m . Applying the model and analyzing the resulting values of φ_i ($i = 1, 2, \dots, m$), one may change the value of m as explained, running the model again until an acceptable solution is obtained.

4.2.4. Selecting an initial set of generator points

In order to construct the Voronoi diagram embedded in the model, an initial set of m generator points $P = \{P_1, P_2, \dots, P_m\}$ must be defined. One possibility is to randomly generate such a set, selecting m points over R with equal probability. Instead, we have adopted a more elaborated selecting process with the objective of reducing computer time.

As seen in Section 2, the power Voronoi diagram polygons are always convex. Since the intersection of two convex sets is a convex set, it is convenient that the external contour of region R be also convex. Otherwise, its intersections with peripheral districts may not be convex polygons, thus requiring special treatment. For this reason, the convex hull of the served points is determined beforehand and it is adopted in the model to represent region R throughout the computing process. Therefore, when region R is hereafter mentioned in the text, we refer to the convex hull just defined above.

A grid of about 10,000 cells of uniform size is set over region R . The demand of service on each cell is obtained by integrating the demand density function over it (see Section 3.3). The cells are ordained according to some pre-established rule and the corresponding cumulative distribution of demand is defined. A Monte Carlo procedure is used to generate a sample of m generator points as follows. First, a cell is randomly generated from the cumulative distribution of demand. The spatial distribution of demand affects the location of the district generators because their distribution tends to be more compact as the density of demand increases. Second, there is a practical minimum limit to the distance between two adjacent generator points. Roughly assuming that the generator points are Poisson distributed

over R , the distance between a generator point to its closest neighbor is represented by a Rayleigh distribution, whose cumulative distribution function is given by [51]

$$\bar{d}_G = \frac{1}{2\sqrt{\delta_G}}, \quad (48)$$

where \bar{d}_G is the expected distance from a generator point to its closest pair, and δ_G is the density of generator points in R , equal to m/A_R . Thus, when the distance from the sampled cell to its closest neighbor is smaller than \bar{d}_G , the randomly generated cell is discarded and another sample is taken. The Monte Carlo process follows until m generator points are obtained. This Monte Carlo technique is only used to get an initial set of generator points. At the next stages of the iterative process the generator points are made coincident with the centers of mass of the Voronoi regions.

4.2.5. The iterative process

The districting problem with obstacles is solved with a power Voronoi diagram

$$\mu_i(\mathbf{X}, \mathbf{P}_i) = [d_{SP}(\mathbf{X}, \mathbf{P}_i)]^2 + w_i \quad (i = 1, 2, \dots, m), \quad (49)$$

where $d_{SP}(X, P_i)$ is the visibility-shortest-path distance between point X and the generator point P_i (Section 2). Let k represent the stage of the iterative process. At each stage the weights $w_i^{(k)}$ ($w_i^{(k)} \geq 0, i = 1, \dots, m$) are modified according to the following convergence rule:

$$w_i^{(k)} = w_i^{(k-1)} + v_i^{(k-1)}, \quad (50)$$

where $v_i^{(k-1)}$ is given by

$$v_i^{(k-1)} = \frac{\varphi_i^{(k-1)} - \bar{\varphi}^{(k-1)}}{d}, \quad d > 0, \quad (51)$$

where $\bar{\varphi}^{(k-1)} = (1/m) \sum_{i=1}^m \varphi_i^{(k-1)}$ and d is a control parameter. The value of d is changed empirically in order to control the convergence of the model. Since $d > 0$, the weight $w_i^{(k)}$ of district i will increase if $\varphi_i^{(k-1)}$ is greater than the mean $\bar{\varphi}^{(k-1)}$. Putting w_i with a positive sign in (49), the dominance region of P_i (the district area) will decrease [20], tending to lead to a more balanced solution. If $\sigma^{(k-1)}$ is the standard deviation of the observed $\varphi_i^{(k-1)}$ ($i = 1, 2, \dots, m$), the value of d is chosen as to guarantee a decreasing sequence of $\sigma^{(k)}$:

