Università degli Studi di Napoli Federico II

Scuola Politecnica e delle Scienze di Base

Department of Industrial Engineering (DII)

PhD in Industrial Engineering (XXXI Cycle)

PhD Thesis

Spatial organization of public services: models and applications

Antonio Diglio

Director:
Prof. Giuseppe Bruno

Co-Directors:
Prof. Stefan Nickel, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
Prof. Francisco Saldanha da Gama, University of Lisbon, Lisbon, Portugal

December 2018

Contents

List of Tables

List of Figures

Introduction

Location decisions are crucial in the spatial organization in both public and private sectors as they can have long term impact on operational performances and on service levels. In the public sector, for instance, the optimal positioning of schools, hospitals and emergency services (i.e. fire and police stations, ambulances, etc.) is fundamental to provide users with good (ideally equitable) accessibility conditions and timely response to intervention requests. The same applies to private firms offering services of general interest (postal offices, banks, pharmacies, etc.) in sectors partially regulated by regional or national bodies. In the private sector, the optimal positioning of plants, warehouses, distribution centers or the optimal identification of delivery areas to be assigned to dedicated operators is a crucial task to reduce operational costs and to increase customers' satisfaction. The main difference between the location of public and private sector lies in the nature of the objectives that decision makers are considering. While costs minimization and capture of larger market shares from competitors are the main goals in the private sector, social cost minimization, universality of services, and equity, in terms of accessibility, are the main objectives in public services contexts. Nevertheless, the enduring trend of public expenditures revision poses, also in the public sectors, the need to pursue objectives of economic efficiency. These aspects underline the multi-objective nature of location decisions in public contexts.

In literature, two families of optimization problems are typically used to address these problems, namely *Facility Location Problems* (FLPs) and *Districting Problems* (DPs).

FLPs aim at identifying the optimal position to assign to one or more structures (*facilities*), in a given space, in order to satisfy a *demand* (actual or potential) coming from a set of customers. Clearly, the meaning of "optimal" strictly depends on the nature of the problem under study, namely in terms of the constraints and of the optimality criteria considered.

DPs are aimed at grouping small geographic areas, called basic areas or territorial units (counties, zip code or company trading areas), generally associated with a set of attributes (i.e. demand, population), into a given number of larger geographic clusters, called *districts*, in a way that the latter are acceptable according to some planning criteria. In service-oriented applications, DPs aim at defining the geographic areas to be served by a set of service facilities (already located or to be located).

In the present thesis we aim at showing how FLPs and DPs can be used to underpin spatial organization processes of public services, providing analytical models able to assist the decision making process. To this end, we will develop mathematical models to address some relevant applications and we will realize extensive computational experiments to show their capability to provide insightful managerial implications. Although they are designed in the context of specific problems, it is worth to underline that the models we propose are totally generalizable as they include features of practical relevance which are common to different areas of application.

The remainder of the present thesis is organized as follows.

In Chapter 1 we introduce the two above classes of problems, i.e. facility location and districting problems, by presenting their main features and the most well-known formulations in the literature. We also present a brief overview on stochastic models, needed to cope with strategic location decisions under uncertain conditions, and we discuss the relevance of using Geographic Information Systems within these classes of problems.

In Chapter 2, we present a facility location model to address the spatial organization problem of Blood Management Systems in regional contexts. The proposed model aims at organizing the network according to a two-level hierarchical structure at the minimal cost. Decisions regarding the dimensioning of the facilities and the distribution of blood units are also considered. Moreover, users are guaranteed to access their closest facility within a maximal distance. The model is then tested on the case study of the Campania Region, characterized by serious inefficiencies under the light of the regulatory framework recently introduced by the national authority.

In Chapter 3 we propose an integrated facility location-districting model for the organization of postal collection operations in urban areas. The problem is motivated by the consistent fall in letter post volumes which is rendering the classical postal collection systems highly inefficient. The model we propose has a twofold objective: reducing the number of postboxes, on the strategic level, and grouping the remaining ones in clusters (collection areas) to be assigned to dedicated operators (postmen) for collection operations, on the tactical level. Equity considerations for users' accessibility, as well as requirements concerning workshift duration and workload balance are also taken into account. To solve the problem, we also devise a constructive heuristic procedure and we test it in the case of the city of Bologna, in northern Italy.

In Chapter 4 we propose a general stochastic programming modeling framework for districting, triggered by uncertainty in the demands. In particular, we propose a two-stage mixed-integer stochastic program. The first stage comprises the decision about the initial territory design: the districts are defined and all the territory units assigned to one and exactly one of them. In the second stage, i.e., after demand becomes known, balancing requirements are to be met. This is ensured by means of two recourse actions: outsourcing and reassignment of territory units. The objective function accounts for the total expected cost that includes the cost for the first stage territory design plus the expected cost incurred at the second stage by outsourcing and reassignment. The study aims at filling an existing gap due to the absence of stochastic models in the field of districting literature.

Finally, some general conclusions and direction for further research are drawn.

Chapter 1

Facility Location and Districting Problems

Summary

In this chapter, we discuss two families of optimization problems, particularly Facility Location and Districting Problems. We start by presenting their main features and the mathematical formulations of the most well-known models in literature. Then, stochastic programming models and linkages between facility location and GIS are also discussed. All the elements presented in this chapter are the first stage for the development of the models and applications presented in the subsequent chapters.

1.1 Facility Location Problems

Facility Location Problems (FLPs) aim at identifying the optimal position to assign to one or more structures (*facilities*), in a given space, in order to satisfy a *demand* (actual or potential) coming from a set of customers. The meaning of ”optimal” strictly depends on the nature of the problem under study, namely in terms of the constraints and of the optimality criteria considered (Laporte et al., 2015).
Although the first evidences date back to the 17th century, the first formulation of location problem is attributable to Weber (1929). However, seminal works of the so-called modern location science are considered the ones from Cooper (1963) and Hakimi (1964, 1965). From that point on, location science has become a very active field of research, as it is testified by the more recent and extensive collections on the topic by Eiselt and Marianov (2011) and Laporte et al. (2015). The relevance of FLPs reside in the long term and strategic nature of location decisions which make them applicable in various contexts of both private and public sectors (Giannikos, 1998; Erkut et al., 2008; Melo et al., 2009; Marianov et al., 2002).
Following Eiselt and Laporte (1995) and ReVelle and Eiselt (2005), fundamental elements of a FLP are:

- *Location Space*, i.e. the space where the facilities have to be located;
- *Facilities* (already existing and new) to be located;
- *Customers* expressing service demand;
- *Interaction* between the objects involved;
- *Objectives*, to be optimized;
- *Constraints*, to be satisfied.

The *location space* generally corresponds to the space where customers are present and facilities are to be located. It is possible to distinguish between *continuous*, *network*, and *discrete* problems. In the first case, facilities may be positioned everywhere in the location space; at most there could be some forbidden zones where locations are not allowed due to geographical obstacles or technical constraints. In the second case, the location of a facility is restricted to the nodes and/or the edges of a network while, in the last one, it has to be chosen within a set of candidate sites.
Facilities are the objects to be located that will provide services and/or goods in order to satisfy the demand. They are typically characterized by some attributes such as the number and type of services they offer, the capacity, the attractiveness, the cost related to their construction and/or operational activities. The number of facilities to be located may be either a given parameter of the problem or a decision variable. Some facilities can have infinite capacity (*uncapacitated problems*), while some others no (*capacitated problems*). Moreover, facilities may differ in terms of typology like plants and warehouses in *hierarchical* and *multi-level problems* (Şahin and Süral, 2007; Ortiz-Astorquiza et al., 2017) and in terms of the number of services or commodities they offer (*single-commodity* vs. *multi-commodity*, Pirkul and Jayaraman, 1998).
Customers are the actors requiring goods/services from the facilities located (or to be located). Also the customers can be characterized by attributes like distribution and demand. In fact, they can located at specific points (called *demand points*) or continuously spread over the location space. Typically, each demand point is associated to a value of demand that can be the same for all the customers or dependent on the specific location. Demand values can be also deterministic or described by a certain distribution function, as in the case of *stochastic FLPs* (Correia and Saldanha-da-Gama, 2015).
Interaction concerns the relations between the customers and the facilities and/or between the facilities themselves. Customers interact with facilities in the sense that they are allocated to them for satisfying their demand. This allocation can be compulsory or based on a preference system and utility functions. Accordingly, we distinguish between *location-allocation* and *location-choice models* (Drezner and Eiselt, 2002). Interaction between facilities refers to the way to compete to capture the largest market shares (Aboolian et al., 2007) or, on the contrary, to the way they cooperate in order to assure a certain level of accessibility to the users (Berman et al., 2011).
Location decisions can be made according to different criteria or *objective functions*. In most applications, location decisions can be driven by cost or covering objectives. As

regards costs, they are typically identified with location and allocation costs which are, respectively, the fixed cost to establish a facility and the variable cost needed for services and/or goods provision (Fernández and Landete, 2015). Covering objectives refer to the total demand that can be served by a facility within a certain distance from its location (García and Marín, 2015). Objective functions can also depend on the nature of the provided service. For instance, the location of undesirable facilities may be targeted by dispersion objectives (Erkut, 1990). Due to the multiplicity of objectives to be optimized, multi-objective settings may also be adopted (Farahani et al., 2010).
Finally, the problem can be characterized by many *constraints*. Typical examples are topological constraints (i.e. minimum and/or maximum distances between facilities, zoning laws), capacity constraints (i.e. maximum demand that each facility can serve), technical and/or technological restrictions, economic and budget constraints.
Depending on the combinations of the above elements, a wide range of mathematical models can be defined. Examples of classification schemes are proposed by Revelle et al. (2008) and Farahani et al. (2012).

1.2 Basic Facility Location Models

In this section, we present some core location models as they represent the starting point for the development proposed in the following chapters. In particular, we refer to discrete problems in which both customers and facilities are located at specified points in a given location space.
To this end, we introduce the following notation:

- I, set of demand points, where customers are located, indexed by i;
- J, set of candidate locations for facilities, indexed by j;
- d_i, demand (weight) associated to node $i \in I$:
- c_{ij}, distance between nodes $i \in I$ and $j \in J$;
- y_j, binary variable equal to 1 if a facility is located in $j \in J$;
- x_{ij}, binary variable equal to 1 if demand point $i \in I$ (customers) is allocated to facility $j \in J$

1.2.1 Median models

The p-median problem finds the optimal location of exactly p facilities, so that the weighted sum of the distances between customers and their assigned facilities is minimized. Since the number $|I|$ of customers is known, by dividing the objective by $|I|$, the minimum average weighted distance between customers and facilities is obtained.
The first formulation, known as the "classical", is attributable to Hakimi (1964) and it is as follows:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i \in I} \sum_{j \in J} d_i c_{ij} x_{ij} && && (1.1)\\
\text{subject to} \quad & \sum_{j \in J} x_{ij} = 1 && i \in I && (1.2)\\
& x_{ij} \leq y_j && i \in I, j \in J && (1.3)\\
& \sum_{j \in J} y_j = p && && (1.4)\\
& x_{ij}, y_j \in \{0, 1\} && i \in I, j \in J && (1.5)
\end{aligned}
$$

Constraints (1.2) force each demand point to be assigned to only one facility. Constraints (1.3) allow demand point i to be assigned to a point j only if there is an open facility in that location. Constraint (1.4) sets the number p of facilities to be located. Finally, Constraints (1.5) define the domain of the introduced decision variables.
Starting from this basic formulation, several extensions have been proposed in the literature to include practical features or to target specific applications. Holmes et al. (1972) presented a formulation that considered that people would not travel beyond a given threshold distance in order to replicate an elastic demand with respect to the distance and applied it to the case day care facilities in Columbus, Ohio. In the same work, the authors also introduce the so called *capacitated p-median* problem, by imposing a maximum capacity level for facilities to be located.
Balancing requirements, obtained by imposing also a lower bound on the capacity level of the facilities, can be also introduced to avoid strong inefficiencies or to obtain a more balanced distribution of customers between facilities, as in Carreras and Serra (1999). The authors applied their formulation to the location of pharmacies in a rural region of Spain.
The p-median problem is proven to be $\mathcal{NP}$-Hard (Kariv and Hakimi, 1979). Therefore, a considerable number of contributions in the literature is devoted to develop tailored solution methods (exact, heuristic and metaheuristic) and more efficient reformulations. In this regard, García et al. (2011) proposed a so-called "*radius formulation*" for solving exactly large scale p-median problems. For a recent review on this family of optimization problems, interested readers may refer to Daskin and Maass (2015).

1.2.2 Covering Models

One of the classical objectives in location modeling is "coverage" which seeks to ensure that each customer is "covered", namely served by a certain facility if the distance between them is lower than a certain threshold value, typically indicated as *covering radius*. In the literature, there are two basic formulations of covering models:

- *Set Covering Location Problem* (SCLP), seeking to minimize the number of facilities needed for full coverage of the population in a given location space;

- *Maximum Covering Location Problem* (MCLP), aiming at maximizing covered population, given a limited number of facilities or budget.

The SCLP was introduced by Toregas et al. (1971) in the location problem of emergency service facilities. In this case, the response time to reach every customer is a crucial feature of the problem. Therefore, considering the following additional notation:

S, target distance for coverage, i.e. covering radius;
N_i, set of all those sites that are within distance S from node i ($N_i = \{j \in J : c_{ij} \leq S\}$).

the (SCLP) can be formulated as follows:

$$\text{minimize} \quad \sum_{j \in J} y_j \tag{1.6}$$

$$\text{subject to} \quad \sum_{j \in N_i} y_j \geq 1 \qquad i \in I \tag{1.7}$$

$$y_j \in \{0,1\} \qquad j \in J \tag{1.8}$$

Objective function (1.6) minimizes the number of facilities to be located. Constraints (1.7) assure that at least one facility is located within distance S from any node i, i.e. every node is covered within distance S. Constraints (1.8) specify the domain of variables y_j.
Differently from the SCLP, the MCLP seeks to maximize the demand covered by locating a fixed number p of facilities. The problem, introduced by Church and ReVelle (1974), considers binary decision variables $x_i, i \in I$ equal to 1 if and only if user i is covered. According to the introduced notation, its formulation can be expressed as follows:

$$\text{maximize} \quad \sum_{i \in I} d_i x_i \tag{1.9}$$

subject to (1.4)

$$\sum_{j \in N_i} y_j \geq x_i \qquad i \in I \tag{1.10}$$

$$x_i, y_j \in \{0,1\} \qquad i \in I, j \in J \tag{1.11}$$

Objective function (1.9) maximizes the total demand covered. As in the p-median model, Constraint (1.4) fixes the number of facilities p to be located. Constraints (1.10) ensure that a demand point i is covered if and only if at least one facility is located within distance S. Constraints (1.11) define the nature of the decision variables. Both the formulations presented above can be derived as special cases of a more general formulation proposed by García and Marín (2015). In that reference, the authors also surveyed theoretical properties of covering models as well as several solution methods proposed in the literature. It also worth underlining that the above models rely on some specific hypotheses regarding the covering objectives, such as:

- *All or nothing coverage*, i.e. a demand node is fully covered if its distance from a located facility is within the covering radius;
- *Individual coverage*, i.e. every demand point can be served by only one facility;
- *Fixed coverage radius*, i.e. the covering radius is a fixed exogenous parameter.

Although these hypotheses are widely used in the location literature, as they allow to design simpler mathematical models, they might be not very applicable to real case-studies. Intuitively, in fact, one can imagine that the concept of coverage tends to be a function of the distance between demand points and facilities, that the covering radius may vary depending on some attributes of the facility itself and that more facilities can cooperate to cover the demand points. Generalizing these aspects, Berman et al. (2010) surveyed classes of covering models derived by relaxing the above assumptions, namely:

- *Gradual cover models*, in which a general coverage function which represents the proportion of demand covered at a certain distance from the facility is considered;
- *Cooperative cover models*, where all facilities contribute to the coverage of each demand point;
- *Variable radius models*, where the coverage radius is an endogenously determined function of the facility cost.

In particular, a new formulation for the MCLP is proposed in Alexandris and Giannikos (2010). The latter is based on the notion of *complementary partial coverage*, i.e. when different facilities, opportunely dispersed in the location space, can cooperate the achieve full nodes coverage. The model is then applied to the location of bank branches in the municipality of Athens.

1.2.3 Center Models

This class of problems involves locating p facilities in such a way that every demand point receives its service from the closest facility and the maximum distance between each demand node and its facility is as small as possible. Therefore, center models belong to the so-called class of *Minimax* problems, differently from the median problems (*Minisum*). Indicating by C the maximum distance between a demand node and its nearest facility, and making use of the notation previously introduced, the problem can be formalized as follows (Hakimi, 1964):

$$\text{minimize} \quad C \tag{1.12}$$

$$\text{subject to} \quad (1.2) - (1.5)$$

$$C \geq \sum_{j \in J} c_{ij} x_{ij} \qquad i \in I \tag{1.13}$$

Objective function (1.12) seeks for the minimization of C which, according to Constraints (1.13) is defined as the maximum distance between any node i and its closest facility j. Clearly, demands d_i can be used as weighting factors in the objective functions.
If the number of facilities to be located is equal to one we obtain the so-called *Absolute center Problems*, introduced by Hakimi (1964). Capacity restrictions on the facilities are considered by Özsoy and Pınar (2006).
Center location problems commonly arise in emergency service location, where the goal of quick response times is significantly more important than any efficiency consideration, related to the cost of delivering that service. For an overview on formulations, solution methods and fields of applications, a recent survey is provided in Calik et al. (2015).

1.3 Districting Problems

Districting Problems (DPs) are aimed at grouping small geographic areas, called basic areas or territorial units (counties, zip code or company trading areas), generally associated with a set of attributes (i.e. demand, population), into a given number of larger geographic clusters, called *districts*, in a way that the latter are acceptable according to some planning criteria. The latter typically refer to balancing, which expresses the need for districts of equitable size in terms of dimension, as well as to topological properties like contiguity and compactness. Contiguity means that in order to travel between Territorial Units (TUs) in the same district there is no need to cross other districts. A contiguity requirement is relevant for dealing appropriately with enclaves (a district within a district). In fact, a good districting plan does not contain enclaves. Compactness indicates that a district is somewhat round-shaped and undistorted (Kalcsics, 2015). Nevertheless, other relevant criteria in DPs include respecting natural boundaries, existing administrative subdivisions, similarity w.r.t existing districting plans, and socio-economic and cultural homogeneity (Bozkaya et al., 2003; Kalcsics et al., 2005; Kalcsics, 2015).
Formulations for DPs resort to the seminal contribution of Hess et al. (1965). To present it, we consider a set I of territorial units that we want to divide into a fixed number, say p, of districts. Each district will have a TU representing it. Hence, when some other TU is assigned to the district we abuse the language by saying that we are assigning a TU to the representative of the district. We note that single-assignment (i.e. *integrity*) is assumed for the TUs as customary in districting problems.
We consider the following parameters defining the problem:

d_i, demand of TU i ($i \in I$);

c_{ij}, cost for assigning TU i to TU j ($i, j \in I$);

α, maximum desirable deviation of the demand in a district w.r.t. the reference value μ;

μ, average demand per district ($\mu = \sum_{i \in I} d_i / |I|$);

and the following decision variables:

$$x_{ij} = \begin{cases} 1, & \text{if TU } i \text{ is assigned to TU } j; \\ 0, & \text{otherwise.} \end{cases} \qquad (i, j \in I)$$

In this case, $x_{jj} = 1$ indicates that TU j is assigned to itself which also indicates that it is selected as the "representative" TU of its district.

Using these decision variables, Hess et al. (1965) proposed the following mathematical model for the territorial districting problem with balancing constraints in the context of political districting:

$$\text{minimize} \quad \sum_{i \in I} \sum_{j \in I} d_i c_{ij}^2 x_{ij} \tag{1.14}$$

$$\text{subject to} \quad \sum_{j \in I} x_{ij} = 1 \qquad i \in I \tag{1.15}$$

$$\sum_{j \in I} x_{jj} = p \tag{1.16}$$

$$(1 - \alpha)\mu x_{jj} \leq \sum_{i \in I} d_i x_{ij} \leq (1 + \alpha)\mu x_{jj} \qquad j \in I \tag{1.17}$$

$$x_{ij} \in \{0, 1\} \qquad i, j \in I \tag{1.18}$$

The objective function (1.14) quantifies the total cost to be minimized. Constraints (1.15) ensure that each TU is assigned to exactly one district whereas Constraints (1.16) guarantee that exactly p districts will be designed. Constraints (1.17) are the balance constraints. Finally Constraints (1.18) define the domain of the decision variables.

In districting problems, the costs c_{ij} are typically related to the distances (Kalcsics, 2015). In particular, denoting by ℓ_{ij} the distance between i and j $(i, j \in I)$, a common cost to consider is $c_{ij} = \ell_{ij}$ or $c_{ij} = \ell_{ij}^2$. This turns the above objective function into a so-called compactness measure known as moment of inertia (Hess et al., 1965). The reader may refer to Kalcsics (2015) for variants of distance-based compactness measures. In particular, in that book chapter, we also observe cost structures that consider the demands as weighting factors. Finally, we note that euclidean distances are often considered (Bergey et al., 2003; Bard and Jarrah, 2009).

The relevance of DPs resides in the strategic and long-term nature of the decisions involved, motivated by a wide spectrum of practical applications arising in different sectors. In fact, apart from political districting (see also Ricca and Simeone, 2008; Ricca et al., 2013), DPs have been extensively applied to tackle problems emerging in the context of strategic service planning and management like health-care (Blais et al., 2003; Benzarti et al., 2013), school systems (Ferland and Guénette, 1990; Schoepfle and Church, 1991; Caro et al., 2004; Bruno et al., 2016a), energy and power distribution networks (Bergey et al., 2003; De Assis et al., 2014; Yanık et al., 2016), police districts (D'Amico et al., 2002), waste collection (Mourão et al., 2009), and transportation (Bruno and Laporte, 2002; Tavares et al., 2007). Other core applications of DPs regard the design of commercial areas to be assigned to a given sales force (Zoltners and Sinha, 2005; Ríos-Mercado and López-Pérez, 2013) and distribution logistics (Zhong et al., 2007). For extensive reviews on DPs readers are encouraged to refer to Kalcsics et al. (2005) and, for a more up-to-date overview, to Kalcsics (2015).

Remark 1. *A close look into the above model proposed by Hess et al. (1965) reveals that it does not solve exactly a districting problem but only a relaxation of it; it solves a location-allocation model with a demand balancing requirement. In fact, the model ignores one important aspect in districting: contiguity (and consequently no enclaves). Therefore, apparently, a solution provided by the model may yield contiguity solutions only by chance. Nevertheless, it should be noted that when we seek to optimize a compactness measure in a districting problem the resulting solution is typically strong in terms of contiguity of the districts. It is also true that the introduction of balancing constraints may disfavor compactness and, accordingly, districts' contiguity. In other words, balancing and compactness measures may easily become conflicting. Despite these facts, most of the authors do not explicitly address contiguity in their models (Kalcsics, 2015) and thus the model proposed by Hess et al. (1965) is still the basic model for researchers working in districting problems. For this reason, we also set it at the core of the developments we propose in chapter 4. In that chapter, we also review the more recent contributions found in the literature.* ◇

1.4 Facility Location Problems under Uncertainty

FLPs involves strategic decisions that must hold for considerable time. During this time, some disrupting event may occur or some parameters may very in an unpredictable manner. In these cases, it is desirable to embed uncertainty in the model, in such a way to somehow anticipate its effects.

Several aspects emerge as crucial when it comes to uncertainty; following Correia and Saldanha-da-Gama (2015) we briefly discuss them next.

A first important aspect is given by how uncertainty can be modelled, i.e. its *representation*. If probabilistic information is available, the uncertain parameters can be represented through random variables. In this case it is possible to use stochastic programming models to deal with the problem. If this is not the case, it is usually considered a measure of robustness to evaluate the performance of the system.