$$\sigma^{(1)} > \sigma^{(2)} > \dots > \sigma^{(k)}. \quad (52)$$

At stage k the iterative process may terminate, in accordance to (43), if

$$\varphi_{\text{Max}}^{(k)} - \varphi_{\text{Min}}^{(k)} < \varepsilon, \quad (53)$$

where $\varepsilon > 0$ is a small tolerance factor. For $k = 1$, the weights $w_i^{(1)}, i = 1, \dots, m$ are set equal to zero, leading to an ordinary Voronoi diagram configuration.

At each stage of the process the power Voronoi diagram is constructed with the quad-tree technique, and the resulting relevant attributes are computed, in special, the center of mass of each district. The center of mass is related here to the concentration of delivery points within the district, since the number of stops is the prevailing variable when dimensioning this kind of service. This is because neither the stop times, nor the delivered quantities vary substantially from point to point. It may occur otherwise in other applications, and the centers of mass of the districts should be determined accordingly. The centers of mass of the districts are then taken as the generator points of the Voronoi diagram for the next stage of the iterative process.

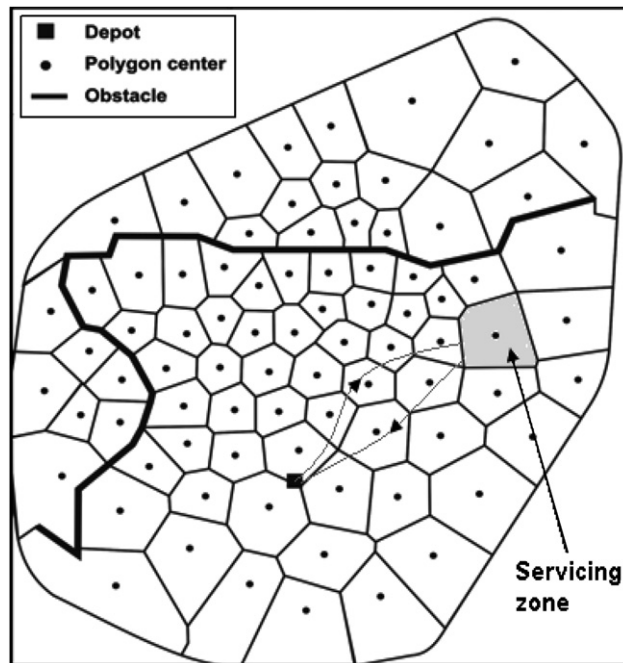


Fig. 6. Power Voronoi diagram logistics districting with barriers [23].

The model was implemented in a Pentium 4, 2.6 MHz computer, and the iterative process converged to a satisfactory solution after 303 iterations. The resulting Voronoi tessellation is shown in Fig. 6.

5. Conclusions

Recent developments in computational geometry opened the way to compute large-scale and more complex Voronoi diagrams applied to transportation and logistics. The utilization of non-ordinary Voronoi diagrams, such as additively weighted and power diagrams, has some advantages when compared to the wedged-shape traditional way. For example, the fitting process leads to more equalized load factors among the districts, meaning the vehicles assigned to the zones will show more balanced utilization levels. The possibility of applications to real-life problems are ample, as shown in the cases analyzed in the text.

Acknowledgments

This research was supported by the National Council for Scientific and Technological Development (CNPq), Brazil, project no. 300886/2005-5.