A complete realization of all uncertain parameters is called a "*scenario*", which is another fundamental element when dealing with uncertainty. This notion is independent of whether probabilistic information is available or not. However, if uncertain parameters can be represented by random variables, some probabilities can be associated with each scenario. Defining scenarios is in itself a major problem. In some situations, scenarios are associated with some driving forces (eg: political conditions in a specific region, economic trends or even technological developments) which, in turn, influence the input of the model that supports decision making. In this case, it is up to the decision maker to understand these driving forces and how they affect the model input. This understanding thus leads to a complete definition of the scenarios.

Furthermore, depending on the problem, one can have a *finite* or *infinite number* of scenarios; this has an impact on the models that can be used.
Another important feature that influences the type of model to consider concerns the decision maker's *attitude towards risk*. Two classes of attitudes are generally considered: *risk neutral* and *risk averse*. In the first case, the decision maker does not consider the risk when making a decision; the utility associated with the decision maker is represented by a linear function. When a probability is associated with each scenario, a risk-neutral decision maker looks for the decision that minimizes the expected cost (or maximizes the expected return). In the second case, the decision maker takes into account the risk when making a decision; the utility associated with the decision maker is represented by a concave function (when the utility is measured on the vertical axis and the monetary value is measured on the horizontal axis). In this circumstance, the decision maker wants to avoid unnecessary risks and therefore the expected value of future activities is no longer an appropriate objective. This decision maker can look, for example, at the solution by minimizing the maximum cost in all scenarios.
Finally, in some classes of problems, there is another aspect that influences the mathematical model to be taken into consideration: the identification of *ex ante* and *ex post* decisions. The ex-ante decisions are the so-called "here and now" decisions, ie the decisions that must be implemented before the uncertainty is revealed. Ex-post decisions, on the other hand, are the decisions to be made after the uncertainty has been revealed. This is a consequence of the strategic nature of these decisions in many problems in the literature, especially in the context of FLPs.
The elements discussed above pose a clear difference between two different ways to deal with uncertainty in optimization problems (and hence, in FLPs), namely *stochastic programming* and *robust optimization*. For more comprehensive overviews on the above topics, interested readers may refer to Birge and Louveaux (2011) (for stochastic programming) and Ben-Tal et al. (2009) (for robust optimization).
For the sake of completeness, we also remind that another class of optimization problems under uncertainty is represented by *chance-constrained* programs. The idea is that one or several constraints of the problem are not required to always hold. Instead, the decision maker is satisfied if they hold with some given probability. Examples of FLPs solved by means of chance-constrained programs are in the works by Lin (2009) and, very recently, by Kınay et al. (2018a,b).
Considering the scope of the present thesis, in the following we put our attention only on stochastic programming and, in particular, we specialize it to the case of FLPs.

1.4.1 Stochastic Facility Location Problems

In this section, we briefly introduce stochastic programming models for FLPs. We follow the clear explanation provided in Correia and Saldanha-da-Gama (2015). For illustrative purposes, let us consider the *Uncapacitated Facility Location Problem* (UFLP), firstly introduced by Balinski (1965). The aim of this problem is to decide the number and the position of facilities to be located in order to satisfy all the demand at the minimal cost, which is given by the sum of location and allocation costs.

Considering fixed costs r_j for locating a facility in j, the UFLP can be formulated as follows:

$$\text{minimize} \quad \sum_{j\in J} r_j y_j + \sum_{i\in I}\sum_{j\in J} d_i c_{ij} x_{ij} \tag{1.19}$$
$$\text{subject to} \quad (1.2), (1.3), (1.5)$$

Remark 2. *It is worth noting that the p-median model (1.1)-(1.5) can be derived as a special case to the UFLP - (1.19), s.t. (1.2),(1.3),(1.5), by introducing Constraints (1.4) if location costs are independent from the specific location ($r_j = r, j \in J$) .* ◇

If we suppose that:

- uncertainty affects both demands d_i and allocation costs (travel times) c_{ij};
- uncertainty can be measured probabilistically;
- we know the joint probability distribution of the vector containing all the random parameter, say $\boldsymbol{\xi}$ ($\boldsymbol{\xi} = [(c_{ij})_{i\in I, j\in J}, (d_i)_{i\in I}]$);

and if we assume *ex ante* location decisions and *ex post* allocation decisions, a stochastic uncapacitated facility location problem can be formulated as follows:

$$\text{minimize} \quad \sum_{j\in I} r_j y_j + \mathcal{Q}(\mathbf{y}) \tag{1.20}$$
$$\text{subject to} \quad y_j \in \{0,1\} \qquad j \in J \tag{1.21}$$

with $\mathcal{Q}(\mathbf{y}) = E_{\boldsymbol{\xi}}[Q(\mathbf{y},\boldsymbol{\xi})]$ and:

$$Q(\mathbf{y},\boldsymbol{\xi}) = \min \quad \sum_{i\in I}\sum_{j\in I} d_i c_{ij} x_{ij} \tag{1.22}$$
$$\text{s. t.} \quad \sum_{j\in J} x_{ij} = 1 \qquad i \in I \tag{1.23}$$
$$x_{ij} \leq y_j \qquad i \in I, j \in J \tag{1.24}$$
$$x_{ij} \in \{0,1\} \qquad i \in I, j \in J \tag{1.25}$$

Model (1.20) - (1.25) is defined for every realization of the random parameter ξ. Location decisions are here-and-now decisions and, hence, are included in the first stage of the model: they will never change regardless of the specific realization of ξ. Allocation decisions are ex post decisions and hence, they are included in the second stage of the model, since they are taken only when uncertainty is disclosed. This is the main reason why they are also called *recourse decisions*. In practice, the above model reproduces the attitude of a decision maker who has to keep strategic decisions by accounting for all the possible features that may happen, as it is expressed by the *recourse function* $\mathcal{Q}(\mathbf{y})$

which links the two stages of the model. The expectation defining the recourse function $\mathcal{Q}(\mathbf{y})$, implicitly conveys a neutral attitude of the decision maker towards risk. Moreover, it has to be also noted that the above model has relatively complete recourse: in fact, for every first-stage feasible decision y_j, $j \in J$, there is always at least one second-stage feasible *completion* (solution) x_{ij}, $i \in I, j \in J$.
If we assume that the support of the random vector $\boldsymbol{\xi}$ is finite, say Ξ, then we can go farther in terms of formulating our problem. In fact, in that case, we can index the different scenarios in a finite set, say $S = \{1, ..., |\Xi|\}$. Moreover, we can index in S the stochastic demands, the assignment costs as well as the second stage decision variables, as follows:

- d_{is}, demand of node $i \in I$ under scenario $s \in S$.
- c_{ijs} allocation cost of demand node $i \in I$ to facility $j \in J$ under scenario $s \in S$;
- x_{ijs} binary variable equal to 1 if demand node demand node $i \in I$ is allocated to facility $j \in J$ under scenario $s \in S$.

If we assume that the probability π_s that scenario $s \in S$ occurs is known (naturally, $\pi_s \geq 0$, $s \in S$ and $\sum_{s \in S} \pi_s = 1$), we can obtain the following extensive form of the deterministic equivalent:

$$\text{minimize} \quad \sum_{j \in J} r_j y_j + \sum_{s \in S} \pi_s (\sum_{i \in I} \sum_{j \in J} d_i c_{ij} x_{ij}) \tag{1.26}$$

$$\text{subject to} \quad (1.21)$$

$$\sum_{j \in J} x_{ijs} = 1 \qquad i \in I, s \in S \tag{1.27}$$

$$x_{ijs} \leq y_j \qquad i \in I, j \in J, s \in S \tag{1.28}$$

$$x_{ijs} \in \{0, 1\} \qquad i \in I, j \in J, s \in S \tag{1.29}$$

Several adjustments may be required if extra features are added to the above model. The introduction of capacity constraints, for instance, may put in crisis the relatively complete recourse of the model. Indeed, it may happen that for some first-stage solutions no feasible completion exists in the second-stage. Correia and Saldanha-da-Gama (2015) show how to cope with this issue by penalizing the non-satisfied demand. Louveaux (1986) considered the case in which capacities are endogenous and, in particular, treated as first-stage decisions. Overturning the perspective at the basis of the previous works, Hinojosa et al. (2014) considered a problem with location decisions made at a tactical or operational level, i.e., location decisions are *ex post* decisions, while distribution decisions are made at a strategic level. Stochastic FLPs triggered by different probability distribution functions describing customers' demand, are explored in Albareda-Sambola et al. (2011) and Bieniek (2015).
In the all the above works it is assumed that uncertainty is revealed in only one occasion. When this is not the case, a *multi-stage* setting seems to be appropriate. Examples of

papers adopting a multi-stage setting include the multi-period facility location problems addressed by Hernandez et al. (2012) and Nickel et al. (2012).

1.5 Using GIS in Location Models

As the aim of FLPs is to locate one or more facilities, in a given space, in order to satisfy a demand coming from a set of customers, the availability of geographically referenced information represents a fundamental prerequisite to model and solve such problems, especially when dealing with real-world applications. In this regard, Geographic Information Systems (GIS) can play a crucial role for supporting decision making in the field of location science (Bruno and Giannikos, 2015).
GIS are information systems that capture, store, control and display geographic information as well as non-spatial information. All GIS perform a set of basic functions including the *management*, *transformation*, *analysis* and *visual presentation* of spatially referenced information. The *management* of spatial and attribute data refers to the need to input, store and handle large amounts of data. The *transformation* of information reflects the need for georeferencing i.e. for linking each data item to its location in a common coordinate system. The *analysis* function enables the application of query, proximity, centrality and other functions to one or more information layers. Finally, the *visual presentation* of data and results has been a core component of all GIS packages and offers tools for the production of digital map, figures and graphic displays.
In the context of FLPs, for instance, information concerning the location space, customers and facilities can be managed by creating appropriate data sets containing specific attributes (i.e. population, areas, etc.). Then, they can be transformed in spatial data sets, called *layers*, by associating each data item to a geometry (points, lines, polygons) and some geographic attributes (e.g. coordinates) required for their proper representation. New layers can be created by aggregating, converting, overlaying or applying some functions on existing layers. For example, it is quite common in location science to discretize a continuos location space, divided in territorial units (e.g. zip areas, municipalities, census tracts, etc.) by extracting their centroids and assuming them as demand points. Other core elements of FLPs and DPs, like distance and adjacency matrices can be also calculated by exploiting GIS capabilities. Finally, solutions obtained can be properly visualized. For an overview on FLPs and GIS linkages, readers can refer to Murray (2010).
A great number of studies appeared in the literature over the years, reporting applications where GIS were employed to tackle a wide range of practical FLPs and DPs. In practice, it is impossible to list all the studies and domains where GIS were applied. Based on Bruno and Giannikos (2015), a non-exhaustive list of broad categories to classify the extant literature is proposed as follows:

- *Land-Use Suitability Analysis*, i.e. the identification of the most appropriate spatial pattern for future land uses in such a way that a set of requirements, properties and preferences are satisfied (Malczewski, 2004). Recent studies deal with GIS-based multi-criteria methodologies applied in the location of wind and solar farms (Sindhu et al., 2017; Ayodele et al., 2018);

- *Waste Management*, i.e. the set of activities associated with the overall chain of managing solid waste in order to reduce its impact on the environment (Higgs, 2006; Wu et al., 2016);

- *Energy Management*, i.e. the identification and selection of marginal lands where energy recovery and production plants can be located, as in the case of biofuel crop production (Niblick et al., 2013), green roofs (Gwak et al., 2017) and waste conversion facilities (Khan et al., 2018);

- *Transportation*, i.e. the optimal positioning of new infrastructures (i.e. roads and highways, parking lots, metro and railway stations, bus lines and stops, intermodal terminals, airports, etc.), based on cost-benefit and users' accessibility analysis (Gutiérrez et al., 2010). Recent applications employing GIS-based multi-criteria analysis are those by Terh and Cao (2018) and Erbaş et al. (2018) dealing, respectively, with cycle paths planning and electric vehicle charging stations location;

- *Private and Public Sector Applications*, i.e. the use of GIS based approach for optimal locating private and public service facilities.

Given the scope of this thesis, we particularly refer to the last category. The applications shown in the following chapters will make extensive use of GIS tools for the problem representation and analysis as well as for solutions visualization phases.

1.6 Contribution of the thesis

For the sake of clarity, we feel the need to remark to the reader the scope of the present work before moving to the specific applications shown in the subsequent chapters.
The research activities at the core of this thesis will focus on problems concerning the spatial organization of public facility services. In this sector, location decisions play a crucial role to provide users with fair (ideally equitable) accessibility conditions and to achieve, at the same time, objectives of economic efficiency due to the general tendency to revise public expenditures. In such a context, it is necessary to find trade-off solutions between these related but, at the same time, conflicting objectives. Therefore, based on the body of literature briefly presented in the previous sections, the purpose of this thesis is to develop analytical frameworks, based on mathematical models integrated with a GIS environment, to support spatial organization processes of public services. In particular, our focus will be posed on the development of optimization models to address emerging real-world applications in the health-care and postal sectors. In this regard, extensive computational experiments will be carried out to validate the proposed models and to demonstrate their capability to provide insightful managerial implications. Moreover, we will also put our effort on the development of stochastic models, especially in the

context of districting problems, which could be potentially useful for practitioners to cope with strategic decisions under uncertainty.

1.7 Conclusion

In this chapter, we presented generalities on Facility Location and Districting Problems. We presented the elements and the formulations of the basic mathematical models known in the literature. An introduction to stochastic models and on the use of GIS in location science is also given. In the next chapters we will present novel mathematical models to address two applications concerning public service facilities and, in the last one, we will introduce a stochastic programming modeling framework for districting. Therefore, in each chapter, a verticalization on specific literature streams will be provided and the gaps we aim at filling will be accordingly outlined.

Chapter 2

A Facility Location Model for the spatial organization of regional Blood Management Systems

Summary

In this chapter, we deal with a first application in the field of public services management concerning the reorganization of regional Blood Management Systems (BMSs).
Blood is a vital resource in the health care context and consequently its efficient and effective collection, management and distribution is a fundamental task. As blood is donated on voluntary basis, an appropriate organization of the collection systems on the territory is crucial; in particular, a widespread presence of collection facilities is important to attract potential donors and assure a given blood collection capacity, but it may have also a strong impact on the total management costs.
This study is motivated by a project aimed at reorganizing regional blood management systems in Italy, in order to reduce total management costs without compromising the self-sufficiency goal, i.e. the goal of satisfying the blood demand coming from the region.
In order to address the problem, we formulate a mathematical programming model and test it on real data related to a specific regional context.
Extensive computational experiments are realized in order to show the capability of the model to handle real-world instances and to act as an analytical framework to assist the strategic redesign of BMSs.

2.1 Introduction

Blood transfusion activities play a crucial role in the context of national health care systems (Williamson and Devine, 2013). Countries throughout the world are facing serious challenges in making sufficient supplies of blood and blood products, while also ensuring their quality and safety. These challenges mainly include the risk of transfusiontransmitted infections, an inadequate number of blood donors, increasing needs for blood and blood products, weak quality systems, and inappropriate and unsafe use of blood and blood products, which may lead to chronic blood shortages, inequitable access, unsafe products and unsound clinical transfusion practices (WHO Expert Group, 2012). In order to tackle such critical issues and make the blood management systems more effective and efficient, in the last decades, most developed countries introduced strict legislation, according to accepted international guidelines and standards.

At EU level, a set of Directives were released in the last years (European Parliament and Council of European Union, 2003, 2004, 2005b,a), aimed at regulating this important sector of health care. Recently, a new document by the European Directorate for the Quality of Medicine and Health care of the Council of Europe (EDQM/CoE), in collaboration with the EU Committee, defined guidelines about quality principles to be adopted and efficiency standards to be met by blood establishments (European Directorate for the Quality of Medicines, 2015). Local governments are asked to implement the European provisions by releasing a set of national guidelines for Regional Authorities.

At regional level, all those activities needed to allow the transfusion of blood from the donors to patients are performed by Blood Management Systems (BMS), composed of a set of dedicated facilities (usually named Blood Centers, BCs), that are linked with hospitals and other types of health centers. In particular, in these facilities blood is collected in an anticoagulant, tested for evidence of infections, processed, stored and then transfused to patients.

The goal of a BMS is to assure that such components are readily available in efficient way to patients whenever and wherever they are needed. First of all, this requires that the total collected amount of whole blood exceeds a given target value, determined on the basis of historical data about yearly regional demand (*self-sufficiency goal*). To achieve this goal, the facilities spread out over the region should be able to attract a significant number of donors. To this aim, their position, in terms of distance from potential donors, plays a crucial role to foster donations; i.e., the higher the presence of such facilities, the higher the possibility to attract donors. But on the other hand, a widespread presence of facilities produces relevant costs due to the need of qualified and specialized staff and of dedicated equipment. In this context, minimizing management costs and satisfying demand represent two conflicting goals, which require the identification of a compromised solution.

In this chapter, we propose an optimization model, based on a mathematical formulation of the problem, aimed at reorganizing the position of these facilities in a regional context. This study is motivated by a real project concerning the reorganization of regional blood management systems in Italy. The project is coherent with a general trend of reducing public expenditures in the public sector, which is common to all western economies. In particular, we formulate the problem as a FLM that is applied to real data regarding the specific regional context of the Campania Region, in southern Italy. We illustrate several scenarios obtained with given combinations of calibration parameters and we show how the proposed model can be used as an analytical framework to support the decision making.

The remainder of the chapter is organized as follows. In Section 2.2 the Italian national organization of BMS is described with the illustration of some data that underline the need of a reform process of the sector. In Section 2.3, a literature review about models and methods used in the context of blood supply chain management is provided, with a particular focus on the existing contributions related to facility location. Then, in Section 2.4, the description of the problem and its formulation is provided. In Section 2.5 the adaptation of the model to the specific regional case is shown while, in Section 2.6 the results provided by the model are analyzed and discussed. Finally, some conclusions are drawn.

2.2 Problem description

In this Section, we provide an analysis of the current organization of BMS in Italy and we discuss the challenges determined by the new regulatory framework. Afterwards, we focus our attention on the case of the Campania Region.

2.2.1 The organizational context of the Italian Blood Management System

In Italy, the BMS is public and part of the National Health System and it is organized according to some principles, similar to those inspiring the BMSs of other EU countries. In particular, the blood collection is performed on the basis of voluntary donations (anonymous and non-remunerated). The distribution of blood components is classified as "essential service"; hence, it has to be performed equitably, impartially, free of charge to any citizen and in such a way to guarantee safety and product quality. The organization of the BMS should be able to satisfy the demand of blood components both at a national and regional level (*self-sufficiency goal*). At this aim, each Regional Authority is free to position facilities over its territory.
In Table 1, for each Italian region, figures related to the total population (column a), the annual blood supply (column b) and the demand (column c), in terms of whole blood units collected and transfused, are reported (Italian Ministry of Health, 2016). Then, the self-sufficiency ratio, calculated as percentage ratio between blood supply and demand, is also reported in order to point out in what measure each region satisfies the self-sufficiency goal (percentage under 100% represents a blood shortage). Moreover, the blood collection and transfusion per capita are indicated; the first indicator, also named donation rate, can be considered a proxy of the propensity of population to donate blood, while the second one can be considered a proxy of the regional blood demand coming from hospitals and health centers. Finally, the average productivity of single BCs, defined as ratio between the total collected blood and the number of BCs (d), is also provided. Data highlight significant differences across regions and some critical aspects at national level. First of all, the system appears quite fragile, as the total supply covers the current demand with a very low surplus (less than 1%), which does not represent a reliable reserve to tackle stochastic or not-ordinary variation of the demand. In addition, there exists a

remarkable unbalance in terms of self-sufficiency across regions (three of them are not self-sufficient), which requires a transfer from regions with excess of blood collection. The differences in terms of transfusions per capita are not related to environmental or socio-economic factors but to a well- known and consolidated phenomenon of patients' migration from hospitals of southern regions towards the ones of northern regions, due to a different perceived quality of the offered health care services. Regional BMSs also reveal strong differences from the efficiency point of view, either in terms of collection (see donation rate values) or productivity (see average productivity per BC). These aspects (almost half of the regions have productivities lower than the average) suggests a twofold interpretation: on one hand, some regions are characterized by a redundant presence of BCs, on the other hand, they have a low capability of blood collection. In particular, as concerns Campania Region, whose BMS is the object of this study, despite a very low level of regional demand per capita (about the 50% of the average national value) it is characterized by a huge number of active BCs (22) with a very low productivity (more than 30% lower than the average national level).

The Italian government implemented the European provisions by releasing a set of national guidelines for Regional Authorities concerning the accreditation of transfusion services and units devoted to the collection of blood and its components (Agreement CSR 149/2012, Italian State-Region Conference, 2012). In particular, some efficiency measures were introduced, according to which consolidation of processing activities is recommended (i.e. blood testing, separation in blood components and transformation in plasma derived products) in a lower number of BCs, so as to guarantee a minimum productivity level per BC equal to 40,000 units of whole blood per year. As shown in Table 1, such target is significantly higher than the average productivity of BCs in each region and this circumstance suggests a reorganization of most of Regional Blood Management Systems. It is worthy to underline that the threshold of 40,000 units has to be intended valid only for those regions characterized by higher supply levels. For example, for regions like Molise and Basilicata, the target of 40,000 blood units has not to be intended as a strict constraint but, of course, they are required to concentrate processing activities in a single BC.

2.2.2 The re-organization of the Campania Blood Management System

In the described context, the motivation of this study arises from the project of reorganization of the BMS of the Campania region undertaken by the Regional Authority, which is responsible for the system's performance.

Campania is the second most populated region in Italy, with about six million of inhabitants (density of 425 inhabitants per km^2); then, it is characterized by high values of annual blood demand, estimated around 150,000 of whole blood cells. The current organization of the BMS is based on the presence of 22 BCs, spread over the region (Figure 2.1), just able to satisfy the self-sufficiency goal. This number is remarkably higher than the other regions (see Table 2.1). Figure 2.2 shows, in details, the productivity per BC. It is possible to notice that, among all the facilities, only BC 4 has a productivity comparable to the new standard of 40,000 blood units, while most of the others are

Region	Population (a)	Annual Collection (b)	Annual Transfusion (c)	No. of BCs (d)	Self sufficiency (b)/(c)	Collection per capita (b)/(a)	Transfusion per capita (c)/(a)	Production per BC
Valle d'Aosta	128,810	6,685	5,236	1	127.7	5.2	4.1	6,685
Piemonte	4,466,509	237,603	205,622	23	115.6	5.3	4.6	10,330
Liguria	1,613,710	74,920	74,520	9	100.5	4.6	4.6	8,324
Lombardia	10,006,710	485,673	470,006	36	103.3	4.9	4.7	13,490
PA di Trento	534,405	22,531	21,704	1	103.8	4.2	4.1	22,531
PA di Bolzano	512,446	25,158	24,060	1	104.6	4.9	4.7	25,158
Friuli	1,235,665	73,607	65,296	5	112.7	6.0	5.3	14,721
Veneto	4,960,336	264,221	247,588	20	106.7	5.3	5.0	13,211
Emilia Romagna	4,464,371	251,839	245,068	12	102.8	5.6	5,5	20,986
Toscana	3,763,076	190,494	188.838	16	100.9	5.1	5.0	11,905
Umbria	909,422	44,825	44,763	4	100.1	4.9	4.9	11,206
Marche	1,569,303	79,437	78,939	12	100.6	5.1	5.0	6,619
Lazio	5,786,715	196,779	227,069	26	86.7	3.4	3.9	7,568
Sardegna	1,674,169	77,367	114,069	12	67.8	4.6	6.8	6,447
Abruzzo	1,345,050	53,973	53,691	6	100.5	4.0	4.0	8,995
Campania	5,834,154	151,584	148,593	22	102.0	2.6	2.5	6,890
Molise	318,646	15,117	14,864	3	101.7	4.7	4.7	5,039
Puglia	4,086,644	161,312	158,755	17	101.6	3.9	3.9	9,489
Basilicata	585,615	28,861	24,751	6	116.6	4.9	4.2	4,810
Calabria	2,008,315	69,722	68,793	12	101.4	3.5	3.4	5,810
Sicilia	5,045,176	193,068	197,674	27	97.7	3.8	3.9	7,151
Total	60,849,247	2,704,776	2,679,899	271	100.9	4.4	4.4	9,981

Table 2.1. Data related to regional BMSs

characterized by values lower than 10,000 units.
As the as-is analysis reveals some inefficiencies, a project of re-organization was undertaken, whose main objectives were fixed, on the basis of the following general principles:

- reduction of the total number of the BCs, also considering the possibility to downgrade a BC in a facility that performs only collection activities (named Blood Station - BS), characterized by significant lower management costs; in this case each BS needs to be allocated to a BC where the collected blood should be processed;

- promotion of campaigns to increase the donation rate. This is motivated by twofold issues: (i) the current low level of the donation rate suggests the possibility of significant margins of increment; (ii) the planned reduction of BCs over the region may reduce the total collection capacity, putting at risk the achievement of the self-sufficiency condition;

- definition of scenarios compatible with possible increases of the total regional blood demand, currently characterized by a low average level of transfusions per capita. Such increase might occur in a medium-long term horizon, by assuming a mitigation of the patients' migration toward other regions.

These principles have been included in the mathematical formulation of the problem, that will be described in Section 2.4, after a brief overview of the literature background.

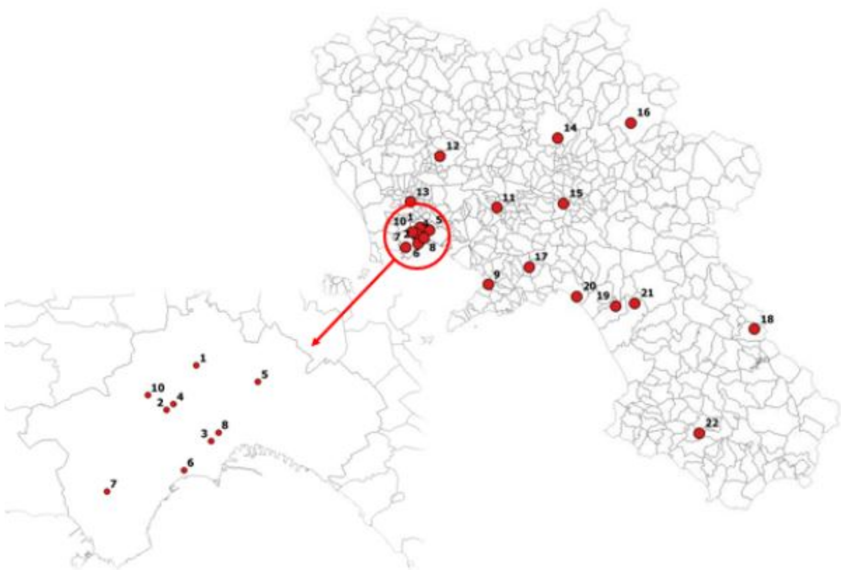


Figure 2.1. Current organization of the BMS in the Campania Region.

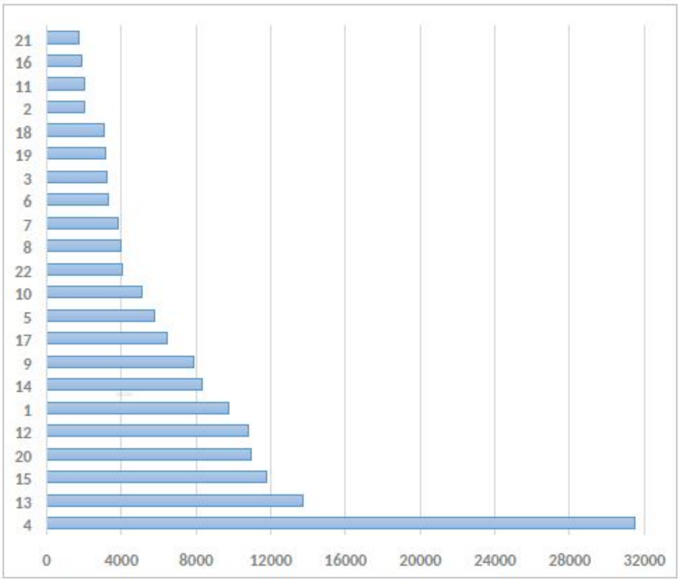


Figure 2.2. Productivity of BCs in the Campania Region.

2.3 Literature background

Scientific literature concerning blood supply chain management was initially focused on issues mainly related to inventory management (Prastacos, 1984) and only in the last two decades the interest was extended toward other aspects, as testified by Pierskalla (2005), Beliën and Forcé (2012) and Osorio et al. (2015). In particular, a significant number of contributions started to focus on outbound problems, i.e. on problems dealing with the distribution of blood products and components (Hemmelmayr et al., 2009, 2010; Delen et al., 2011), the analysis and forecasting of blood supply and demand (Glynn et al., 2002; Bosnes et al., 2005; Custer et al., 2005; Schreiber et al., 2006; Godin et al., 2007), the design and the optimization of the whole supply network (Katsaliaki and Brailsford, 2007; Nagurney et al., 2012) and the impact of innovation technologies on operations' performances (Butch, 2002; Li et al., 2008; Davis et al., 2009).
From a strategic point of view, the problem of optimally locating blood facilities is crucial, as such decisions may strongly affect logistic and operational performances of the whole supply chains. The class of models traditionally used in the Operational Research literature to address problems related to the territorial organization of private and public services is Facility Location Models – FLMs (Laporte et al., 2015). These models may be effectively employed in order to locate new facilities and/or to modify the configuration of existing facilities, by closing or relocating some of them and/or by downsizing/redistributing their operating capacities (Wang et al., 2003; ReVelle et al., 2007; Sonmez and Lim, 2012; Bruno et al., 2016a,b, 2017b).
Location problems have been widely applied in the context of health care sector; the most popular applications concerned the location of facilities like hospitals, preventive

care services, primary care and specialized centers (Daskin and Dean, 2005; Güneş and Nickel, 2015). Recently, Ahmadi-Javid et al. (2017) reviewed all the contributions related to the location of healthcare facilities, over the period 2004-2016. Among the identified research gaps, they explicitly included the extension of location models to the optimal design of blood supply networks. Indeed, despite the great importance of blood facilities within health systems, only a few papers analysed the problem of locating blood banks. Among them, it is possible to distinguish papers that deal with the design of efficient blood supply chains in ordinary conditions, i.e. able to face the ordinary demand at minimum cost (Şahin et al., 2007; Cetin and Sarul, 2009; Elalouf et al., 2015; Zahiri et al., 2015), from those ones aimed at designing emergency blood supply chains, able to be resilient under different disaster scenarios Jabbarzadeh et al. (2014); Fahimnia et al. (2017). Considering the scope of our work, we focus on the papers specifically devoted to the design of supply chain in ordinary conditions, in which the demand mainly comes from patients hospitalized within the region under consideration for surgery purposes or specific deseases. Within this class, Cetin and Sarul (2009) addressed a multi-objective model to locate a set of blood banks and determine the allocation of hospitals, so as to minimize the total cost for locating facilities and for the distribution of blood derivatives towards hospitals. Nevertheless, well-consolidated and valuable approaches to tackle the problem reside in the design of hierarchical models (Ortiz-Astorquiza et al., 2017). Şahin et al. (2007) considered a hierarchical structure, made up of four different levels; at the highest one, regional blood centers (RBCs) perform all functions related to blood transfusion (collection, testing, processing, storage and distribution) as well as peculiar coordination functions of the lower-level units; at a lower level, blood centers (BCs) perform the same core transfusion activities by RBCs with no coordination role and, finally, blood stations (BSs) and mobile units (MUs) support collection activities. The authors decomposed the overall design problem into three sub-problems. In the first one, a two-level uncapacitated facility location model was formulated to identify the optimal locations of q RBCs and p BCs, with the aim of minimizing the sum of the weighted distances between demand points (donors) and facilities and between facilities themselves. Such model was applied at national level to regionalize the blood management system in Turkey, composed of 23 existing blood centers. In taking such decisions, blood supply, coming from donors, was aggregated in 81 demand points, corresponding to Turkish provinces, with no realistic consideration about distances between donors and facilities and, hence, their interaction. Moreover, decisions concerning lower-level facilities (location of BSs and assignment of MUs) were taken through subsequent and independent problems.

Elalouf et al. (2015) considered a three level structure, made up of: *clinics*, where blood is collected, *centrifugation centers*, where it is separated into components and a *centralized testing laboratory*, were blood samples are analysed. In the initial configuration, each clinic was assigned to a centrifugation center, while the model evaluated the possibility of equipping clinics with machines for in-house centrifugation and it decided the optimal configuration, in such a way that a profit function was maximized under a budget constraint. In this case, donors were not considered as the position of their access

point to the network did not change (clinics).
Zahiri et al. (2015) addressed the problem of locating fixed and temporary blood facilities over a multi-period horizon at the minimum cost, in order to meet a given level of blood demand per period. Fixed facilities were located at the beginning of the time horizon and their position could not be changed anymore; while temporary facilities could be relocated in each time period. Authors considered donors willing to cover different covering radii to patronize the considered types of facility and a further radius was introduced to regulate the allocation mechanism of temporary to fixed facilities. From a practical point of view, the problem was tested on instances, based on real data, composed of a limited number of candidate locations (six and nine for fixed and temporary facilities,respectively) and donors groups (14). Hence, its applicability is limited to address local management problems, at the level of cities or small regions.
The model we propose in this chapter deals with the reorganization of regional blood management systems in Italy, under ordinary conditions. In particular, each regional system, currently made up of blood banks performing all the activities related to transfusion, has to be reorganized according to a 2-levels hierarchical structure, in which the lower-level facilities (BSs – blood stations) have to perform only collection activities, while the higher-level ones (BCs – blood centers) have to perform both blood collection and processing. Similarly to Zahiri et al. (2015), collection is considered capacitated and different covering radii are considered, one for the allocation of donors to both type of facilities and another for the assignment between BSs and BCs. The above features renders out model de facto attributable also to the class of hierarchical covering problems with set covering objectives, i.e. oriented to costs minimization (Marianov and Serra, 2001). In particular, since we guarantee a strong relationship between the catchment areas of BCs and BSs, we also seek to design a coherent network structure (Serra, 1996). The original contribution of our formulation concerns the assumptions related to the processing activities. Indeed, besides the maximum capacity, also a minimum efficiency requirement is introduced, in terms of minimum amount of blood units, that represents the driving force of the real reorganization process being considered. Moreover, we also introduce decisions regarding the dimensioning of the processing capacity, which may also be extended at an extra-cost, through the establishment of additional modules, in order to achieve economies of scale at production sites. From the practical point of view, in the considered application donors are aggregated at level of single municipalities, thus obtaining an effective representation of their real distribution throughout the region and of the interaction with the facilities to be located. For this reason, the scenarios provided by our model may provide useful indications to the decision maker. By summarizing, the contribution that this work aims at giving to the literature is twofold. On one side, we introduce a new modeling framework in the context of the design of two-level capacitated blood supply chain, that considers minimum and maximum capacity for processing activities, and modular capacities. Moreover, a distinguishing feature of our model is the consideration of the self-sufficiency goal. On the other, we show, by means of several computational experiments, the capability of the introduced model to handle real-world instances and provide interesting managerial implications. The last,

but not the least, the possibility to solve the introduced model with a commercial solver (CPLEX) in limited computational times, can represent a valuable element for all the practitioners interested in this kind of problems.

2.4 A mathematical model for the reorganization of a regional BMS

In this section we provide a more formal description of the problem. In particular, in Section 2.4.1 we discuss the hypotheses we rely on to formulate the mathematical model, which is then presented in Section 2.4.2.

2.4.1 Model assumptions

With reference to a given region, with a yearly demand D of whole blood to be collected, we suppose the presence of a set J of existing facilities that perform both blood collection and processing activities (BCs). Moreover, we consider a set I of discrete nodes where potential donors are located and we assume that each donor is willing to reach a collecting facility if and only if its distance from the closest one is within a given distance r.

According to guidelines released by the national government, facilities performing processing activities are required to meet a minimum productivity target P_{min}, in terms of processed blood per year. Assuming that, in the current configuration, the system does not satisfy such efficiency requirement, we consider a possible reorganization aimed at consolidating processing activities in a lower number of facilities, while maintaining a widespread collection over the study region. In this process each existing facility can be (i) kept open as BC, (ii) downgraded in a BS, by deactivating blood processing and performing solely collection activities, or alternatively (iii) it can be closed and not used anymore. Of course, blood collected at each BS has to be moved to an active BC, to be processed. Due to some technical constraints related to the degradation phenomenon of blood properties, such allocation is assumed to be feasible if and only if the distance between a BS and its assigned BC is within a given radius d_{max}. Moreover, each BS can be assigned to a single BC, thus avoiding the possibility of splitting its collection among different BCs, as, from a managerial point of view, it is desirable to define the cluster of BSs under the competence of each single BC, in order to better coordinate the associated flows of materials and information.

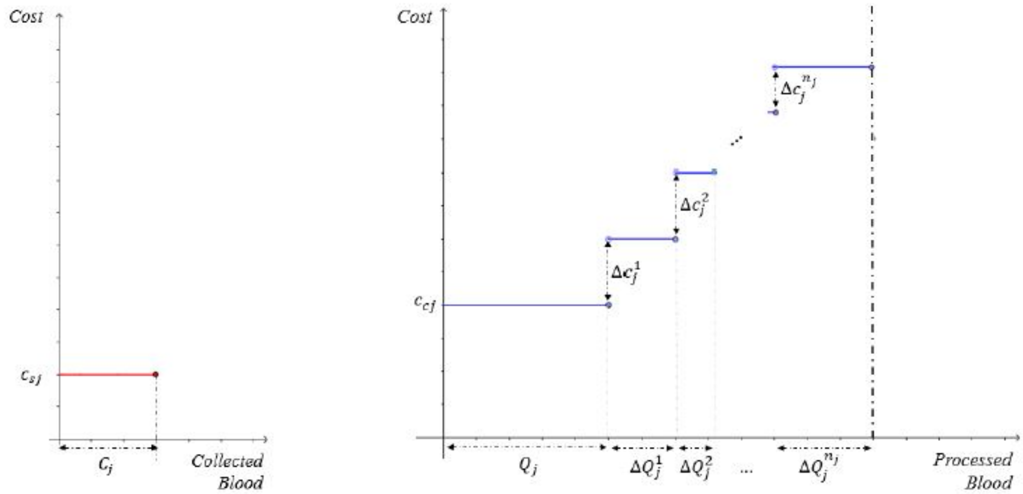
Then, the problem consists in locating, among the existing facilities, BCs and BSs with the aim of satisfying a set of requirements in terms of total collected blood (*self-sufficiency goal*) and productivity per single BC (*efficiency goal*). Among all the feasible alternatives, the model will select the configuration that minimizes the operational costs per year.

As concerns the evaluation of costs to be incurred in the final configuration, we assume that both collection and processing activities are *capacitated*; for this reason, each fa-

cility $j \in J$ is characterized by a collection capacity C_j and a processing capacity Q_j, indicating the maximum number of units that, respectively, it is able to collect and process per year. We assume that, while the processing capacity of each BC may be gradually extended by discrete amounts or modules of capacity, no capacity expansion is allowed for collection activities. This assumption seems to be realistic as, on one side, the decision maker is interested in the consolidation of processing activities in a lower number of facilities, to meet the above described efficiency goal. Hence, within certain limits, he may explore the possibility to reallocate (human and material) resources from closed facilities or BSs towards larger BCs, in order to exploit possible economies of scale. On the contrary, as concern collection activities, consolidation could not be an appropriate strategy, as the decision maker will most likely need to distribute facilities throughout the region, with the aim to attract a consistent number of donors and to meet the self-sufficiency goal.
In order to include this aspect in the model, we indicate with n_j the maximum number of modules that can be added to expand the processing capacity Q_j of the generic facility j and with ΔQ_j^k the size of the k-th module that can be added (with $k = 1, ..., n_j$). Consequently, the cost associated to a BC is modeled through a step function, in which, starting from the initial cost c_cj (to be incurred for an initial setting of the collection capacity C_j and the processing capacity Q_j), an additional cost δc_j^k is incurred whenever a module ΔQ_j^k of processing capacity is introduced, while the cost associated to a BS is identically equal to c_{sj}, within the limit of capacity C_j. These assumptions find firm roots in the literature. As it will be clarified later, we consider in the model only the *avoidable costs*, i.e. the ones that the BMS can save or avoid as a result of reorganization decisions. According to Telser (1978), the plant's avoidable cost is typically identified with the variable costs associated to a certain output level, which can be considered fixed within certain limits of capacity. As a consequence, examples of stepped fixed avoidable costs, in line with the above introduced cost structure, are found in Van Boening and Wilcox (1996) and Telser (2005). In Figure 2.3, a representation of the introduced cost functions is provided. Finally, also the transportation cost to transfer collected blood from BSs to BCs is considered and it is assumed proportional to the covered distances.

2.4.2 Model formulation

On the basis of the above described assumptions, we introduce the following notation in order to formulate the model:



(a) BS **(b)** BC

Figure 2.3. Cost functions

.

I	set of nodes where potential donors are located;
J	set of nodes where existing facilities are located;
a_i	amount of donations potentially coming from node $i \in I$;
d_{ij}	distance between nodes $i \in I$ and $j \in J$;
r	maximum distance that donors are willing to cover to reach a collection site (*BS covering radius*);
N_i	set of nodes $j \in J$ whose distance from $i \in I$ is lower than r ($N_i = \{j \in J : d_{ij} \leq r\}$);
$d_{jj'}$	distance between facilities $j, j' \in J$;
d_{max}	maximum feasible distance between a BS and its assigned BC (*BC covering radius*);

M_j set of nodes $j' \in J$ whose distance from $j \in J$ is lower than d_{max} ($M_j = \{j' \in J : d_{jj'} \leq d_{max}\}$);
D total regional demand of whole blood;
P_{min} minimum amount of whole blood units to be processed by each BC;
C_j collection capacity, in terms of whole blood units, of facility $j \in J$;
Q_j processing capacity, in terms of whole blood units, of facility $j \in J$ ($Q_j > P_{min}$);
n_j maximum number of processing capacity modules that can be added to facility $j \in J$;
ΔQ_j^k size of the k-th processing capacity module that can be added to facility $j \in J$;
c_{sj} cost to be incurred if a BS is active at node j;
c_{cj} cost to be incurred if a BC is active at node j, with a processing capacity equal to Q_j;
δc_j^k extra-cost for expanding the processing capacity of a BC in j through the addition of the k-th module ΔQ_j^k ($k = 1, ..., n_j$);
ϵ unit transportation cost, i.e. cost per unit of covered distance to transfer the blood collected at a given BS toward its assigned BC.

The decision variables are as follows:

y_{sj} binary variables equal to 1 if and only if a BS is active in the node j in the final configuration (0 otherwise);
y_{cj} binary variables equal to 1 if and only if a BC is active in the node j in the final configuration (0 otherwise);

x_{ij} binary variables equal to 1 if and only if all the potential donations from node i are assigned to the collection site j (0 otherwise);
$q_{jj'}$ binary variables equal to 1 if and only if all the amount of blood collected by facility in j is processed by the facility in j' (0 otherwise);
z_j^k binary variables equal to 1 if and only if the processing capacity of facility j is expanded through the addition of k-th capacity module (0 otherwise).

Then, the Integer (Binary) Programming model is formulated as follows:

$$\text{minimize} \quad \sum_{j\in J} c_{sj}y_{sj} + \sum_{j\in J} c_{cj}y_{cj} + \sum_{j\in J}\sum_{k=1}^{n_j} \Delta c_j^k z_j^k + \epsilon \sum_{j\in J}\sum_{j'\in J} d_{jj'}q_{jj'} \tag{2.1}$$

$$\text{subject to} \quad y_{sj} + y_{cj} \leq 1 \qquad j \in J \tag{2.2}$$

$$x_{ij} \leq y_{sj} + y_{cj} \qquad i \in I, j \in J \tag{2.3}$$

$$\sum_{j\in(J\setminus N_i)} x_{ij} = 0 \qquad i \in I \tag{2.4}$$

$$|N_i| \sum_{j\in N_i} x_{ij} \geq \sum_{j\in N_i} (y_{sj} + y_{cj}) \qquad i \in I \tag{2.5}$$

$$\sum_{j'\in N_i} d_{ij'}x_{ij'} + (F - d_{ij})(y_{sj} + y_{cj}) \leq F \qquad i \in I, j \in N_i \tag{2.6}$$

$$q_{jj'} \leq y_{cj'} \qquad j, j' \in J \tag{2.7}$$

$$\sum_{j\in M_j} q_{jj'} = y_{sj} + y_{cj} \qquad j \in J \tag{2.8}$$

$$q_{jj} \geq y_{cj} \qquad j \in J \tag{2.9}$$

$$\sum_{i\in I}\sum_{j\in J} a_i x_{ij} \geq D \tag{2.10}$$

$$\sum_{i\in I} a_i x_{ij} \leq C_j \qquad j \in J \tag{2.11}$$

$$\sum_{j'\in J} (\sum_{i\in I} a_i x_{ij'}) q_{jj'} \geq P_{min} y_{cj} \qquad j \in J \tag{2.12}$$

$$\sum_{j'\in J} (\sum_{i\in I} a_i x_{ij'}) q_{jj'} \leq Q_j y_{cj} + \sum_{k=1}^{n_j} \Delta Q_j^k z_j^k \qquad j \in J \tag{2.13}$$

$$z_j^{k+1} \leq z_j^k \qquad j \in J, k = 1, ..., n_j \tag{2.14}$$

$$y_{sj}, y_{cj}, z_j^k, x_{ij}, q_{jj'} \in \{0, 1\} \qquad i \in I, j, j' \in J, k = 1, ..., n_j \tag{2.15}$$

The objective function (2.1) represents the total operational cost associated to the active facilities in the final configuration. In particular, it is given by the sum of the costs of active BSs ($\sum_{j\in J} c_{sj}y_{sj}$) and BCs ($\sum_{j\in J} c_{cj}y_{cj} + \sum_{j\in J}\sum_{k=1}^{n_j} \Delta c_j^k z_j^k$), including potential extra-costs incurred for the additional capacity modules. The last term of the objective function ($\epsilon \sum_{j\in J}\sum_{j'\in J} d_{jj'}q_{jj'}$) represents the transportation cost to transfer blood units from BSs to BCs.

Constraints (2.2) assure that a single type of facility can be located in each node $j \in J$ i.e., if a BS is active in j ($y_{sj} = 1$), a BC cannot be active in the same node ($y_{cj} = 0$) and viceversa.