References

- [1] In: Drezner Z, editor. Facility location: a survey of application and methods. Berlin: Springer; 1995.
- [2] In: Drezner Z, Hamacher H, editors. Facility location: applications and theory. Berlin: Springer; 2002.
- [3] Hale TS, Moberg CR. Location science research: a review. *Annals of Operations Research* 2003;123:21–35.
- [4] Preparata FP, Shamos MI. Computational geometry—an introduction. New York: Springer; 1985.
- [5] Hojati M. Optimal political districting. *Computers & Operations Research* 1996;23(12):1147–61.
- [6] Mehrotra A, Johnson EL, Nemhauser GL. An optimization based heuristic for political districting. *Management Science* 1998;44(8):1100–14.
- [7] Boskaya B, Erkut E, Laporte G. A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research* 2003;144:12–26.
- [8] Williams Jr. JC. Political redistricting: a review. *Papers in Regional Science* 1995;74:13–40.

- [9] Schoepfle OB, Church RL. A new network representation of a “classic” school districting problem. *Socio-Economic Planning Science* 1991;25(3):189–97.
- [10] D’Amico SJ, Wang SJ, Batta R, Rump CM. A simulated annealing approach to police district design. *Computers & Operations Research* 2002;29(6):667–84.
- [11] Zoltners AA, Sinha P. Sales territory alignment: a review and model. *Management Science* 1983;29(11):1237–56.
- [12] Boots B, South R. Modeling retail trade areas using higher-order, multiplicatively weighted Voronoi diagrams. *Journal of Retailing* 1997;73(4):519–36.
- [13] Muyldermans L, Cattrysse D, Van Oudheusden D, Lotan T. Districting for salt spreading operations. *European Journal of Operational Research* 2002;139(3):521–32.
- [14] Zhou G, Min H, Gen M. The balanced allocation of customers to multiple distribution centers in the supply chain network: a genetic algorithm approach. *Computers & Industrial Engineering* 2002;43:251–61.
- [15] Langevin A, Saint-Mleux Y. A decision support system for physical distribution planning. *Revue des Systèmes de Décisions* 1992;1(2–3):273–86.
- [16] Novaes AG, Gracioli OD. Designing multi-vehicle tours in a grid-cell format. *European Journal of Operational Research* 1999;119:613–34.
- [17] Novaes AG, Souza de Cursi JE, Gracioli OD. A continuous approach to the design of physical distribution systems. *Computers & Operations Research* 2000;27(9):877–93.
- [18] Langevin A, Mbaraga P, Campbell JF. Continuous approximation models in freight distribution: an overview. *Transportation Research B* 1996;30(3):163–88.
- [19] Dasci A, Verter V. A continuous model for production-distribution system design. *European Journal of Operational Research* 2001;129:287–98.
- [20] Okabe A, Boots B, Sugihara K. Spatial tessellations concepts and applications of Voronoi diagrams. 2nd ed., Chichester: Wiley; 2000.
- [21] Suzuki A, Okabe A. Using Voronoi diagrams. In: Drezner Z, editor. *Facility location: a survey of applications and methods*. New York: Springer; 1995. p. 103–18.
- [22] Galvão LC, Novaes AG, de Cursi JES, Souza JC. A multiplicatively-weighted Voronoi diagram approach to logistics districting. *Computers & Operations Research* 2006;33:93–114.
- [23] da Silva ACL. Districting strategies in logistics problems using Voronoi diagrams with obstacles. Doctoral Dissertation, Federal University of Santa Catarina, Brazil; 2004 [in Portuguese].
- [24] Aurenhammer F. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys* 1991;23(3):345–405.
- [25] Hoff III K, Culver T, Keyser J, Lin MC, Manocha D. Interactive motion planning using hardware-accelerated computation of generalized Voronoi diagrams. In: *Proceedings of the 2000 IEEE international conference on robotics & automation*. San Francisco, CA; 2000. p. 2931–2937.
- [26] Okabe A, Suzuki A. Locational optimization problems solved through Voronoi diagrams. *European Journal of Operational Research* 1997;98:445–56.