Constraints (2.3)–(2.6) reproduce the allocation of donors to active facilities. Constraints (2.3) impose that donors can be assigned to any node j where a BS or a BC is activated in the final configuration ($y_{sj} + y_{cj} = 1$), as both types of facilities perform collection activities. Constraints (2.4) avoid the allocation of any node i to facilities j farther than the

covering radius r ($j \notin N_i$). Constraints (2.5) impose that each node i has to be assigned at least to one facility, if there exists at least an active BS or BC within its covering radius r ($\sum_{j \in N_i}(y_{sj} + y_{cj}) \geq 1$), while Constraints (2.6) impose that node i is allocated, within the radius r, to its closest active facility. Indeed, by setting $F = max_{i \in I, j \in J}\{d_{ij}\}$, Constraints (2.6) become redundant for all those nodes in which no facility is located ($j \in N_i : y_{sj} + y_{cj} = 0$) while, in relation to the active facilities, assign each node i to the node j' at the minimum distance.
Constraints (2.7)–(2.9) reproduce the allocation of BSs to BCs. In particular, Constraints (2.7) assure that a node $j \in J$ can be assigned only to facilities $j' \in J$ where a BC is open ($y_{cj} = 1$); on the other hand, Constraints (2.8) impose that any node j where a BS or a BC is activated in the final configuration ($y_{sj} + y_{cj} = 1$) has to be assigned to one and only one facility within the maximum feasible distance d_{max} ($j' \in M_j$). Finally, Constraints (2.9) guarantee that each BC is assigned to itself.
Constraints (2.10) assures that the whole BMS collects an amount of blood units higher (or equal) than the regional blood demand D (self-sufficiency goal). Constraints (2.11) represent capacity requirements for collection activities. Constraints (2.12)–(2.14) concern the productivity of BCs. In particular, Constraints (2.12) impose that the total amount of processed blood by each BC j exceed the minimum target P_{min}. Constraints (2.13) impose that the total amount of blood units processed by each BC j is lower (or equal) than its capacity Q_j, potentially expanded with additional capacity modules ($\sum_{k=1}^{n_j} \Delta c_j^k z_j^k$). Constraints (2.14) assure that the $(k+1)$-th capacity module may be activated at a BC if and only if the k-th module has been already activated.
Finally, Constraints (2.15) define the domain of the introduced decision variables.
Model (2.1)–(2.15) is not linear, due to the presence of constraints (2.12)–(2.13). They have been linearized by introducing a new set of binary decision variables $w_{ijj'}$ ($w_{ijj'} = x_{ij'} q_{j'j}$) and the following set of constraints:

$$w_{ijj'} \leq x_{ij} \qquad i \in I, j, j' \in J \qquad (2.16)$$
$$w_{ijj'} \leq q_{jj'} \qquad i \in I, j, j' \in J \qquad (2.17)$$
$$w_{ijj'} \geq x_{ij} + q_{jj'} - 1 \qquad i \in I, j, j' \in J \qquad (2.18)$$

Constraints (2.16)–(2.18) guarantee that the new binary variables $w_{ijj'}$ are equal to 1 if and only if donations from i are assigned to facility j ($x_{ij} = 1$) and such blood is moved to j' to be processed ($q_{jj'} = 1$). This way the model may be formulated as follows:

$$\begin{aligned} &\text{minimize} && (2.1) \\ &\text{subject to} && (2.2)-(2.11),(2.14),(2.16)-(2.18) \\ & && \sum_{j'\in J}\sum_{i\in I} a_i w_{ijj'} \geq P_{min} y_{cj} \qquad j \in J && (2.19) \\ & && \sum_{j'\in J}\sum_{i\in I} a_i w_{ijj'} \leq Q_j y_{cj} + \sum_{k=1}^{n_j} \Delta Q_j^k z_j^k \qquad j \in J && (2.20) \\ & && y_{sj}, y_{cj}, z_j^k, x_{ij}, q_{jj'}, w_{jj'} \in \{0,1\} \qquad i \in I, j, j' \in J, k = 1, ..., n_j && (2.21) \end{aligned}$$

In the next section, we analyze the application of the model to the case study of Campania Region, described in Section 2.2.

2.5 The case study of the Campania Region

As introduced above, we apply the proposed model to the BMS of the second most populated region in Italy (Campania), that is currently characterized by the presence of 22 facilities over the territory, with an average productivity level significantly lower than the minimum threshold P_{min} imposed by the EU and national regulations (see Figures 2.1 and 2.2).

In order to apply the model to the selected case study, we adopt a discretization of the study area, based on the territorial subdivision of the region in municipalities (551). Donations coming from each municipality per year are assumed to be proportional to the living population, according to a coefficient α (*donation rate*), that is considered uniformly distributed within the region. Such parameter should not to be interpreted as percentage of donors but as the number of yearly donations over the total population. According to such assumption, without loss of generality, it can be considered that each donor accounts for a single donation and no further assumptions about donors' habits and behaviors are required. In particular, from the analysis of historical data provided by the regional authority about the yearly regional donations, it emerged that almost 35 donations every 1000 inhabitants were observed in the last 5 years ($\alpha = 0.0350$).

Figure 2.4 shows the adopted zoning system along with the distribution of the population among municipalities (ISTAT, 2011). As the municipality of Naples, i.e. the regional chief town, is characterized by a population level much higher than the others (almost 962,000 inhabitants) and by the presence of nine BCs (Figure 2.1), it has been in turn subdivided in 10 sub-areas, thus obtaining 560 Territorial Units (TUs). Potential donors from each municipality are concentrated in a single node, corresponding to the centroid of the municipality itself. The distances between the 560 centroids and the 22 existing facilities and the distances between each pair of facilities are calculated as shortest routes (in km) on the road network (considering motorways, national and regional roads).The BC covering radius d_{max}, i.e. the maximum distance between each active BS and its assigned BC, are estimated considering the technical constraints related to the degradation phenomenon of blood properties, according to which the time elapsing from the collection and the processing time cannot exceed five hours. On the basis of such time limit and some assumptions on the possible transportation strategies indicated by the decision maker, it is reasonably to assume a maximum distance d_{max} almost equal to 40 km.

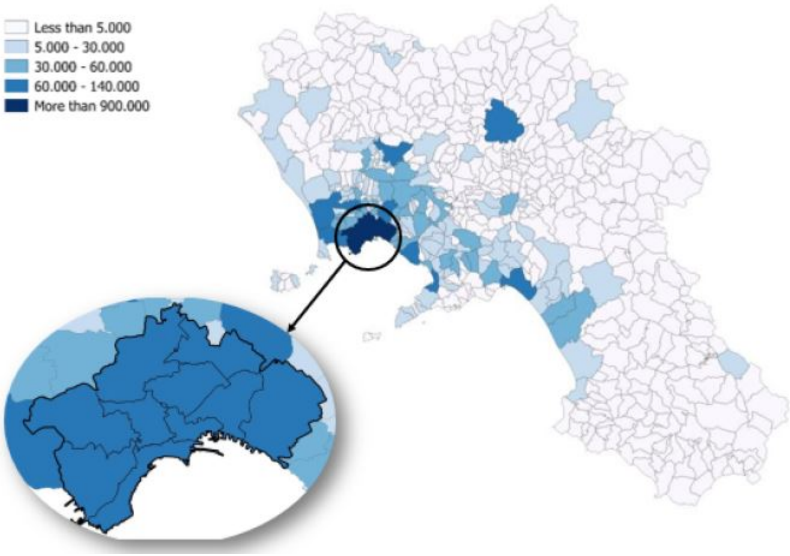


Figure 2.4. Zoning system and population distribution among TUs.

The other parameters characterizing the BMS (i.e., capacities, costs, etc.) have been estimated through an analysis conducted in collaboration with the Regional Authority. Currently, all the existing facilities are able to collect and process almost 30,000 and 50,000 blood units, respectively, per year ($C_j = C = 30,000, j \in J; Q_j = Q = 50,000, j \in J$). The processing capacity is constrained by the amount of employed *human resources* (i.e. doctors and technicians), that are completely saturated, while *equipment* presents a residual capacity of almost 30,000 blood units per year. For this reason, we assume that the processing capacity of each facility may be eventually expanded with a module ($n_j = n = 1, j \in J$) of 30,000 units ($\Delta Q_j^1 = \Delta Q^1 = 30,000, j \in J$), obtained by providing it with additional human resources that allow it to fully utilize the residual equipment capacity.
As regard costs, we consider in the model only the avoidable costs, i.e. the ones that the BMS can save or avoid as a result of reorganization decisions. In particular, as concern collection activities, the parameter c_{sj} includes costs for *human resources* (i.e. for example, *administrative staff*, that accepts and registers donors; *doctors*, that check donors' health conditions; *nurses*, that perform blood draws), as they can be saved by the BMS if the collection activity is not performed anymore at a given location. The costs for medical devices (i.e., needles, reagents, etc.) are considered *unavoidable* as they depend on the total amount of collected blood at regional level, regardless of the location where activities are performed. On the other side, as concern processing activities, the avoidable costs are represented by the costs incurred for *human resources* (i.e. doctors

and technicians) and *equipment*. As the two types of activities do not share any resource, the operating cost of a BS corresponds to the cost to be incurred for collection activities, while the cost of BC to the sum of costs for both the activities; moreover such costs have been assumed to be equal among them and independent from the specific location ($c_{sj} = c_s; c_{cj} = c_c, j \in J; c_c \approx c_s + c_{processing}$). Since the avoidable costs incurred for processing activities are almost double the costs required for the collection phase, we set $c_{processing}$ and $c_s = 1$. Details about these costs are provided in Appendix A - Table .2. Finally, as the employee cost accounts for 80% of the total cost of processing activities, the cost for the activation of an additional module of capacity at any facility has been fixed equal to 0.80 ($\Delta c_j^1 = \Delta c^1 = 0.80 \times c_{processing}, j \in J$). These assumptions are also in line with the results provided by Santini et al. (2013). The unit transportation cost ϵ has been fixed equal to 0.001.
Moreover, in accordance with the regional authority, we have varied four calibration parameters, that are associated to peculiar strategies that may be undertaken in a medium-long term horizon:

- the BS covering radius r, in order to reproduce the effects of a more efficient and widespread collection in the region, thanks to the use of mobile units managed by voluntary associations (r = 20, 25, 30 km);

- the BC covering radius d_{max}, in order to evaluate the effects of different transportation strategies of collected blood from BSs towards their assigned BCs (d_{max} = 40, 50 km);

- the donation rate α, in order to simulate the effects of potential regional campaigns aimed at increasing the propensity of population to donate blood (α = 0.0350, 0.0375, 0.0400);

- the total regional blood demand D, in order to reproduce the effects of potential regional campaigns aimed at fighting the above mentioned phenomenon of patients' migration toward other regions, that would produce an increasing of blood demand coming from the hospitals within the region (D = 150,000, 165,000, 172,500, 180,000).

The test problems, generated by combining the calibration parameters above introduced, have been solved using a commercial solver (CPLEX 12.8) on an Intel(R) Celeron(R) with 1.50 GHz and 4 GB of RAM. Through a linkage with a GIS (Geographic Information System), each provided scenario has been represented on a map, by reporting information about the position of the operating facilities, their classification in BSs and BCs, the attractiveness area of each facility in terms of donors and the assignment of BSs to BCs for the blood transformation. In the next section, the scenarios produced by the model are introduced and analyzed.

2.6 Analysis of the results

In this section, a deep analysis of the scenarios provided by the model is reported. First of all, examples of scenarios are shown and analyzed in detail in section 2.6.1; then, all the solutions, produced by varying the above introduced calibration parameters, will be compared on the basis of a set of key indicators (section 2.6.2). Finally, a post-optimality analysis will be reported, aimed at evaluating the robustness of solutions, with reference to cost parameters (2.6.3) and the impact of the reorganization decisions on the distribution costs towards hospitals (2.6.4).

2.6.1 Illustrative examples

Figure 2.5 and Table 2.2 show a first scenario produced by the model, by setting $r = 20$ km, $D = 150{,}000$, $\alpha = 0.0350$. In this case, the model activates, among the 22 existing facilities, three BCs and five BSs, while the remaining 14 facilities are closed. TUs are grouped according to collection sites (BS or BC) they are assigned to; the borders of the districts thus obtained are represented with double edge in the figure. Moreover, as each BS is assigned to a single BC, the seven districts are then grouped in three clusters, reported with different colors in the map.

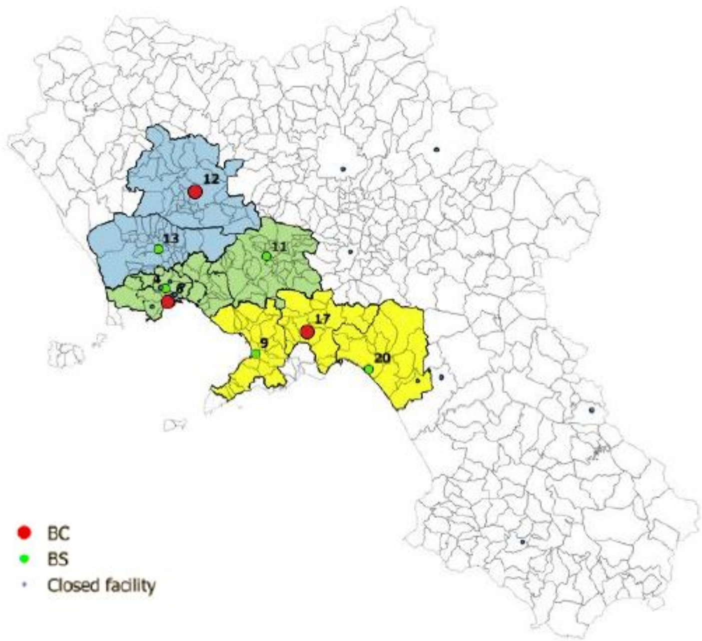


Figure 2.5. Map of Scenario 1 ($r = 20$ km, $D = 150{,}000$, $\alpha = 0.0350$, $d_{max} = 40$ km)

In Table 2.2 the total amount of blood collected by each active facility is reported; moreover, for each BC, an indication of its assigned facilities is provided along with the total amount of processed blood, given by the sum of the collection performed by the single facilities within the related cluster. For example, blood processed by BC 17 is

BCs		6		13			17		Total
BSs	4	6	11	12	13	9	17	20	
Collection	29,866	26,211	12,714	14,120	26,979	18,919	10,696	10,544	150,049
Processing		68,791		41,099			40,159		

Table 2.2. Characteristics of Scenario 1.

given by the sum of the amounts collected by BSs 9 and 20 and by BC 17 itself. It is interesting to notice that the self-sufficiency goal is met as the total collected blood is equal to 150,049 and that each BC satisfies the minimum productivity requirement of 40,000 blood units, with a single BC (6) requiring an expansion of the initial capacity. In Figure 2.6, a second scenario obtained by varying the BC covering radius d_{max} from 40 to 50 km is shown. Compared with scenario 1, this solution is characterized by the same number of active facilities (equal to 8) but a lower number of BCs. This is reasonable since the active BCs are now able to cover farther BSs and thus attract higher amounts of blood. In this situation, the model chooses to consolidate processing activities in a lower number of BCs, as this produces economies of scale and a consequent reduction of the objective function. As it can be noticed from Table 2.3, both the active BCs require the activation of the additional capacity module, as the production levels are in both cases close to the maximum capacity of 80,000 units (71,251 and 79,533 for BC 13 and 17, respectively).

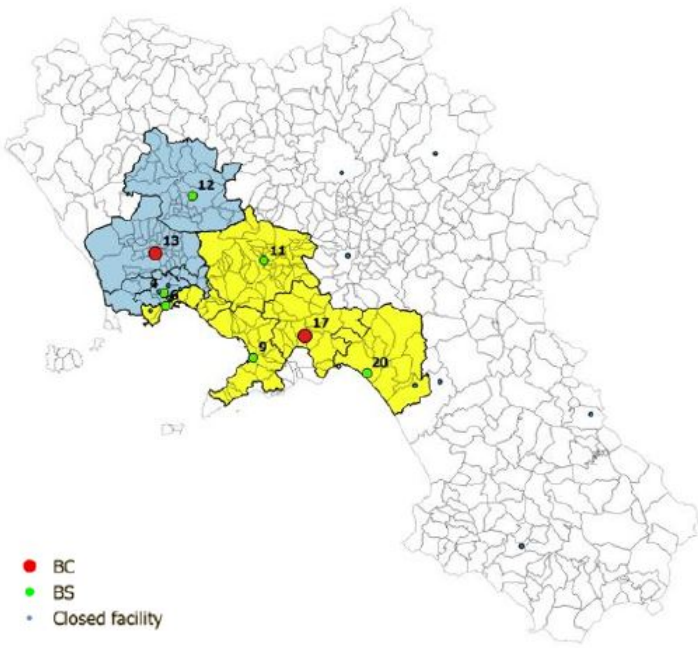


Figure 2.6. Map of Scenario 2 (r = 20 km, D = 150,000, α = 0.0350, d_{max} = 50 km)

Moreover, also a third scenario is shown in Figure 2.7, obtained by varying the BS

BCs		13				17			Total
BSs	4	12	13	6	9	11	17	20	
Collection	29,866	14,120	27,264	26,211	18,919	13,164	10,696	10,544	150,784
Processing		71,251				79,533			

Table 2.3. Characteristics of Scenario 2.

covering radius r from 20 to 30 km. In this case, the main difference with scenario 1 consists in a significant reduction of the total number of active facilities, from eight to six, due to the fact that facilities are able to attract donors from farther TUs and, consequently, to collect higher amounts of blood. As it can be noticed from Table 2.4, the amounts of collected blood by single facilities are significantly higher than the values reported in Table 2.1, with an average increase of almost 30%.

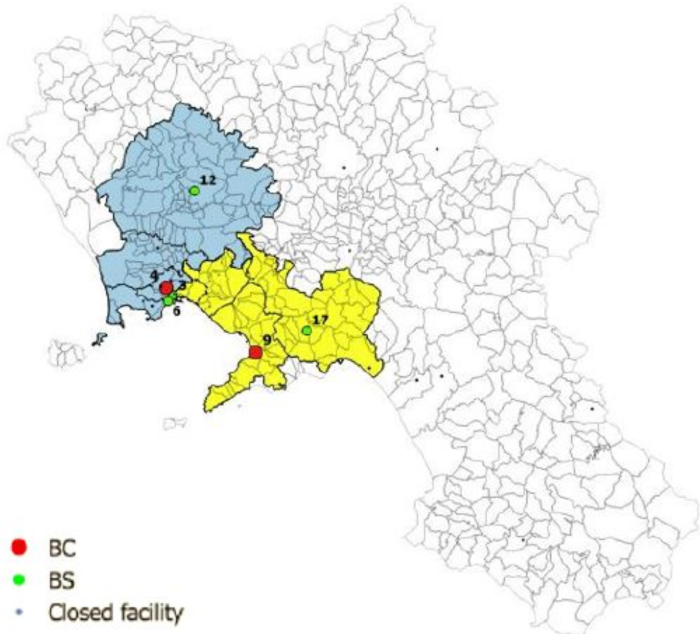


Figure 2.7. Map of Scenario 3 (r = 30 km, D = 150,000, α = 0.0350, d_{max} = 40 km)

BCs		4			9		Total
BSs	4	6	12	3	9	17	
Collection	28,802	26,351	23,383	28,931	22,981	19,859	150,308
Processing		78,536			71,772		

Table 2.4. Characteristics of Scenario 3.

2.6.2 Extensive results

The solutions obtained for any combination of calibration parameters (α, D, r, d_{max}) are compared, in Figure 2.8, in terms of active facilities (BSs and BCs). In particular, values obtained for d_{max} equal to 40 and 50 km are reported on the top and in the bottom of the graph, respectively.
It can be noticed that only 50 out of the 72 generated tests provide feasible solutions. The comparative analysis of the feasible scenarios may provide useful information and insights to decision maker. By fixing all the other calibration parameters α, D, d_{max} the total number of active facilities (BSs and BCs) tends to decrease by increasing the covering radius r. Indeed, this way each facility is able to attract donors from wider areas and, then, less facilities could be needed to meet the self-sufficiency goal. For example, with $D = 150,000, \alpha = 0.0350$ and $d_{max} = 40$ km, the model reduces the total number of active facilities (from eight to six) when r increases from 20 to 30 km. In particular, since BCs are able to attract more donors by themselves, the model tends to reduce the number of BSs as no need of activating so many dedicated collection sites is there anymore. Notice how, for $D = 165,000, \alpha = 0.0375$ and $d_{max} = 40$ km, the number of BSs tends to decrease when r increases, while the number of BCs remains equal to three. The same trend may be observed when the donation rate α or the BC covering radius d_{max} increase, being all the other values of calibration parameters equal. In the first case, this is due to an increased level of donations within the same covered areas. In the second case, the motivation is related to the capacity of BCs to cover farther BSs. In particular, this latter aspect has twofold implications. On one side, processing activities can be concentrated in a lower number of BCs; see, for example, the scenario obtained for $(\alpha, D, r) =$ (0.0400, 180,000, 25), in which the number of BCs decreases from four to three by increasing d_{max} from 40 to 50 km. On the other side, an increased value of d_{max}, may produce a reduction in BSs, as they can be more dispersed among each other and cover more effectively donors in the study region; see, for example, the scenario obtained for $(\alpha, D, r) =$ (0.0400, 172,500, 30), in which the number of BSs decreases from five to four by increasing d_{max} from 40 to 50 km while the number of BCs remains equal to three.
Finally, the number of active facilities tends to increase, by increasing the total regional demand D, both in terms of BSs and BCs, due to the presence of capacity constraints.

In Figure 2.9, the obtained scenarios are compared on the basis of the production levels of the active BCs. In particular, this representation allows the decision makers to understand, for each solution, how many BCs require the addition of a capacity module of 30,000 units and how many do not. Moreover, the distances of such points from the two lines, corresponding to $Q =$ 50,000 and $Q =$ 80,000, give also an indication about the saturation rate of such capacities. For example, notice that the solution obtained for $(\alpha, D, r) =$ (0.0400, 172,500, 25) presents 3 BCs, 2 of which with an expanded capacity that is underutilized; while the solution with an increased radius ($r = 30$ km) presents the same number of BCs but with a completely different distribution of production levels; in this case, only one additional module is activated and almost saturated, as well as the capacities of the other two BCs.

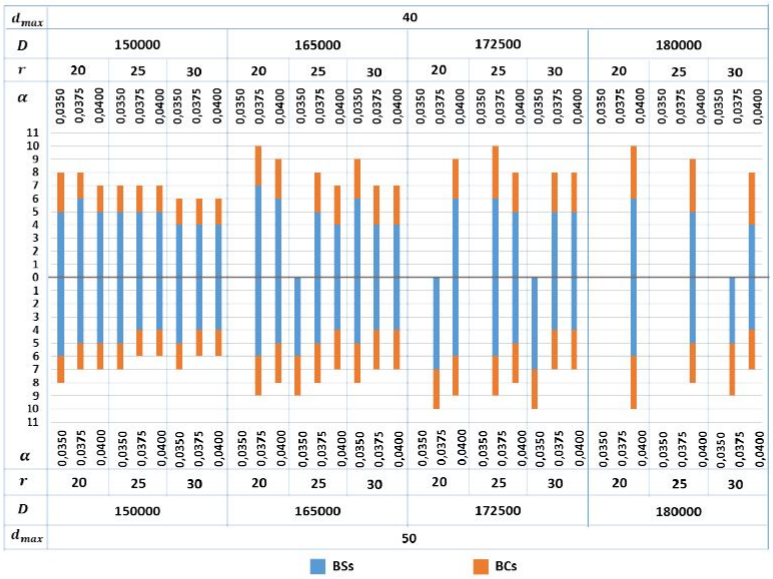


Figure 2.8. Nubmer of activated facilities in the produced scenarios

Extensive details about the overall computational experience are provided in Appendix C - Table .4.

The main objective of the proposed model is to support the decision maker by providing a set of possible feasible scenarios (according to the variation of some parameters), that can be evaluated and compared on the basis of a set of appropriate performance indicators. Among all the feasible obtained solutions, the decision maker should be aided at selecting the most appropriate scenario, by tuning the calibration parameters, according to his priorities and expectations.

As it emerges from results depicted in Figure 2.8, for instance, choosing the lowest value of D (the lower bound for the self-sufficiency goal), solutions with lower number of active facilities can be identified, so obtaining lower cost scenarios. However, these solutions are not able to capture higher levels of potential demand without significantly increasing the value of α (donation rate) and/or the value of r (covering radius of donors). On the other hand, if the self-sufficiency aspect is considered a main priority, by setting higher values of D (for instance 162,500 or 175,000), feasible solutions can be found also with a limited increase of active facilities and of the values of other parameters. In general, a key finding that clearly emerges from the above computational experiments is the crucial role played by the donation rate α. Indeed, a higher donation rate allows to obtain a wider range of feasible solutions, regardless of any possible setting of the other characteristic parameters of the model. For this reason, a consolidated recommendation to decision-makers would be to increase, especially locally, the propensity of population to donate blood more than putting effort in capturing donors from larger areas. Clearly, the combined achievement of both goals would foster very efficient solutions in terms of total number of located facilities. Hence, the choice of the final scenario cannot be

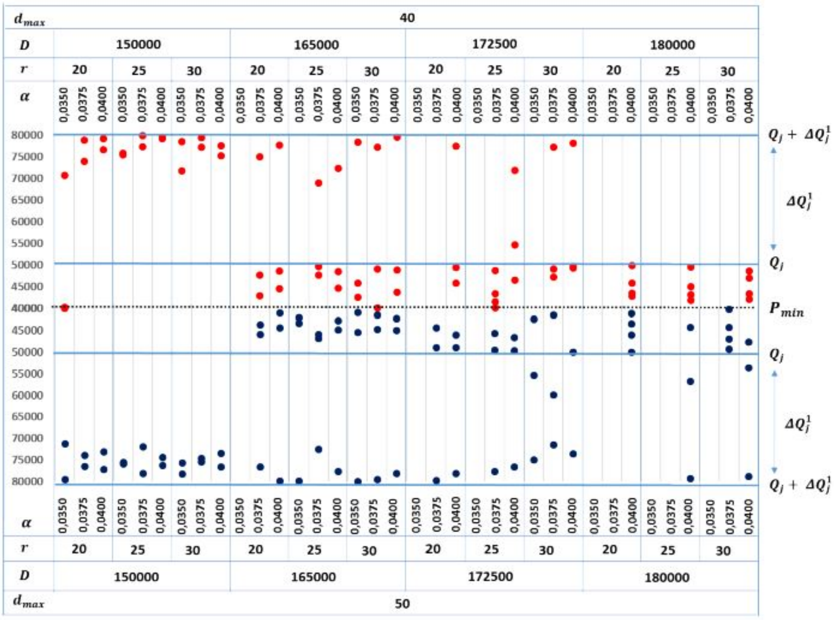


Figure 2.9. Production levels of BCs in the produced scenarios

made without taking into account how much a decision maker is confident to increase the current donation rate. Indeed, an “optimistic” decision maker, which is sure to increase the donation rate up to 0,0400, would be able to choose the lower cost re-organization solutions according to the threshold value D, by setting an appropriate covering radius r. On the contrary, a “prudential” decision maker is suggested to look for scenarios achievable with the current donation rate (0.0350). In this condition, higher values of D (165,000 or 172,500) can be met only by increasing the covering radius r and d_{max}. In this situation, our consolidated recommendation would be to adopt the scenario obtained by setting r = 30 km, d_max = 50 km, which is the more robust in terms of self-sufficiency goal, being able to reach at least 172,500 blood units.

2.6.3 Sensitivity analysis

In order to verify the robustness of the solutions provided by the model, we performed a sensitivity analysis on the cost coefficient of collection activities of the single facility j (c_{sj}); this way, on one hand, for each facility j included in the final solution ($y_{sj} + y_{cj}$ = 1), we determined the extra cost Δc^{+}_{sj} beyond which that facility is closed and, on the other hand, for each facility j not included in the solution ($y_{sj} + y_{cj} = 0$), we calculated the minimum reduction of cost Δc^{-}_{sj} under which that facility is included in the solution. For example, considering the solution of Scenario 1, obtained results show that the configuration of the BMS remains unmodified outside the area of the municipality of Naples, as the system is constrained to activate some centers, regardless of their costs, in order to reach farther donors and meet the self-sufficiency goal ($\Delta c^{+}_{sj} =$

$+\infty; \Delta c^-_{sj} = -\infty$). Figure 2.10 shows the obtained values Δc^+_{sj} and Δc^-_{sj} associated to the facilities located within the municipality of Naples, in terms of percentage variation of cost coefficient c_{sj} with reference to its intial value. In this case the provided solution, characterized by the presence of facilities 4 and 6, may change for a minimum variation of collection cost. In particular, facility 3 is the best candidate to enter in the solution as it is characterized by the lowest value of Δc^-_{sj}. This aspect is due to the fact that facilities in this area are very close and then slight variations in the collection costs can modify the configuration of the solution, with the entering facility replacing another one. In Figure 2.11 we compare the values provided by the sensitivity analysis in the case of scenarios 1 and 2, which differ for the values of d_{max} (40 and 50 km respectively). The increase of the value of covering radius makes competitive also facilities outside the municipality of Naples (see facilities 13, 15 and 17), that may enter or leave the solution depending on the associated collection cost.

Figure 2.10. Sensitivity analysis for Scenario 1 ($r = 20$ km, $D = 150{,}000$, $\alpha = 0.0350$, $d_{max} = 40$ km)

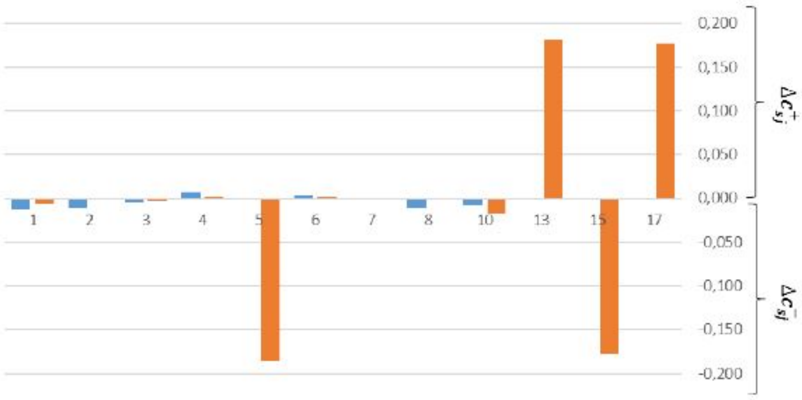


Figure 2.11. Comparison between sensitivity analysis in scenarios 1 and 2

The sensitivity analysis suggests that a deeper analysis of cost structure of a subset of existing facilities could be useful to obtain more refined solutions. Moreover, the low values of Δc_{sj} characterizing the facilities within the municipality of Naples,suggest that different indicators may be adopted along with geographic proximity to select within such area the facilities to be kept open.

2.6.4 Impact of reorganization decisions on blood distribution costs

In the model formulation, the third tier of the blood supply chain, represented by hospitals, has not been explicitly considered. This choice is motivated by the fact that the distribution of blood and its components or derived products towards hospitals is a tactical problem that can be better described as an inventory-routing problem and that is only slightly influenced by the facility location decisions taken at a strategic level, in terms of total costs. However, in order to give evidence of the impact of the location decisions taken by the model on the distribution cost towards hospitals, a specific analysis is conducted in this sub-section.
The public health-care system of the Campania region currently accounts for 50 hospitals, whose distribution is shown in Figure 2.12. The ones represented with red dots in the map are the hospitals where BCs are currently located. Geographical coordinates of the considered facilities are available in Appendix B - Table .3. For each given solution, we estimated the blood distribution costs from active BCs to hospitals, by considering the average distance between any hospital and its closest active BC. Hence, denoted by H the set of hospitals locations ($J \subset H$) and by l_jh the road distances between facilities $j \in J$ and $h \in H$, the proxy of distribution cost is defined as $DC = (\sum_{h \in H} \min_{j \in J: y_{cj}=1}\{l_{hj}\})/(|H|)$. In order to compare obtained cost DC with a benchmark value and, hence, evaluate the impact of the provided solution on the distribution activities, we calculated a lower bound $LB(DC)$, i.e. the minimum cost that has to be incurred by the system in order to deliver blood from BCs towards hospitals. To this end, we ran, for each combination of the parameters (α, r, d_max, D) a p-median model that decides the best location for p BCs in order to minimize the average distance between hospitals and their assigned BCs. In particular, we set p equal to the maximum number of BCs that can be located according to the minimum requirement of blood units to be processed ($p = \lfloor D/P_{min} \rfloor$), in such a way to reproduce a situation in which a less consolidation of processing activities is induced. Moreover, we considered as candidate locations only those nodes $j \in J$ able to attract, through donors directly assigned to themselves or through those assigned to BSs within the radius d_{max}, a minimum amount of blood units equal to P_{min} ($J' = \{j \in J : \sum_{k \in J: d_{jk} \leq d_{max}} (\sum_{i \in I: d_{ik} \leq r} a_i) \geq P_{mim}\}$).

By setting the same combination of calibration parameters of Scenario 1 (r = 20 km, D = 150,000, α = 0.0350, d_{max} = 40 km), the p-median model suggests to locate three BCs (3, 12 and 20), 2 of which different from the ones activated by our model, and provides an average allocation distance almost six km lower than the value associated to Scenario 1 ($LB(DC)$ = 32,6 km; DC= 38,6 km). A comparison between the cumulative distribution functions of the allocation distances between hospitals and their assigned BC is reported in Figure 2.13. As it can be noticed, the curves show significant differences only for a subset of hospitals, almost 30%.

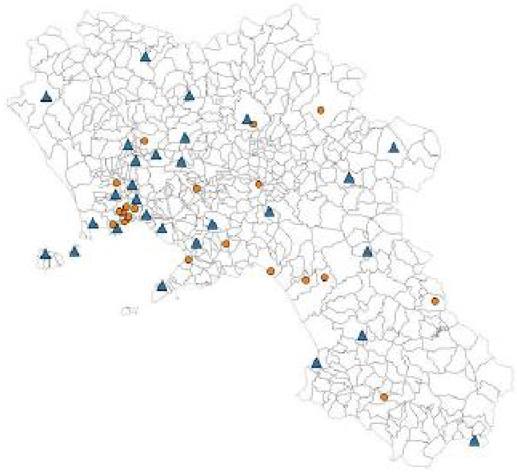


Figure 2.12. Hospitals locations in the Campania Region

By extending the analysis to all the set of produced scenarios, we obtained the results shown in Figure 2.14 (for further details, readers are encouraged to consult Appendix D - Table .5). The graphic highlights that the difference between the two measures DC and $LB(DC)$ ranges from a minimum of about five km up to a maximum of almost 22 km. Moreover, in 50% of the considered scenarios, this value is contained within 13 km. It is also worth noticing that the p-median model provides solutions to a relaxed problem, where the donors-to-facility allocation mechanism is absent. Therefore, the presented gaps could be even smaller given the potential infeasibility of these solutions. This analysis confirms that the impact on the distribution costs is limited. This result is partially due to the structure of the formulated model; indeed, the model will tend to locate BCs in areas characterized by a high density of population (in order to attract donations) and hospitals (due to the geographical proximity to BSs). As BCs represent the depots from which blood components and derived products will depart towards hospitals, the choice should positively affect the overall distribution costs. Moreover, distribution costs could be further optimized by accurately designing routes and schedules of blood units' provisions depending on the inventory levels planned at each hospital; and, also in this sense, the consolidation of processing activities could be beneficial as a lower number of larger BCs, that control the stock of blood products and components at regional level, should foster inventory optimization and reduce the risk of blood products shortage.

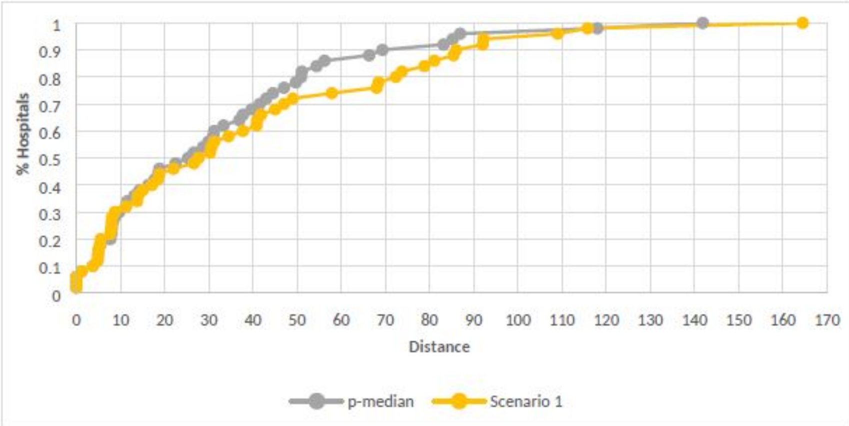


Figure 2.13. Cumulative distribution function of allocation distances by the fraction of hospitals

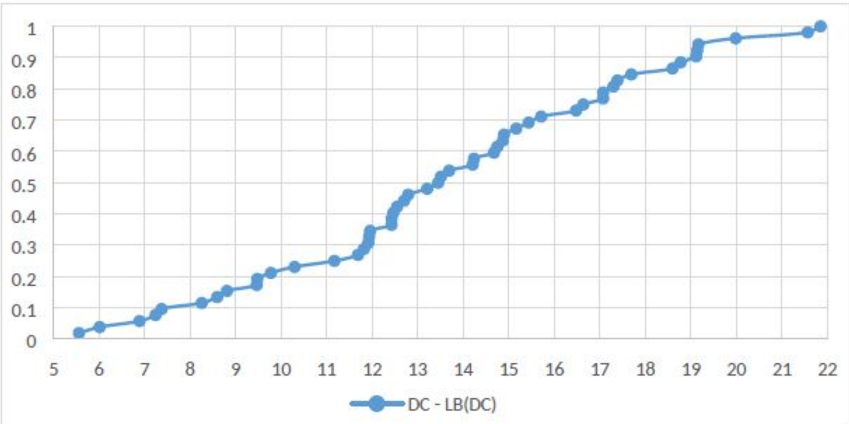


Figure 2.14. Difference between average allocation distances by the fraction of scenarios

2.7 Conclusions

In this chapter, we analyzed a strategic problem concerning the territorial reorganization of regional blood management systems, aimed at closing or reconverting a set of operating facilities already located in a region, in order to find a trade-off solution between the aim of attracting donors (self-sufficiency goal) and the aim of containing the management costs (efficiency goal). To this end we proposed and developed an optimization model, getting inspiration from the literature on location models. The main objective of the proposed model is to support the decision maker by providing a set of possible feasible scenarios (according to the variation of some calibration parameters), that can be evaluated and compared on the basis of a set of appropriate performance indicators. The model was tested on the real case study of the Campania Region. Obtained results show that the model is able to produce realistic scenarios, that can be evaluated by the Regional Authority to guide an informed process for the definition of possible plan to gradually modify the system. The proposal certainly represents an original approach to tackle the specific problem, but it could also be considered to solve similar problems in the health care context.
In the next chapter, we focus our attention to postal service and, in particular, on the reorganization of postal collection operations. Similarly to the case analyzed in this chapter, we will propose a possible approach to tackle the problem by facing the trade-off determined by two inherent but, at the same time, conflicting goals: users' accessibility, on one hand, and management costs reduction, on the other.

Chapter 3

An integrated Facility Location-Districting Model for the organization of postal collection operations

Summary

In this chapter, we deal with a real problem concerning the reorganization of the collection system of the Italian postal service provider, based on the reduction of the number of postboxes currently located in an urban area. This study is motivated by the consistent fall in postal volumes, generated by the substitution of traditional letter posts by electronic forms of communication, that has rendered the collection of postal items highly inefficient. To tackle the problem, we propose solving a mathematical programming model, embedded in a constructive heuristic procedure, in order to reduce the number of postboxes and to create associated collection areas, i.e., clusters of postboxes to be assigned to operators. Considering the crucial role of postboxes as main access points of users to the postal network, equity is also taken into account. We also consider workload balance and shift duration requirements in the determination of the collection areas. Several computational experiments are conducted based on real data from the city of Bologna, in northern Italy. The resulting scenarios show the capability of the model to support the decision making process towards the redesigning of the postal collection system.

3.1 Introduction

Service organizations are experiencing several profound changes driven by the wide diffusion of information and communication technologies (ICTs), which have significantly modified customer behaviour and increased competitiveness. In particular, ICTs had a strong impact on organizational design and performance, thus constituting the main driver towards an overall reshaping of the traditional concept of service (Barrett et al., 2015). In this context, public services are required to reach higher levels of efficiency and to adopt more flexible structures in order to better respond to unpredictable market changes (Phillips and Wright, 2009). We are interested in postal services, which have been affected by two significant trends: *e-substitution*, i.e., the progressive substitution of conventional letter mails by electronic forms of communication, and *e-commerce*, which has led to a significant growth of parcels volumes distribution against a crisis in the letter post segment (Hong and Wolak, 2008; Nikali, 2008). Indeed, in the 2002–2012 decade, the total traffic of letter post items has known an annual decrease of 2.0% at a global level and, starting from 2013, it has also gone down, with significant differences between countries, depending on their industrialization level and geographic position. In the same period, the annual growth of traffic of parcels has been 3.1%, with an increase of 6.4% in 2014–2015 (Universal Postal Union, 2015). This traffic currently accounts for more than 20% of the total income of postal providers, doubling its percentage share with respect to 2005. Such trends are expected to consolidate in the next few years and, consequently, the letter post segment is expected to lose its historical leading position as a major contributor to postal operators' revenues. The growth of e-commerce has been well documented in some recent publications (Difrancesco et al., 2017; Ding et al., 2017; Zhang et al., 2017).

This transformation has also been accompanied by an evolution of the regulatory framework, which has promoted a gradual liberalization and globalization of the market. At the European level, the member states have been required to regulate postal markets with the goal of ensuring an efficient, reliable and good-quality service at affordable prices to citizens and enterprises (European Parliament and Council of European Union, 1998, 2002, 2008).

In this environment, considering the general economic and political background, characterized by growing cuts to public expenditures, postal companies have been and are still involved in reengineering processes aimed at innovating their organizational, logistic and technological systems, and at improving their overall operational performance (Morganti et al., 2014; Cardenas et al., 2017).

Our work focuses on a pilot study provided by the main Italian Postal Service provider (Poste Italiane S.p.A), dealing with the reduction of postboxes located over a specific area and the definiton of new *collection areas*, i.e. groups of postboxes to be assigned to a single postman for the daily operations (routing and letters pick-up). The problem is motivated by the fact that in the described context, the average daily amount of mail accumulated in each postbox is actually very low and consequently, the overall organization of the collection activities are highly inefficient. However, any reorganization decision cannot be taken by neglecting the nature of the service being considered, defined by EU as a "universal service" (European Parliament and Council of European Union, 1998), meaning that it has to be accessible to all users, regardless of their geographic position. In the Italian case, the regulatory authority set two main standard quality requirements for the territorial organization of postboxes: i) a *spatial criterion*, indicating

maximum distances that users should cover in order to reach their closest postbox; ii) a *demographic criterion*, indicating the maximum number of inhabitants who have to be served by each postbox and then, the ideal number of postboxes to be kept open over a study region (AGCOM, 2014).
In order to tackle this problem, we introduce a mathematical programming model and we solve it on real data from the city of Bologna. In particular, to solve the problem, we also devise a constructive heuristic procedure. The proposed solution procedure is applied to a real case study with the aim of showing its ability to support the decision making process. The remainder of the chapter is organized as follows. In Section 3.2, after a short description of a typical postal logistic system, we review the state-of-the-art of models and algorithms specifically devoted to postal applications. The formal description of the problem, the mathematical model and the proposed solution procedure are given in Section 3.3. Section 3.4 focuses on the description and solution of the case study. Conclusions follow in Section 3.5.

3.2 Literature background

Postal services are provided through a dedicated logistics network capable of performing all typical operations such as reception, collection, transportation, sorting, and delivery of postal items over a given territory (Figure 3.1). *Reception* is the phase in which users access facilities of the postal network in order to send postal items, while *collection* is the phase in which operators visit these facilities to collect the items and transfer them to departure sorting centers (DSCs), where they are clustered on the basis of their final destination. *Transportation* to arrival sorting centers (ASCs) is performed through different single or combined modes (airplanes, trains, vehicles), according to the volumes, the distance, and the delivery deadlines. At ASCs postal items are sorted and prepared for the final door-to-door distribution phase, performed through daily tours assigned to postmen. All these activities must be completed within deadlines which depend on the specific services required by users.

The problems associated with the planning, organization and management of this complex system have stimulated a rich variety of models and algorithms. Regarding the *planning aspect*, a first proposal specifically focused on postal applications is due to Labbé and Laporte (1986) who solved the problem of locating postboxes with the aim of minimizing the total distance traveled by users and the routing cost associated with emptying the postboxes. More recently, Blagojević et al. (2013) developed a novel approach to locate the optimal number of postal units, based on the generation of fuzzy rules from numerical data, while Higgs and Langford (2013) investigated the spatial impact of post office closures on user accessibility in rural and urban areas.
Several studies focused on *personnel scheduling problems* arising in various phases of the system (Jarrah et al., 1994a; Bard et al., 2003; Bard, 2004; Bard and Wan, 2008; Zhang et al., 2009). The performance improvement of machines and equipment used to support the operations has also been the focus of several papers (Jarrah et al., 1992, 1994b; Zhang and Bard, 2005). Wang et al. (2005) considered the problem of sequenc-

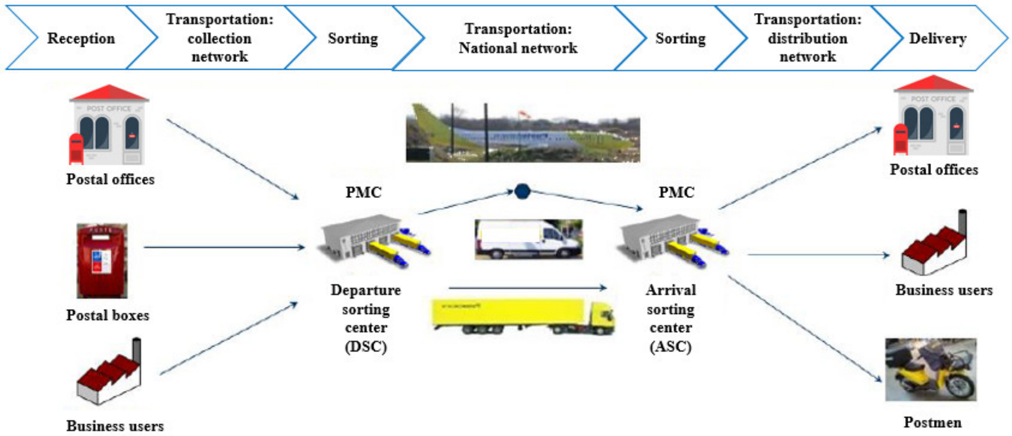


Figure 3.1. Representation of a postal logistics network.

ing the processing of incoming postal items at DSCs or at ASCs, with the objective of minimizing the unused capacity of outbound trucks.

Concerning the *transportation phase*, McWilliams and McBride (2012) studied the parcel hubs scheduling problem, namely the scheduling of inbound trailers for the minimization of the timespan between unloading and loading docks. More recent contributions related to the inbound and outbound truck scheduling were those of Jarrah et al. (2014), Boysen et al. (2017) and Zenker and Boysen (2018).

Grünert and Sebastian (2000) provided an overview of the tactical problems related to several problems encountered in the final *distribution phase*, while Irnich (2008) proposed mathematical formulations and solution methodologies for real-world postal problems arising in letter mail delivery. Algorithms for the integrated management of collection and distribution activities were presented in the studies by Mitrović-Minić et al. (2004), Jung et al. (2006) and Qu and Bard (2012). Abbatecola et al. (2016) described a decision support system for the management of postal deliveries in an urban area based on a vehicle routing model. Rosenfield et al. (1992), Novaes and Graciolli (1999) and Novaes et al. (2000) considered continuous approximation models (see Franceschetti et al., 2017 for an extensive review) for the determination of the optimal size of service territories and of the total transportation cost.