- [27] Fortune S. A sweepline algorithm for Voronoi diagrams. *Algorithmica* 1987;2:153–74.
- [28] Samet H. Hierarchical data structures and algorithms for computer graphics. *IEEE Computer Graphics & Applications* May 1988; 48–68.
- [29] Klein R. *Concrete and abstract Voronoi diagrams*. Berlin: Springer; 1989.
- [30] Daganzo CF. The distance traveled to visit N points with a maximum of C stops per vehicle: an analytic model and an application. *Transportation Science* 1984;18:331–50.
- [31] Newell GF, Daganzo CF. Design of multiple-vehicle delivery tours—I. A ring-radial network. *Transportation Research B* 1986;20B(5):345–63.
- [32] Han AFW, Daganzo CF. Distributing nonstorable items without transshipments. *Transportation Research Record* 1986; (1061):32–41.
- [33] Daganzo CF. *Logistics systems analysis*. Berlin: Springer; 1996.
- [34] Boor C. *A practical guide to splines*. New York: Springer; 2001.
- [35] Bathe K. *Finite element procedures in engineering analysis*. Englewood Cliffs: Prentice-Hall; 1982.
- [36] Hamacher HW, Liebers A, Schöbel A, Wagner D, Wagner F. Locating new stops in a railway network. *Electronic Notes in Theoretical Computer Science* 2001;50(1) 11pp.
- [37] Laporte G, Mesa JA, Ortega FA. Locating stations on rapid transit lines. *Computers & Operations Research* 2002;29:741–59.
- [38] Saka AA. Model for determining optimum bus-stop spacing in urban areas. *Journal of Transportation Engineering* May–June 2001; 195–9.
- [39] Sullivan S, Morrall J. Walking distances to and from light-rail transit stations. *Transportation Research Record* 1538, Washington, DC: Transportation Research Board; 1996.
- [40] Hooke R, Jeeves TA. Direct search solution of numerical and statistical problems. *Journal of the Association for Computing Machinery* 1962;8:212–29.
- [41] Pinter J. *Global optimization in action*. Boston: Kluwer Academic; 1995.
- [42] Souza de Cursi JE. Minimisation stochastique de fonctionnelles non convexes en dimension infinie. Publications du service de Mathématiques de l’Ecole Centrale Nantes, France; 1992. Available at (<http://meca.insa-rouen.fr/~souza/ra92.pdf>).
- [43] Pogu M, Souza de Cursi JE. Global optimization by random perturbation of the gradient method with a fixed parameter. *Journal of Global Optimization* 1994;5:159–80.
- [44] Autrique L, Souza de Cursi JE. Une méthode mixte stochastique/déterministe pour l’identification en vulcanisation. *APII(RAIRO)* 1994;28(3):263–82.
- [45] Autrique L, Souza de Cursi JE. On stochastic modification for global optimization problems: an efficient implementation for the control of the vulcanization process. *International Journal of Control* 1997;67(1):1–21.
- [46] Cortes MBS, Souza de Cursi JE. Approximate Gaussian distributions in optimization by random perturbation methods. *Applied Numerical Mathematics* 1999;30:23–30.

- [47] Gonçalves MB, Souza de Cursi JE. Parameter estimation in a trip model by random perturbation of a descent method. *Transportation Research, part B* 2001;35(2):137–61.
- [48] Bouhadi M, Ellaia R, Souza de Cursi JE. Stochastic perturbation methods for affine restrictions. In: Pardalos P, Floudas N, editors. *Advances in convex analysis and global optimisation*. Boston: Kluwer Academic; 2001. p. 489–500.
- [49] Bouhadi M, Ellaia R, Souza de Cursi JE. Global optimization under nonlinear restrictions by using stochastic perturbations of the projected gradient. In: Pardalos P, Floudas N, editors. *Advances in convex analysis and global optimization*. Boston: Kluwer Academic; 2004. p. 541–62.
- [50] Newell GF, Daganzo CF. Design of multiple-vehicle delivery tours—II other metrics. *Transportation Research B* 1986;20B(5):365–76.
- [51] Larson RC, Odoni AR. *Urban operations research*. Englewood Cliffs: Prentice-Hall; 1981.