We focus on the reorganization of postal collection operations. To this end, we propose a mathematical model integrating two decision levels: location and districting. The location part of the model aims at identifying the set of postboxes to be kept active so as to ensure an equitable access to the users (Marsh and Schilling, 1994; Talen and Anselin, 1998). In practice, a covering paradigm is applied in our model (García and Marín, 2015): indeed, we ensure that a certain fraction of users is covered (i.e. has at least one postbox) within a given accessibility distance. The goal of the districting part is to group the located postboxes in a given number of areas to be visited by dedicated operators. In the existing literature, several studies resort to the use of districting models (Ricca et al., 2013; Kalcsics, 2015; Bruno et al., 2017b) to cope with the strategic problem of designing pickup and delivery areas in distribution logistics (Galvão et al., 2006; Haugland et al., 2007; Zhong et al., 2007; Jarrah and Bard, 2012; Lei et al., 2012, 2015, 2016). Notably, in all the extant studies, the districting phase involves a set of points that already located in a given area. In contrast, we introduce location decisions concerning the points (i.e. postboxes) to be clustered. This element represents a distinctive feature of our model.

Another valuable element of the study lies in the novelty of the application. In fact, to the best of our knowledge, the only work tackling a problem similar to the one analysed in this chapter is by Labbé and Laporte (1986). As in this reference, we consider the distances between customers and their allocated postbox, as well as the routing cost incurred to empty the postboxes. However, we particularize the problem by splitting the set of users according to their accessibility conditions and we group the located postboxes into a fixed number of clusters, so that they can be visited by dedicated operators within specified deadlines.

In order to take the latter constraint into account, we propose a constructive decomposition heuristic in which the model is iteratively solved until a feasible solution, in terms of work shift duration and workload balancing requirements, has been obtained. This way, we are able to show, by means of the conducted computational experiments, the capability of the introduced model to handle real-world instances and to provide meaningful managerial implications.

3.3 Problem description and mathematical model

The logistics network of Poste Italiane currently contains more than 50,000 postboxes spread throughout the national territory to provide users local access to the network and comply with the quality of service requirements (AGCOM, 2014). In recent years, the drastic drop in letter volumes has led to a general underutilization of such postboxes, thus making the current organization of the collection service unsustainable. Indeed, as long as a postbox is open, it has to be maintained and visited each day by an operator, thus impacting on the total operational cost. As a result, the postal provider is now evaluating the opportunity of reorganizing the collection system with the aim of reducing the number of existing postboxes and, hence, the number of operators devoted to collection activities. With this aim in mind, the following decisions should be made:

- the identification of the postboxes to be closed, at the strategic level;
- the definition of proper *collection areas*, i.e. groups of postboxes, to be visited by the same operator.

When tackling this problem, specific constraints must be taken into account.

From a strategic point of view, the reduction of postboxes may yield a deterioration of users' accessibility. The universal nature of the postal service imposes that such decisions be made by preserving the access of users to the logistics network, according to the quality standards fixed by the regulatory authority. In the Italian case, two main criteria have to be considered: i) a *spatial criterion*, indicating maximum distances that users should cover in order to reach their closest postbox; ii) a *demographic criterion*, indicating the maximum number of inhabitants who have to be served by each postbox and then, the minimum desired number of postboxes (AGCOM, 2014). As concerns the first criterion, proper equity conditions on the distances to be covered by the users in the final configuration have been included in the model (Marsh and Schilling, 1994; Talen and Anselin, 1998; Barbati and Piccolo, 2016; Barbati and Bruno, 2017). In particular, we imposed that extra travel costs may be incurred only by those users located within a preset distance from their closest postbox, while the accessibility condition of the most disadvantaged users cannot be worsened any further. As concerns the *demographic criterion*, we did not include any explicit condition in the model regarding the minimum number of postboxes to be kept open. This choice is due to the fact that the relaxation of this criterion is currently under discussion in the context of a negotiation process between Poste Italiane and the national regulatory authority, as it is no longer considered a proper quality standard for service provision. Indeed, demographic conditions were originally introduced as a sort of capacity constraint to avoid the risk of having few access points in the network compared to the global demand, especially for densely populated areas. However, due to the decreasing trend characterizing the letter post segment, the daily demand has become negligible compared with the collection capacity and the service can be substantially considered to be uncapacitated. This criterion has therefore become obsolete and its enforcement could yield a high proliferation of underutilized postboxes over the national territory, with a consequent damage for the postal providers, in terms of competitiveness.

From a tactical point of view, the definition of collection areas should be made by taking into account aspects related to the work shift duration of the operators. In particular, the workload assigned to each operator should be feasible in terms of duration of the tour performed to visit and empty the postboxes within its assigned collection area. Moreover, the workload balance of the operators should also be taken into account.

To model the problem, we consider a set I of nodes at which users are located, and a set J of existing postboxes partitioned into a subset of *standard* postboxes, i.e. postboxes that can be removed, and a subset J_0 of *special* postboxes that need to be kept active, being located close to post offices or to crucial locations such as hospitals and universities. We denote by d_{ij} the distance between nodes $i \in I$ and $j \in J$ and, assuming that each user patronizes the closest postbox, we adopt the minimum distance $d_i^{min} = \min_{j \in J}\{d_{ij}\}$ as a measure of the accessibility of users to the network. By fixing a given threshold distance $\overline{d}$, the set of users I may be partitioned into two subsets I' and $I \setminus I'$, which respectively include the users located further or closer than $\overline{d}$ from their nearest postbox. In order to take equity into account, we impose that the postboxes patronized by users in I' cannot be closed.

The proposed model aims at identifying the postboxes to be closed and at partitioning the remaining ones into a given number p of clusters, called *collection areas*, so as to optimize the workload of the operators and to preserve the accessibility of the worst served users. Our objective function is the minimization of a compactness measure

of the created collection areas, since it is reasonable to assume that a more compact cluster of nodes most likely requires a lower routing time. It is important to stress that although compactness is widely acknowledged as a proxy measure of the travel times within a district (Kalcsics, 2015; García-Ayala et al., 2016), it does not necessary lead to shorter routes. We provide in Appendix G - Figure .14 the correlation obtained between the two measures in our computational experiments.
We use the following notation:

I	set of nodes where users are located, indexed by i;
J	set of nodes where postboxes are currently located, indexed by j;
$J_0 \subseteq J$	subset of special postboxes;
c_{jk}	distance between postboxes $j, k \in J$;
d_{ij}	distance between nodes $i \in I$ and $j \in J$;
d_i^{min}	distance between node $i \in I$ and its closest postbox $j \in J$ ($d_i^{min} = \min_{j \in J}\{d_{ij}\}$);
$\overline{d}$	equity distance;
I'	subset of users in the worst accessibility condition, i.e. further than $\overline{d}$ from their closest postbox ($I' = \{i \in I : d_i^{min} > \overline{d}\}$);
J'_i	subset of postboxes located within a distance $\overline{d}$ from node $i \in I$ ($J'_i = \{j \in J : d_{ij} \leq \overline{d}\}$);
$J^\star$	subset of postboxes patronized by disadvantaged users $i \in I'$ ($J^\star = \{j \in J : d_{ij} = d_i^{min}, i \in I'\}$);
p	number of groups of postboxes or collection areas to be created;
M	a very large number.

The decision variables are as follows:

x_{ij}	binary variable equal to 1 if and only if user in $i \in I$ patronizes postbox $j \in J$;
y_j	binary variable equal to 1 if and only if postbox $j \in J$ remains open;
z_{jk}	binary variable equal to 1 if and only if postbox $j \in J$ is assigned to the postbox $k \in J$.

The model is then formulated as follows:

$$\text{minimize} \quad \sum_{j \in J} \sum_{k \in J} c_{jk} z_{jk} \tag{3.1}$$

$$\text{subject to} \quad x_{ij} \leq y_j \qquad i \in I, j \in J \tag{3.2}$$

$$\sum_{j \in J'_i} x_{ij} = 1 \qquad i \in I \setminus I' \tag{3.3}$$

$$\sum_{j \in J \setminus J'_i} x_{ij} = 0 \qquad i \in I \setminus I' \tag{3.4}$$

$$\sum_{t \in J'_i} d_{it} x_{it} + (M - d_{ij}) y_j \leq M \qquad i \in I \setminus I', j \in J'_i \tag{3.5}$$

$$y_j = 1 \qquad j \in J_0 \cup J^\star \tag{3.6}$$

$$z_{jk} \leq z_{kk} \qquad j, k \in J \tag{3.7}$$

$$\sum_{k \in J} z_{jk} = y_j \qquad j \in J \tag{3.8}$$

$$\sum_{k \in J} z_{kk} = p \tag{3.9}$$

$$(1-\beta)UBz_{kk} \leq \sum_{j \in J} c_{jk} z_{jk} \leq UB \qquad k \in J \tag{3.10}$$

$$x_{ij}, y_j, z_{jk} \in \{0,1\} \qquad i \in I, j, k \in J. \tag{3.11}$$

The objective function (3.1) is defined as the total radial distance of the active postboxes $j \in J$ from their assigned cluster center $k \in J$, thus representing a compactness measure of the created collection areas. Constraints (3.2) guarantee that each node $i \in I$ is assigned only to active postboxes j. Constraints (3.3) ensure that each node $i \in I \backslash I'$ is assigned to one and only one postbox located within $\overline{d}$, whereas Constraints (3.4) avoid the allocation to a postbox further than $\overline{d}$. Constraints (3.5) assign the users of $I \setminus I'$ to their closest active postbox within $\overline{d}$ (Berman et al., 2009). Constraints (3.6) ensure that special postboxes, as well as postboxes patronized by users $i \in I'$, are kept open. Constraints (3.7)–(3.10) regulate the aggregation of active postboxes in collection areas or clusters. In particular, Constraints (3.7) impose that a postbox $j \in J$ can only be assigned to postboxes k selected as cluster centers ($z_{kk} = 1$). Constraints (3.8) impose that only active postboxes j are assigned to a single postbox $k \in J$. Constraint (3.9) fixes the number of cluster centers equal to p. Constraints (3.10) impose a balancing requirement on clusters, by setting a lower and an upper bound on the internal compactness. In particular, the lower bound is set as a maximum deviation β from the upper bound UB. Finally, Constraints (3.11) define the domain of the decision variables.

3.3.1 Solution methodology

The proposed mathematical model does not incorporate routing decisions and, hence, it lacks an explicit formulation of the workshift duration and workload balance constraints. Instead, it makes use of a proxy measure of the routing costs within each collection area, coinciding with the internal compactness or radial distance t_k associated to the single clusters k ($t_k = \sum_{j \in J} c_{jk} z_{jk}, k \in J$). For this reason, we adopt a decomposition strategy according to which we initially solve the model, and we then evaluate the actual routing costs by solving a Travelling Salesman Problem (TSP) within each cluster, to check their feasibility, in terms of duration. The above strategy is embedded within a constructive heuristic, depicted in Figure 3.2, aimed at finding a scenario characterized by the minimum number of feasible clusters, i.e. having a Collection Time (CT) not exceeding a maximum workshift duration T.

In a generic step of the solution procedure, we solve the model (3.1)–(3.11) for a specific setting of parameters $\overline{d}$, p and UB. We also set $\beta = 1.00$, as it makes the left hand side of the balancing constraint (3.10) inactive. If the model leads to infeasibility, we run it with an increased number of collection areas ($p \leftarrow p + 1$), otherwise we solve a TSP within each collection area to compute the optimal length of the routes and then the collection times (CT_k) obtained by considering an average speed s and a unit time to collect items at each postbox v. If the maximum collection time ($\max_{k=1,\ldots,p}\{CT_k\}$) lies within the given duration T, included a given tolerance of 10%, we stop the procedure by setting $p = p^\star(\overline{d})$ and $UB = UB^\star(\overline{d})$. Otherwise, we run the model with a more stringent value of the upper bound UB. In particular, we set UB as 95% of the maximum compactness recorded in the obtained scenario ($UB \leftarrow 0.95 \times \max_{k=1,\ldots,p}\{t_k\}$).

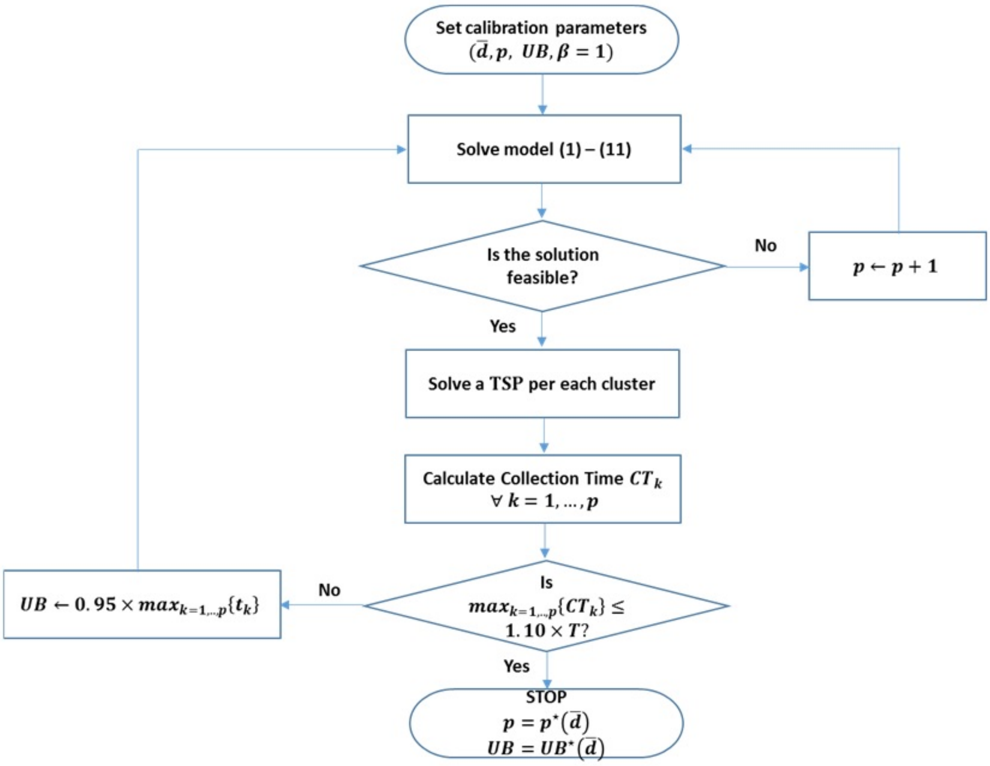


Figure 3.2. The proposed solution procedure.

In other words, according to our solution procedure, $p^\star(\overline{d})$ and $UB^\star(\overline{d})$ represent, respectively, the minimum number of collection areas to be activated and the maximum value of internal compactness to be set in order to have tours of feasible duration (not exceeding T).

Once a feasible solution is obtained, a specific procedure is triggered to achieve workload balance. In particular, we solve model (3.1)–(3.11) with $p = p^\star(\overline{d})$ and $UB = UB^\star(\overline{d})$ by decreasing the parameter β. In fact, imposing a more stringent condition on the minimum compactness per cluster is expected to lead to a more homogeneous distribution of the collection times.

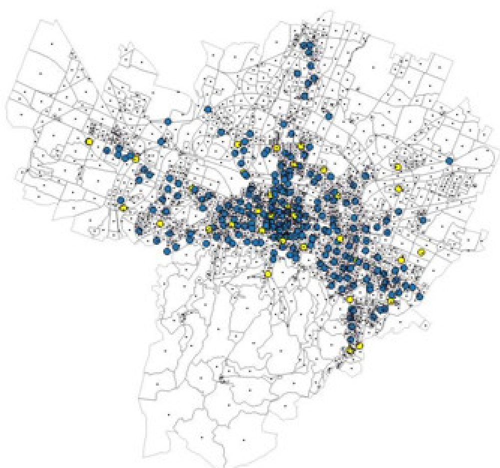
3.4 The Bologna case study

In this section, we report on the computational experiments performed to test the solution procedure proposed for the reorganization of the postal collection operations. We first describe the test data used. We then present and discuss the obtained results.

3.4.1 Test data

The solution procedure was applied to a real case study concerning the city of Bologna (371,217 inhabitants), located in the northern part of Italy. In order to apply the proposed procedure to this case, we discretized the location space into 2,333 territorial units, corresponding to the city census tracts (ISTAT, 2011), and we assumed that all users located in a census tract are concentrated in its centroid (Figure 3.3). We are conscious that any aggregation of demand points may introduce a bias error in the evaluation of users' accessibility, but "there is no agreement on to how measure error" (Francis and Lowe, 2015) and hence on the best way to discretize a study region. Nevertheless, in Appendix E we provide a detailed analysis aimed at justifying the adopted aggregation, in terms of its representativeness of the real demand distribution. Figure 3.3b shows the current location of the 272 postboxes J, highlighting, in yellow, the positions of the special postboxes ($|J_0| = 36$). The distances d_{ij} between the centroids and the existing postboxes, and distances c_{jk} between pairs of postboxes, were determined as the shortest paths on the road network.

(a) Census tracts



(b) Centroids and postboxes locations

Figure 3.3. Data related to the case study of Bologna (provided by Poste Italiane)

.

The solution provided by the model highly depends on the calibration of the parameter $\overline{d}$, which contributes to the determination of the set I' of users to be preserved from the reallocation mechanism. In order to set appropriate values for $\overline{d}$, we first calculated

the cumulative distribution of users' distances from their closest postboxes. In particular, considering the population a_i living in each census tract $i \in I$, we calculated the percentage of the total population A_{tot} having a distance from the closest postbox not exceeding a given $\overline{d}$ $(\alpha = F(d_i^{min} \leq \overline{d}) = \frac{1}{A_{tot}} \sum_{i \in I: d_i^{min} \leq \overline{d}} a_i)$; see Figure 3.4. This way, the equity distance $\overline{d}$ may be set according to the portion $1-\alpha$ of the population that the decision maker wants to preserve $(\overline{d} = F^{-1}(1-\alpha))$. In practice, in the case under analysis, for $\alpha = 1$, i.e., when no user is preserved and anyone is free to be reallocated, regardless of their current distance to the closest postbox $(I' = \emptyset)$, we obtain a threshold distance of $\overline{d} = 5,626$ m; by decreasing the value of parameter α the distance $\overline{d}$ decreases as well. For $\alpha = 0$, i.e., when all users are preserved $(I' = I)$, no reallocation is feasible. Table 3.1 reports the values of $\overline{d}$ obtained for several values of α, whereas Figure 3.5 shows how the partition of the demand nodes I in the subsets of *disadvantaged* users I' (in red) and *non-disadvantaged* users $I \setminus I'$ (in white) varies according to the value of α (and hence $\overline{d}$).
It is worth noting that the spatial criteria fixed by the Italian regulatory authority require that, in each municipality, at least the 75% of the total inhabitants be covered within a maximum distance of three km, 92.5% within five km and 97.5% within six km (AGCOM, 2014). Table 3.1 shows that for any value of α, the mechanism introduced in the model is definitely consistent and even more stringent with respect to the current requirements.
Although the above considerations are drawn with respect to the adopted aggregation, in Appendix F - Figure .13 we show that they still hold when considering the real distribution of demand points.

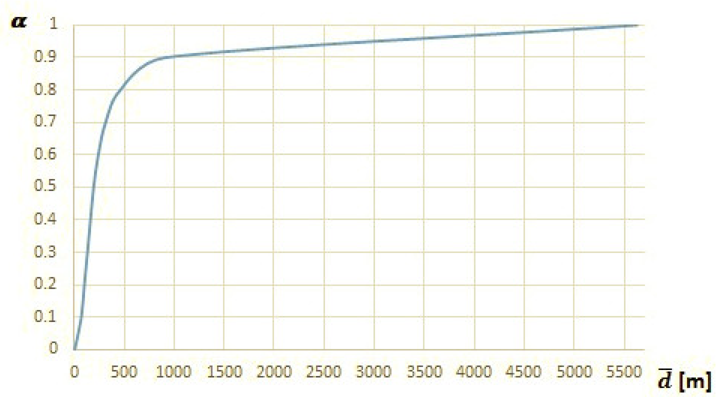


Figure 3.4. Distribution of the population by the distance from the closest facility.

The time to visit and empty all the postboxes within any area, indicated by Collection Time (CT) in the proposed solution procedure, was calculated by assuming a speed s of 10 km/h and a time v of 3.0 minutes to collect items at each postbox. Moreover, on the basis of an analysis conducted by Poste Italiane, the amount of time T dedicated to

α	1.00	0.90	0.80	0.70	0.60	0.00
$\overline{d}[m]$	5625.50	913.15	461.59	313.27	240.59	4.15

Table 3.1. Values of calibration parameter $\overline{d}$ selected for testing the model.

(a) $\alpha = 0.70$

(b) $\alpha = 0.80$

(c) $\alpha = 0.90$

Figure 3.5. Partitioning of the users into a set I' of disadvantaged users (red) and a set $I \setminus I'$ of non-disadvantaged users (white) by varying α

.

collection activities by each postman was fixed at 120 minutes.
In order to trigger our solution procedure, we need to set the values of the calibration parameters UB and p. The parameter UB imposes an upper bound on the compactness of the clusters to be created. Since clusters compactness is measured as the sum of radial distances of postboxes from their assigned cluster center, it coincides with half of the length of a star-shaped tour constructed within each group of postboxes. Therefore, feasible duration tours (i.e. at most equal to $1.10 \times T$) could be more likely obtained by setting $UB = 1.10 \times T \times s = 22,000$m.
The parameter p fixes the number of clusters. In order to make our solution procedure more efficient, we identify, for each value of α, a different initialization value $p_{min}(\alpha)$. In fact, according to the introduced notation, the minimum number n_{min} of postboxes to be located is equal to $|J_0| + |J^\star| + |\overline{J}|$, where $\overline{J}$ is the set of postboxes located in an optimal solution to a Set Covering Problem solved for users $i \in (I \setminus I')$, by setting a covering radius equal to $\overline{d}$. Hence, a lower bound L to the optimal tour through these postboxes is given by $L = \sum_{j\in(J_0\cup J^\star\cup\overline{J})} \min_{k\in J}\{c_{jk}\}$. Therefore, we can compute $p_{min}(\alpha) = (L/s + n_{min} \times v)/(1.10 \times T)$. The resulting values for $p_{min}(\alpha)$ are reported in Table 3.2.
Finally, in order to activate the balancing mechanism in Constraints (3.10) we also set parameter $\beta = 0.20, 0.40$.

α	1.00	0.90	0.80	0.70	0.60	0.00
$p_{min}(\alpha)$	2	3	5	6	8	9

Table 3.2. Values of calibration parameter $p_{min}(\alpha)$
.

The proposed solution procedure was coded in Python 3.6 and the TSP were solved using the open source VRP Spreadsheet Solver (Erdoğan, 2017). The embedded model was optimally solved using CPLEX 12.8 for each value of α, with $\alpha = 1.00, 0.90, 0.80, 0.70, 0.60, 0.00$. Details about computational times are provided in Appendix H - Table .6.

3.4.2 Experimental results

We now show how the proposed methodology can be used to support the decision maker in the reorganization process of the collection activities.
We first analyse in detail the scenarios obtained by setting $\beta = 1.00$, i.e. when the workload balancing requirement is not explicitly considered. Even in this case, the solutions can represent interesting scenarios for the decision maker. Indeed, even if not balanced, the workshifts of the postmen will be characterized, on average, by a lower saturation degree, which could offer more flexibility in managing their daily tasks.
Table 3.3 summarizes all the scenarios obtained by setting $\beta = 1.00$ and by varying α. This table reports, for each value of α, the optimal number of created clusters $p^\star(\alpha)$, the total number of active postboxes ($n_{tot} = \sum_{j\in J}\sum_{k\in J} z_{jk}$)) and the maximum collection time per cluster. As expected, a higher percentage of users to be preserved, in terms

of the accessibility condition, results in a higher number of facilities to be kept open; indeed, the number of facilities increases by decreasing α. Similarly, the number of created clusters $p^\star(\alpha)$ increases when α decreases, passing from three for $\alpha = 1.00$ to 13 for $\alpha = 0.60$. This is reasonable since with a higher number of facilities, more clusters have to be created in order to complete the tours within the time limit T. Interestingly, the solution with $\alpha = 0.00$ basically represents the 'status quo', because no user may be reallocated with such setting of calibration parameters and, hence, only four postboxes are closed.

α	1.00	0.90	0.80	0.70	0.60	0.00
$p^\star(\alpha)$	3	6	8	10	11	13
n_{tot}	39	69	125	183	224	268
Max CT	131.52	126.31	130.51	131.80	131.93	131.46

Table 3.3. Characteristics of the produced scenarios ($\beta = 1.00$)
.

Figure 3.6 shows the solution obtained by fixing $(\alpha, p^\star) = (0.90, 6)$. It can be seen that 203 postboxes out of 272 have been removed, while the remaining 69 are grouped into six clusters, depicted with different colours. In Table 3.4, each cluster k is characterized by its *number of postboxes* ($n_k = \sum_{j\in J} z_{jk}$), its internal *compactness*, i.e., the total radial distance of postboxes from their assigned cluster center ($t_k = \sum_{j\in J} c_{jk}z_{jk}$), and its *collection time* (CT_k), i.e., the time needed to visit and empty all its postboxes. It can be seen that the sizes of the six clusters are quite different, with a number of postboxes ranging from nine to 18, and a collection time ranging from 50 minutes to almost two hours.

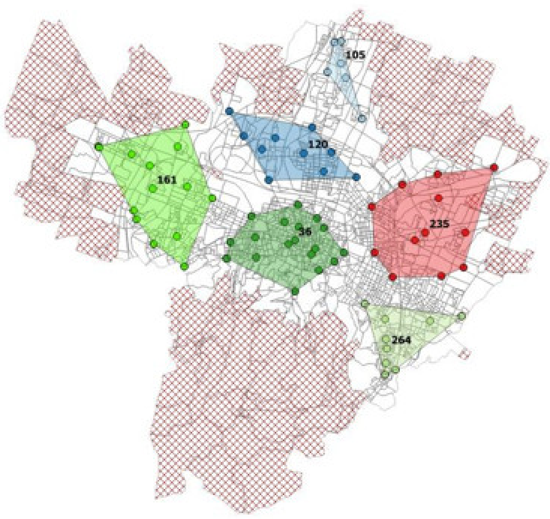


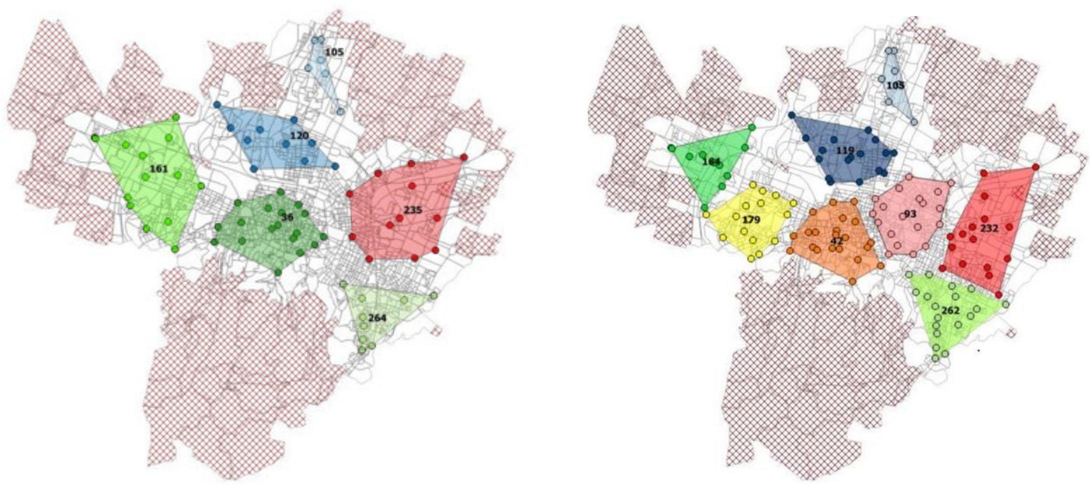
Figure 3.6. Scenario 1
($\alpha = 0.90$; $p^\star = 6$) .

It is interesting to analyse the effect of α on the solutions provided by the model.

Cluster center	n_k	t_k [m]	CT_k [min]
36	18	18,971.55	126.31
105	6	4,055.11	49.82
120	10	12,105.06	92.14
161	14	20,144.17	122.83
235	12	16,926.49	119.64
264	9	8,539.26	76.53
Total	69	80,741.64	587.27
Average	11.50	13,456.94	97.88

Table 3.4. Characteristics of scenario 1 ($\alpha = 0.90$; $p^\star = 6$)

Figure 3.7 illustrates the solution obtained for $(\alpha, p^\star)$=(0.80, 8). When the constraints are more stringent, the number of postboxes increases from 69 to 125, and the resulting number of collection areas passes from six to eight. As reported in Table 3.5, a higher number of postboxes has a positive impact on accessibility as the average distance to the closest postboxes is within 429 m, about 100 m less than the average accessibility condition obtained with $\alpha = 0.90$. In order to illustrate how user accessibility is affected by the solution provided by the model, we show in Figure 3.8 the new distributions of minimum distances d_i^{min} yielded by the two scenarios. Of course, the distribution associated to the first scenario shows the same values as the initial distribution for $\alpha \geq 0.90$, which is consistent with the fact that the distance from the closest facility does not change for the disadvantaged users; the same happens with $\alpha \geq 0.80$ for the second scenario. For the non-disadvantaged users, the average accessibility distance increases, as testified by the fact that the curves progressively move to the right by increasing α.



(a) Scenario 1 ($\alpha = 0.90, p^\star = 6$) **(b)** Scenario 2 ($\alpha = 0.80, p^\star = 8$)

Figure 3.7. Maps of scenarios 1 and 2

.

α	$\overline{d}$	Number of postboxes	Average user accessibility distance [m]	Average objective function per cluster [m]	Minimum collection time [min]	Maximum collection time [min]	Average collection time [min]
0.80	461.90	125	428.94	15,236.88	84.43	130.51	109.88
0.90	913.50	69	529.60	13,456.94	49.83	126.32	97.88

Table 3.5. Comparison between scenarios 1 and 2

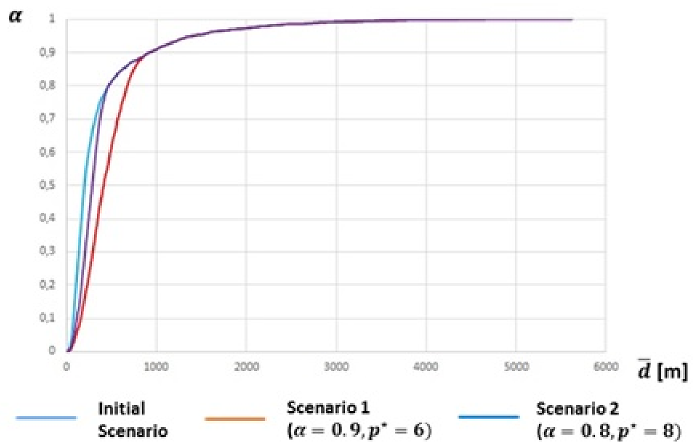
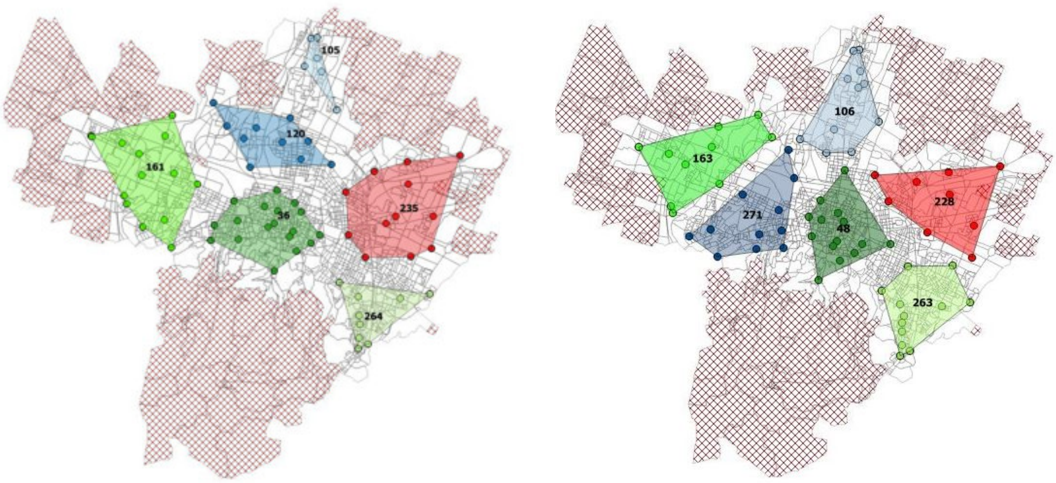


Figure 3.8. Comparison of the users accessibility measures for the three scenarios

From Table 3.5, it can be seen that the collection times in the single areas are not well balanced since there exists a significant gap between the minimum and maximum times under both scenarios. For this reason, according to the proposed solution procedure, we solve for each value of α the model (3.1)–(3.11) with $(p, UB) = (p^\star(\alpha), UB^\star(\alpha))$, by decreasing the value of the parameter β in order to show how the introduction of the constraint regulating the balance condition can help generate more balanced solutions. As an example, Figure 3.9b depicts the solution obtained with $(\alpha, p^\star, UB^\star) = (0.90, 6, 22,000)$, by setting $\beta = 0.40$. Tables 3.6 highlights significant differences between the solutions. Despite very similar number of postboxes (69 vs. 68), the objective function increases by 6.23% (from 80,741.64 to 85,813.27 m). However, Scenario 3 is certainly more balanced in terms of estimated collection times, as reflected by the standard deviation values. Finally, Table 3.7 shows the minimum and maximum collection times in the produced scenarios.

It can be observed that the gaps tend to be reduced when a more stringent bound is applied. We conclude that the objective function actually works well as a proxy measure for the cluster collection times. It therefore constitutes an additional decision making lever on which to act in order to meet the workload balancing requirements without dramatically compromising the computational complexity of the problem.



(a) Scenario 1 ($\alpha = 0.90, p^{\star} = 6, \beta = 1.00$)

(b) Scenario 3 ($\alpha = 0.90, p^{\star} = 6, \beta = 0.40$)

Figure 3.9. Maps of scenarios 1 and 3

.

Cluster center	n_k	t_k $[m]$	CT_k $[min]$
36	18	18,971.55	126.31
105	6	4,055.11	49.82
120	10	12,105.06	92.14
161	14	20,144.17	122.83
235	12	16,926.49	119.64
264	9	8,539.26	76.53
Total	69	80,741.64	587.27
Average	11.5	13,456.94	97.88
St. dev.	4.18	6,349.38	30.67

(a) Scenario 1 ($\alpha = 0.90, p^{\star} = 6, \beta = 0.40$)

Cluster center	n_k	t_k $[m]$	CT_k $[min]$
48	14	13,822.80	110.70
106	11	13,586.33	91.50
163	11	16,736.72	105.86
228	9	13,267.51	101.45
263	12	13,262.46	92.93
271	11	15,137.55	102.48
Total	68	85,813.27	604.93
Average	11.33	14,302.21	100.82
St. dev.	1.63	1,379.72	7.42

(b) Scenario 3 ($\alpha = 0.90, p^{\star} = 6, \beta = 0.40$)

Table 3.6. Characteristics of scenarios 1 and 3

	$\beta = 1.00$		$\beta = 0.40$		$\beta = 0.20$	
α	Min	Max	Min	Max	Min	Max
1.00	120.47	131.52	120.47	131.52	120.47	131.52
0.90	49.83	126.32	91.50	110. 70	93.30	100.47
0.80	84.43	130.51	85.26	127.98	101.14	119.40
0.70	60.13	131.80	88.36	131.06	99.29	130.63
0.60	105.71	131.94	106.95	129.79	108.30	128.41
0.00	94.46	131.46	105.86	131.93	108.29	131.93

Table 3.7. Minimum and maximum collection times [min] in the produced scenarios

3.5 Conclusions

In this chapter, we addressed a real problem related to the reorganization of the collection system of the Italian postal service provider, based on the reduction of the number of postboxes currently located in an urban area. However, since postboxes are the main access points of users to the postal service, any reorganization decision should aim to avoid an uncontrolled worsening of their accessibility. In such a multi-objective context, it becomes important to develop models capable of handling several criteria. To this end, we developed a mathematical programming model aimed at reducing the number of postboxes and at creating clusters of postboxes to be assigned to operators. We also devised a constructive heuristic procedure, based on a decomposition approach, in order to take into account specific requirements related to workshift duration and workload balance. The model, and hence the solution procedure, were tested on the Bologna data, and several accessibility analyses and graphical visualizations were performed by using a GIS software This enabled us to provide a large set of scenarios that can support the decision maker towards the reorganization process of postal collection operations.
So far, we proposed mathematical models characterized by a deterministic setting; however, strategic problems like those analyzed in this thesis may involve decisions that must hold for some considerable time. During this time, changes may occur in the underlying conditions. Hence, it becomes crucial for decision makers to hedge against uncertainty when perfect information about some parameters is not available. Therefore, in the next chapter, we propose a stochastic programming modeling framework for a specific family of optimization problems, strictly related to facility location, namely districting problems.

Chapter 4

A stochastic programming modeling framework for Districting Problems

Summary

In this chapter, we focus our attention on districting problems. This class of problems, which is strictly related to facility location, is widely used to cope with the strategic design of services. In particular, in this chapter, a stochastic districting problem is investigated. Demand is assumed to be represented by a random vector with a given joint distribution function. A two-stage mixed-integer stochastic programming modeling framework is proposed. The first stage comprises the decision about the initial territory design: the districts are defined and all the territory units assigned to one and exactly one of them. In the second stage, i.e., after demand becomes known, balancing requirements are to be met. This is ensured by means of two recourse actions: outsourcing and reassignment of territory units. The objective function accounts for the total expected cost that includes the cost for the first stage territory design plus the expected cost incurred at the second stage by outsourcing and reassignment. The (re)assignment costs are associated with the distances between territory units which means that the focus is on the compactness of the solution. The new modeling framework proposed is tested using 96 instances built using real geographical data from a province in northern Italy. Two of these instances are also used for illustrative purposes. The results show the relevance of capturing uncertainty in a districting problem. Several extensions to the investigated problem triggered by several features of practical relevance are also discussed.

4.1 Introduction and literature review

Districting Problems (DPs) aim at partitioning a set of basic geographic areas, named Territorial Units (TUs), into a set of larger clusters, called districts, according to some planning criteria. As introduced in Section 1.3, interested readers are encouraged to refer to Kalcsics et al. (2005) and, for a more up-to-date overview, to Kalcsics (2015) for extensive reviews on DPs. Taking these works into account, in this section, our effort is devoted to reviewing only the more recent contributions found in literature.

Bianchi et al. (2016) and Kim et al. (2017) explored an interesting application of DPs, which consists of the territory design of functional regions. In this kind of problems the hypothesis of complete and exclusive assignment of TUs to districts were relaxed and only clusters with a strong spatial interaction were created. Accordingly, the authors seek to optimize a distance-based compactness measure. The latter was frequently used as an objective function in DPs. In fact, it represents a good proxy of users' accessibility to public facilities (Bruno et al., 2017b) and it typically leads to shorter travel times when designing distribution areas for service provision (Bender et al., 2016; García-Ayala et al., 2016).

The minimization of maximum territory dispersion, namely the maximum distance between any TU and the center of the districts they are assigned to, was considered in Ríos-Mercado (2016) and Ríos-Mercado and Escalante (2016). Due to the multiplicity of planning goals to be simultaneously achieved, some works adopt a multicriteria setting (Camacho-Collados and Liberatore, 2015; Camacho-Collados et al., 2015; Xie and Ouyang, 2016).

Clearly, the objective to be optimized can be also specifically tailored according to the application. For instance, De Fréminville et al. (2015) introduced the so-called Financial Product Districting Problem, where the goal was to partition a set of geographical units in such a way that a cost homogeneity is achieved within the designed territories. Such homogeneity was expressed in terms of the cost variance. Bruno et al. (2016b) defined a model to support the rationalization process of public facilities aimed at optimizing the total cost needed to the activation of additional capacities to satisfy the reallocated demand generated by the closure of some facilities. Lin et al. (2017) proposed a mixed-integer programming formulation for a problem emerging in the context of home health-care services: the Meals-On-Wheels service districting problem. The goal was to design the minimum number of districts covering all the basic units. Akdoğan et al. (2018) considered the minimization of the mean response time in a problem involving the location of emergency services.

One aspect of practical relevance in districting problems concerns the need to cope with changing demand. This may stem for instance from the expansion of urban areas or migration movements. Depending on the particular problem we are dealing with, different possibilities emerge. One is to to assume a reactive posture and solve a so-called redistricting problem. This is an optimization problem that aims at redesigning existing districts in some geographical area. De Assis et al. (2014) tackled such a problem in the context of meter reading in power distribution networks. A bi-objective problem was considered. The objectives were related to compactness and homogeneity of districts. The authors developed a heuristic algorithm to approximate the Pareto front. Other works dealing with redistricting problems were those by Caro et al. (2004) and Bruno et al. (2017b). In particular, Caro et al. (2004) proposed a mathematical model and a GIS-based approach to solve a school redistricting problem, whereas Bruno et al. (2017b) presented several formulations to address a redistricting problem emerging in the redesign of administrative boundaries of the Italian provinces.

One alternative to cope with demand changes is to become proactive and make a decision that directly accounts for such changes. When accurate forecast for the demand are available, we can embed time in the optimization models. To the best of our knowledge, the only papers addressing multi-period territory design were those by Lei et al. (2015), Bender et al. (2016), Lei et al. (2016), and Bender et al. (2018). A multicriteria optimization framework was considered in the first and third papers.

Finally, if demand changes are uncertain then embedding uncertainty in the models may be desirable. Assuming that uncertainty can be measured using a probability function a stochastic optimization model emerges as appropriate.

So far, the research on stochastic districting has been mainly conducted in the context of vehicle routing problems. In those works, the authors treated the problem as a two-stage stochastic program with recourse where districts are designed in the first stage and routing decisions are planned in the second stage, once demands (Haugland et al., 2007) or customers (Lei et al., 2012) are revealed. Those were problems in which the districting decisions were driven by the need to build "good" routes for visiting the customers. In the above studies, tailored heuristic and metaheuristic solution methods were proposed. Stochastic vehicle problems based on efficient partitioning procedures of the service region were also exploited by Carlsson (2012) and Carlsson and Delage (2013).

In this chapter we introduce a Stochastic Districting Problem with Recourse (SDPR) whose aim is to partition a given set of TUs into a prefixed number of clusters in order to maximize the overall compactness and to meet balancing constraints, expressed in terms of average demand per district. Demands associated to each TU are modeled as random variables. The problem is treated as two-stage stochastic program with recourse. Districts are created in the first stage by allocating the basic areas to those TUs chosen as representative (centers) of the districts. Then, in the second stage, two recourse actions are considered: The first simply aims at overcoming demand shortage or surplus via outsourcing; the second consists of solving a redistricting problem.

The use of the new modeling framework we propose can be useful to solve practical problems emerging in the context of service districting, where the need to provide users with high service levels and fair accessibility conditions is a top priority for decision makers (schools, hospitals, postal and emergency services) as well as in the redesign of political and administrative boundaries or in the planning of sales territory where demands for goods and services can be highly unpredictable. Indeed, changing conditions in the labor market, the phenomenon of migratory flows and the strong impact of technology development on customers' habits and behaviors, for instance, may push towards profound reorganization processes to meeting new socio-economic and cultural homogeneity requirements and future demand trends (ESPAS, 2015). In all these cases, a strategic planning able to work well against all the possible future scenarios that may occur is necessary.

The potential contribution to the literature is fourfold: (i) to introduce a new modeling framework for a two-stage stochastic districting problem; (ii) to embed redistricting decisions as a way to hedge against uncertainty; (iii) to show the relevance of capturing stochasticity in districting problems; (iv) to show that the new models proposed in this chapter make sense i.e., lead to plausible solutions.

The optimization models used in districting contain key components of facility location-allocation models and thus the same occurs if uncertainty is captured. We refer the reader to the book chapter by Correia and Saldanha-da-Gama (2015) and to the references therein for an overview on discrete stochastic facility location problems. Throughout the chapter we will emphasize those aspects capturing in the context of districting that that are usual components of facility location models.

The remainder of the chapter is organized as follows: in Section 4.2 we propose the modeling frameworks for our problem. Then, in section 4.3 the formulation of the SDPR is presented. In section 4.4 the most common measures known in literature to assess the relevance of considering a stochastic framework are reviewed. Section 4.5 presents the computational tests performed using a commercial solver for solving the model proposed in Section 4.3. Possible extensions of the SDPR are discussed in Section 4.6. Then, the chapter ends with a summary of the work done, some conclusions drawn from it, and some directions for further research.

4.2 Formulations for districting problems

For the sake of clarity, we recall the reader the basic formulation for DPs (Hess et al., 1965) as it is the core of the developments proposed within this chapter.
We consider a set I of territorial units (TUs) that we want to divide into a fixed number, say p, of districts. Each district has a TU representing it. Hence, when some other TU is assigned to the district we abuse the language by saying that we are assigning a TU to the representative of the district. We note that single-assignment is assumed for the TUs as customary in districting problems.
We consider the following parameters defining our problem:

- d_i, demand of TU i ($i \in I$);
- c_{ij}, cost for assigning TU i to TU j ($i, j \in I$);
- α, maximum desirable deviation of the demand in a district w.r.t. the reference value μ.

and the following single assignment decision variables:

$$x_{ij} = \begin{cases} 1, & \text{if TU } i \text{ is assigned to TU } j; \\ 0, & \text{otherwise.} \end{cases} \qquad (i, j \in I)$$

In this case, $x_{jj} = 1$ indicates that TU j is assigned to itself which also indicates that it is selected as the "representative" TU of its district.

Using these decision variables, Hess et al. (1965) proposed the following mathematical model for the territorial districting problem with balancing constraints:

$$\text{minimize} \quad \sum_{i \in I} \sum_{j \in I} c_{ij} x_{ij} \tag{4.1}$$

$$\text{subject to} \quad \sum_{j \in I} x_{ij} = 1 \qquad i \in I \tag{4.2}$$

$$\sum_{j \in I} x_{jj} = p \tag{4.3}$$

$$(1-\alpha)\mu x_{jj} \leq \sum_{i \in I} d_i x_{ij} \leq (1+\alpha)\mu x_{jj} \qquad j \in I \tag{4.4}$$

$$x_{ij} \leq x_{jj} \qquad i, j, \in I \tag{4.5}$$

$$x_{ij} \in \{0, 1\} \qquad i, j \in I \tag{4.6}$$

The objective function (4.1) quantifies the total cost to be minimized. Constraints (4.2) ensure that each TU is assigned to exactly one district whereas constraints (4.3) guarantee that exactly p districts will be designed. The constraints (4.4) are the balance constraints. Inequalities (4.5) state that we can only assign TUs to representatives of a district. In the above model, these constraints are in fact implied by (4.4). However, we decided to keep them since in our stochastic programming models to be presented later they have a role they will be relevant for ensuring the feasibility of the solutions. Finally (4.6) define the domain of the decision variables.

Next, we introduce a stochastic variant of the districting problem above described.

4.2.1 A stochastic districting problem

A natural source of uncertainty in districting problems concerns the demands d_i, $i \in I$. In this case we may still think of organizing a territory into districts. However, if this is a here-and-now decision (i.e. made before uncertainty is revealed) it is desirable that it hedges against uncertainty.

Looking for a solution that accounts (i.e. is feasible) for all the possible future observations (scenarios) of the demand vector $\boldsymbol{\xi} = [d_1, \ldots, d_{|I|}]$ may be impossible or if that is not the case may render a too "fat" solution since we may end up planning for realizations that are too extreme although occurring with a very low probability. An alternative is to devise a plan that takes uncertainty into account without being too strict in terms of imposing its feasibility for every "future" we may have to deal with but considering the implementation of some recourse action in case the initial solution is rendered infeasible for the demands actually observed.

The above setting can be casted within the context of two-stage stochastic programming when $\boldsymbol{\xi} = [d_1, \ldots, d_{|I|}]$ is a random vector with some known joint cumulative distribution function (e.g. estimated using historical data). This is what we assume hereafter. As we show next, the deterministic model presented in the previous section can be reformulated in order to capture the stochasticity being assumed or the demand.

Under the uncertainty setting we are assuming, the balancing constraints (4.4) are not well-defined before the actual demands become known. Therefore, we relax such constraints when looking for a here-and-now solution and assume that extra costs are incurred if, upon observing the actual demand, we realize that in some district it is above [below] the maximum [minimum] threshold. These costs may correspond for instance to some outsourced activity, to extra costs incurred for assuring that we have conditions for appropriately supplying the demand, or to additional investments caused by the unmet minimum demand needed to justify the establishment of the district itself.

Let us denote by g_j (> 0) the extra cost in district j for every unit of demand above the maximum threshold and by h_j (> 0) the extra cost for every unit of demand below the minimum threshold (w.r.t a here-and-now solution). Additionally, let us consider two sets of auxiliary variables accounting for the "shortage" and "surplus" demand in each district w.r.t the thresholds initially set:

φ_j = demand shortage w.r.t. the minimum threshold, $j \in I$;

ψ_j = demand surplus w.r.t. the maximum threshold, $j \in I$;

The new problem we are dealing with will be called the Stochastic Districting Problem with Auxiliary Recourse (SDPAR); it can be formulated mathematically as follows:

$$\text{minimize} \quad \sum_{i\in I}\sum_{j\in I} c_{ij}x_{ij} + \mathcal{Q}(\mathbf{x}) \tag{4.7}$$

$$\text{subject to} \quad (4.2), (4.3), (4.5), (4.6)$$

with $\mathcal{Q}(\mathbf{x}) = E_{\boldsymbol{\xi}}[Q(\mathbf{x}, \boldsymbol{\xi})]$ and

$$Q(\mathbf{x}, \boldsymbol{\xi}) = \min \quad \sum_{j\in I} g_j\psi_j(\boldsymbol{\xi}) + \sum_{j\in I} h_j\varphi_j(\boldsymbol{\xi}) \tag{4.8}$$

$$\text{s. t.} \quad (1-\alpha)\mu x_{jj} \le \sum_{i\in I} d_i x_{ij} + \varphi_j(\boldsymbol{\xi}) - \psi_j(\boldsymbol{\xi}) \le (1+\alpha)\mu x_{jj} \qquad j \in I \tag{4.9}$$

$$\varphi_j(\boldsymbol{\xi}) \ge 0 \qquad j \in I \tag{4.10}$$

$$\psi_j(\boldsymbol{\xi}) \ge 0 \qquad j \in I \tag{4.11}$$

In the above model, the first stage problem seeks a here-and-now territory design (possibly violating the balancing constraints for some—or all—observations of the demand vector). The second stage model (4.8)–(4.11) accounts for the extra cost incurred due to the actual demand observed and given a first-stage decision. In fact, in the second stage model the values of x_{ij} are constants. Although a feasible solution to the second stage problem may have $\varphi_j > 0$ and $\psi_j > 0$ for some $j \in I$ due to the fact that $g_j, h_j > 0$ it is easy to see that every optimal solution to that problem will have at most one of those variables above zero. Furthermore, this will occur only if we need to set one such variable above zero in order to ensure the feasibility of the corresponding balancing constraint (4.9).

We note that $Q(\mathbf{x}, \boldsymbol{\xi})$ is a random variable since the quantities d_i and μ are in fact random variables. By considering its expected value for defining the recourse function $\mathcal{Q}(\mathbf{x})$, we are assuming a so-called neutral attitude of the decision maker towards risk which in our opinion defines a reasonable starting point for the study of stochastic districting problems. In fact, other attitudes towards risk often lead to measures that generalize the one we are considering and thus their analysis should be performed as follow-up to what we are presenting in the current chapter.

The above model has complete recourse. In fact, for every here-and-now decision there is a feasible completion in the second stage. This will be fully clear below when we show that the optimal values of the second stage decisions are fully determined by the first stage decision and by the actual demand observed. Finally, we note that like the original model proposed by Hess et al. (1965) this model solves in fact a stochastic location allocation problem with balancing requirements. Nevertheless, as the computational results presented in Section 4.5 show, it renders plausible solutions to the problem.

If we assume that the support of the random vector $\boldsymbol{\xi}$ is finite, say Ξ, then we can go farther in terms of formulating our problem. In fact, in that case, we can index the different scenarios in a finite set, say $S = \{1, ..., |\Xi|\}$. Moreover, we can index in S the stochastic demands, the assignment costs as well as the second stage decision variables, as follows:

- d_{is}, demand of TU $i \in I$ under scenario $s \in S$.
- φ_{js} demand shortage in district $j \in I$ w.r.t. the minimum threshold under scenario $s \in S$;
- ψ_{js} demand surplus in district $j \in I$ w.r.t. the maximum threshold, $j \in I$ under scenario $s \in S$;

We assume that the probabilities associated with the different scenarios are known (for instance estimated using historical data). In particular, we define π_s the probability that scenario s occurs, $s \in S$. Naturally, $\pi_s \geq 0$, $s \in S$ and $\sum_{s \in S} \pi_s = 1$. Additionally, we denote by $\bar{\mu}$ the reference value to be used in the balancing constraints.

We can finally write the so-called extensive form of the deterministic equivalent that we call model (M_1):

$$\text{minimize} \quad \sum_{i \in I} \sum_{j \in I} c_{ij} x_{ij} + \sum_{s \in S} \pi_s \left(\sum_{j \in I} g_j \psi_{js} + \sum_{j \in I} h_j \varphi_{js} \right) \tag{4.12}$$

subject to (4.2), (4.3), (4.5), (4.6)

$$(1 - \alpha)\bar{\mu} x_{jj} \leq \sum_{i \in I} d_{is} x_{ij} - \psi_{js} + \varphi_{js} \leq (1 + \alpha)\bar{\mu} x_{jj} \quad j \in I, s \in S \tag{4.13}$$

$$\varphi_{js} \geq 0 \qquad j \in I, s \in S \tag{4.14}$$

$$\psi_{js} \geq 0 \qquad j \in I, s \in S \tag{4.15}$$

The objective function (4.12) can be written in a different way:

$$\sum_{i\in I}\sum_{j\in I} c_{ij}x_{ij} + \sum_{j\in I} g_j \left(\sum_{s\in S} \pi_s \psi_{js}\right) + \sum_{j\in I} h_j \left(\sum_{s\in S} \pi_s \varphi_{js}\right) \tag{4.16}$$

Remark 3. *A close look shows that the variables φ_{js} and ψ_{js} ($j \in I$, $s \in S$) are, in fact, auxiliary variables (which explain the name given to the problem). In fact, the knowledge about the values of the x-variables as well as of the occurring scenario immediately determines their values:*

$\varphi_{js} = \max\{0, (1-\alpha)\bar{\mu}x_{jj} - \sum_{i\in I} d_i x_{ij}\},\ j \in I,\ s \in S;$

$\psi_{js} = \max\{0, \sum_{i\in I} d_i x_{ij} - (1+\alpha)\bar{\mu}x_{jj}\},\ j \in I,\ s \in S.$

Hence, we could easily reformulate our problem as a single-stage problem. However, we omit this reformulation since it is not helpful when it comes to solving the problem.

4.3 A stochastic districting-redistricting problem

In the stochastic models presented in the previous section, a here-and-now decision is made and there is no effective recourse decision. In fact, the second stage decision variables that were considered simply help us to account for the costs of having demand surplus or shortage w.r.t. the minimum and maximum thresholds set for the districts.

In practice, it may be relevant to be proactive. This means that upon observing a violation in the balancing constraints in some scenarios, instead of simply taking recourse actions that overcome infeasibilities of the here-and-now solution (and thus accounting for the corresponding costs), we may try to adapt slightly the territory design (first stage solution) to the occurring demands. In other words, we may think of performing a redistricting (Caro et al., 2004; De Assis et al., 2014; Bruno et al., 2017b). This is what we propose next. We continue assuming that we wish to build a fixed number p of districts. This is a major decision that keeps being done here-and-now. However, as before, we may observe a scenario for which the demand in some district is above or below the thresholds initially settled. In this case, in addition to considering straightforward recourse actions as before we also consider a "redistricting" recourse decision for some (hopefully just a few) territories. This means that such territories would be satisfied by some other district as a reaction to such a scenario. This type of recourse action makes sense since it corresponds to a "re-distribution" of the demand in order to get an overall balanced solution. The only TUs that cannot be reassigned are those that were set as district representatives by the first stage decision.

In what follows, we keep considering that the support of the random vector underlying the problem is finite and also that the demands are indexed in the scenarios. Furthermore, we consider that reassigning a territory incurs an extra cost. In particular, we define by r_{ijs} the cost we pay for reassigning the demand of TU i to the district represented by TU j, under scenario $s \in S$. Like for the previous models, the reassignment costs r_{ijs} should be typically related to the distances; demands can be considered or not as weighting factors. Nevertheless, in this case we wish to somehow "penalize" reassignments.

In order to formulate the new extension of the problem we consider one additional set of decision variables:

$$w_{ijs} = \begin{cases} 1, & \text{if TU } i \text{ is assigned to district } j \text{ under scenario } s; \\ 0, & \text{otherwise.} \end{cases} \qquad (i, j \in I,\ s \in S)$$

In addition to these variables, we consider another set of auxiliary variables that help us to preset a linear model: for every $i, j \in I,\ s \in S$, v_{ijs} is a binary variable equal to 1 if the assignment of TU i to TU j under scenario s corresponds in fact to a reassignment. More formally, we can define these variables as:

$$v_{ijs} = \begin{cases} 1, & \text{if } w_{ijs} = 1 \text{ and } x_{ij} = 0; \\ 0, & \text{otherwise.} \end{cases} \qquad (i, j \in I,\ s \in S)$$

The new problem will be called the Stochastic Districting Problem with Recourse (SDPR). Given all the aspects already discussed as well as the notation above introduced we can directly present the extensive form of the deterministic equivalent that we call model (M_2).

$$\text{minimize} \quad \sum_{i \in I} \sum_{j \in I} c_{ij} x_{ij} + \sum_{s \in S} \pi_s \left(\sum_{i \in I} \sum_{j \in I} r_{ijs} v_{ijs} + \sum_{j \in I} g_j \psi_{js} + \sum_{j \in I} h_j \varphi_{js} \right) \tag{4.17}$$

subjet to (4.2), (4.3), (4.5), (4.6), (4.14), (4.15)

$$\sum_{j \in I} w_{ijs} = 1 \qquad i \in I,\ s \in S \tag{4.18}$$

$$(1-\alpha)\bar{\mu} x_{jj} \le \sum_{i \in I} d_{is} w_{ijs} - \psi_{js} + \varphi_{js} \le (1+\alpha)\bar{\mu} x_{jj} \quad j \in I,\ s \in S \tag{4.19}$$

$$\sum_{\ell \neq j} w_{j\ell s} \le 1 - x_{jj} \qquad j \in J,\ s \in S \tag{4.20}$$

$$v_{ijs} \ge w_{ijs} - x_{ij} \qquad i, j \in I,\ s \in S \tag{4.21}$$

$$v_{ijs} \ge 0 \qquad i, j \in I,\ s \in S \tag{4.22}$$

$$w_{ijs} \in \{0, 1\} \qquad i, j \in I,\ s \in S \tag{4.23}$$

In this formulation, the objective function (4.17) accounts for the initial territory design plus the expected cost for redesigning the territory, and the expected costs for demand shortage and surplus in each district (w.r.t. the minimum and maximum thresholds, respectively). Constraints (4.18) and (4.19) seek a territory redesign (dependent on the observed scenario). Constraints (4.20) guarantee that a district that was selected by the first stage decision as a district representative is not reassigned in the second stage. Constraints (4.21) and (4.22) are an alternative way of writing that $v_{ijs} \geq \max\{0, w_{ijs} - x_{ij}\}$, $i, j \in I$, $s \in S$. On the other hand, due to the non-negativity of costs r_{ijs} we know that in every optimal solution we will observe $v_{ijs} = \max\{0, w_{ijs} - x_{ij}\}$, $i, j \in I$, $s \in S$. Therefore, we are paying the reassignment cost for the demand that is actually reassigned. Finally constraints (4.23) define the domain of the w-variables.

In principle, the above model allows all the territories to be reassigned for some scenario. However, since there are extra costs for reassigning and for demand shortage and surplus, an optimal solution to the problem will seek to reassigning as little demand as necessary trying to achieve an overall balancing.

Remark 4. *The above model is a generalization of model* min (4.16) *s. t.* (4.2), (4.3), (4.5), (4.6), (4.13) − (4.15)*. In fact, in case we set the costs $r_{ijs} = \infty$, $i, j \in I$, $s \in S$, all the v-variables become equal to zero in an optimal solution which leads to the redundancy of all constraints involving the latter variables.* ◇

Remark 5. *Due to the presence of the auxiliary variables ψ_{js} and φ_{js} ($j \in I$, $s \in S$) there is always a feasible solution to the above model. Hence, for every feasible first stage solution, there is always a feasible completion at the second stage, which means that we are dealing with a stochastic programming with complete recourse.* ◇

4.4 The relevance of considering a stochastic modeling framework

In the previous sections we have considered stochastic programming modeling frameworks for a districting problem under uncertainty. One important question in this case concerns the relevance of considering such frameworks instead of simpler ones (e.g. deterministic) . Two measures are usually considered for evaluating the relevance of capturing stochasticity in an optimization problem: the value of the stochastic solution (VSS) and the expected value of perfect information (EVPI). The reader can refer to Birge and Louveaux, 2011 for further details. Next we specialize these measures to the more general problem that we considered: SDPR.

Consider the model min (4.17), s. t.(4.2), (4.3), (4.5), (4.6), (4.14), (4.15), (4.18)−(4.23). Denote by SP its optimal value.

We can consider now the corresponding single-scenario problem in which the "single scenario" is the one induced by replacing all the random variables by their expected values. By solving that model, we obtain a first stage solution, say $\hat{\mathbf{x}}$ that is also feasible for the stochastic problem. As an output of the expected value problem, we can also denote it as Expected Value (EV) solution. Accordingly, we can evaluate its cost as a feasible solution to the stochastic problem. To do so, we just need to fix the values of the x-variables according to $\hat{\mathbf{x}}$ in the problem min (4.17), s. t.(4.2), (4.3), (4.5), (4.6), (4.14), (4.15), (4.18)−(4.23) The resulting value is denoted by EEV.

The value of the stochastic solution is computed as VSS=EEV-SP. This non-negative value indicates how good is the optimal solution to the expected value problem as an approximation to the optimal solution to the stochastic problem. A small VSS indicates that the latter can be very well approximated by the former which questions the relevance of considering a stochastic programming modeling framework in that case.

The EVPI is a measure of how much a decision maker would be willing to pay to get access to perfect information. A high EVPI indicates that having access to that information is quite relevant which means that uncertainty plays an important role in the problem. In order to compute the expected value of perfect information we start by solving the single scenario problem for every possible scenario. If we denote by $\mathcal{V}_s$ the corresponding optimal value, we can compute the so-called wait-and-see solution value: WS=$\sum_{s\in S}\pi_s\mathcal{V}_s$. The expected value of perfect information is computed as EVPI=SP-WS.

We consider the above measures in the next section.

4.5 Computational experiments

In this section, we report on the computational tests performed to test the general modeling framework proposed for SDPR. Such tests aimed at checking whether the new models work, i.e., whether they render plausible solutions. With this purpose, we focus the computations on a data set that makes use of real geographical data.

In Section 4.5.1 we describe the test data used. In Section 4.5.2 we focus on two specific instances in order to analyze in detail the plausibility of the solutions obtained and also to illustrate the relevance of capturing stochasticity when making a decision. Finally, in Section 4.5.3 we present the value of the stochastic solution and the expected value of the perfect information for all the instances considered in this study

4.5.1 Test data

Our stochastic programming modeling framework has been tested using real geographical data corresponding to the province of Novara, in northern Italy. This is a province divided into 88 municipalities (ISTAT, 2011) that we take as the reference TUs for our study. Making use of the GIS software QGIS we discretized the region by determining the centroids of the municipalities. Afterwards, we identified set I with the set of the extracted centroids, where we suppose that demands d_i are concentrated. Euclidean distances ℓ_{ij} between all pairs of centroids $i, j \in I$ have been computed. The province of Novara, its municipalities, and their centroids are depicted in Figure 4.1.

Concerning the demand vector $\boldsymbol{\xi}$, we worked with three scenarios. Since no real data could be found in terms of the demand for some real services we decided to generate randomly the missing data. In particular, a so-called intermediate scenario was obtained by randomly generating the 88 demands according to a uniform distribution in the range $[1, 10]$. Then, two additional scenarios were computed by considering a 20% positive and negative deviation respectively, from the *intermediate* scenario. The generated demands

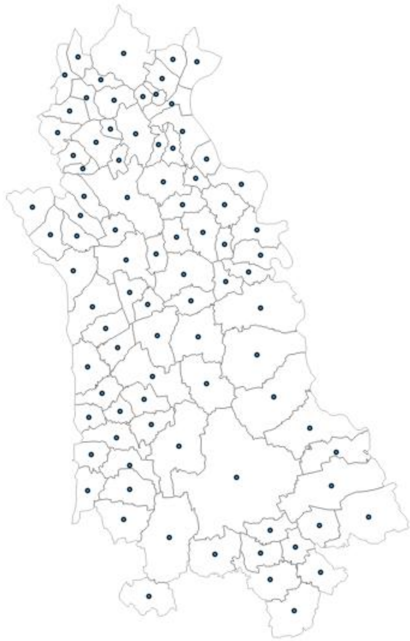


Figure 4.1. Novara province: the municipalities and the corresponding centroids.

and the coordinates of the centroids are available in Appendix I - Table .7.

The reason for considering only three scenarios has to do with the purpose of our computational tests: to check the relevance of the models proposed in the previous sections. In fact, if such relevance can be observed when only three scenarios are considered, then it should be even stronger in a larger setting.

Four probability distributions, indexed in $K = \{1, 2, 3, 4\}$, were considered inducing an equal number of base data sets for the computational tests. In particular, we denote by π_{sk} the probability of scenario s according to the k-th probability distribution ($k \in K$). The information is provided in Table 4.1.

s (scenario)	π_{s1}	π_{s2}	π_{s3}	π_{s4}
1 (20% below intermediate)	1/3	1/6	1/6	2/3
2 (intermediate)	1/3	1/6	2/3	1/6
3 (20% above intermediate)	1/3	2/3	1/6	1/6

Table 4.1. Probability distributions across the scenarios.

A parameter that needs to be define for specifying an instance is the reference value $\overline{\mu}$ to be used in the balancing constraints. In our case we set this value dependent on the underlying probability distribution as follows:

$$\bar{\mu}_k = \frac{1}{p} \sum_{i \in I} \sum_{s \in S} \pi_{sk} d_{is}, \quad k \in K.$$

Demands have been chosen as weighting factors in the computation of the assignment costs. In particular for the instance associated with probability distribution k ($k \in K$) we defined:

$$c_{ij}^{(k)} = \ell_{ij} \sum_{s\in S} \pi_{sk} d_{is}, \quad i,j \in I.$$

Concerning the re-assignment costs, they were defined as follows:

$$r_{ijs} = \omega\, d_{is}\, \ell_{ij}, \quad i,j \in I, s \in S.$$

We make the above costs dependent on the observed demand, the euclidean distance between the centroids of interest and a parameter ω that accounts for the penalty due to the reassignment. The chosen expressions for generating the reassignment costs resemble those used in the redistricting problems addressed by Caro et al. (2004) and Bruno et al. (2017b).

Remark 6. *In case $\omega = 1$ we have $c_{ij}^{(k)} = \sum_{s\in S} \pi_{sk} r_{ijs}$. This means that we have a relation between the initial assignment costs and the second stage (re)assignment costs. This relation allows us to quantify the expected "total second stage (re)assignment cost", which can be accomplished by considering the expression $\sum_{i\in I}\sum_{j\in I}\sum_{s\in S} \pi_{sk} r_{ijs} w_{ijs}$. We recall that w-variables define the second stage territory design in which some TUs remain assigned like in the first stage (the w-variable coincides with the corresponding x-variable but some other TUs are assigned to other TU (a reassignment occurs). This expression is directly comparable with $\sum_{i\in I}\sum_{j\in I} c_{ij} x_{ij}$.*

In other words, if $\omega = 1$ we can directly compare the (expected) compactness of the second stage territory design with the compactness of the first stage one. ◇

In our experiments, in addition to $\omega = 1$ (to allow the comparison detailed in the above remark) we also considered $\omega = 2$.

The unitary costs for shortage and surplus demand (w.r.t. the reference value) were defined as follows:

$$g_j^{(k)} = h_j^{(k)} = M_k = \max_{i,j\in I}\{\ell_{ij} \sum_{s\in S} \pi_{sk} d_{is}\}, \qquad j \in I,\ k \in K.$$

The above values may seem too high. However, we think they are appropriate to foster a reallocation mechanism by adapting the solution to the occurring demands.

Finally, in order to better test the behavior of our modeling framework, parameter α was varied from 0.05 up to 0.30 with a step of 0.05, while p was set equal to 4 and 6.

In total, we have generated 96 instances: four values for k, two values for ω, six values for α and two values for p.

All the instances were solved using the commercial solver IBM CPLEX v12.8 on an Intel(R) Celeron(R) with 1.50 GHz and 4 GB of RAM.

Next, we report on the computational results obtained by the implementation of the models proposed for the SDPR. In particular, we focus on model (M_2) which includes (M_1) as a particular case. For the data being considered, this is a model with 7744 binary variables, 23760 continuous variables and 31857 constraints.

4.5.2 First observations

We start by analyzing one specific instance to illustrate the relevance of capturing uncertainty in districting problems.

The instance considered at this stage is the one defined by $p = 4$, $k = 3$, $\omega = 1$, and $\alpha = 0.20$. The results obtained by using the model proposed for the SDPR are depicted in Figure 4.2. In this Figure, we associate different colors to different districts. The TUs selected as districts' centers or representatives are depicted in yellow. We also note that each TU is labeled with a unique ID code.

According to the parameters' setting, the model groups all the TUs in to four districts. In Figure 4.2a we can observe that the stochastic model rendered a first stage solution with very compact clusters. The cost associated with this territory design is $\sum_{i\in I}\sum_{j\in I} c_{ij}x_{ij} = 3,299.38$. Additionally we observe that apart from TU 44, contiguity holds although it was not explicitly imposed.

In Figures 4.2b–4.2d we can observe the second stage solutions—one for each possible realization of the demand vector. Looking into these figures we conclude that when demand is disclosed, the model "suggests" some reassignments in order to meet balancing constraints without incurring in too high penalty costs. For instance, when scenario 1 ("below intermediate") is observed, 6 TUs (16, 50, 53, 59, 74, 87) are reallocated from the orange and cyan districts to the dark blue one. This is a way to let the latter to comply with the minimum demand threshold. Similarly, when scenario 3 ("above intermediate") occurs 7 TUs (19, 53, 59, 74, 79, 83, 87) are reassigned to avoid above-threshold demand in the districts. To accomplish this, for example, the model "suggests" the reallocation of TUs 79 and 83 to the orange district. According to the chosen probability distribution ($k = 3$), the first stage solution remains unchanged when scenario 2 ("intermediate") is considered.

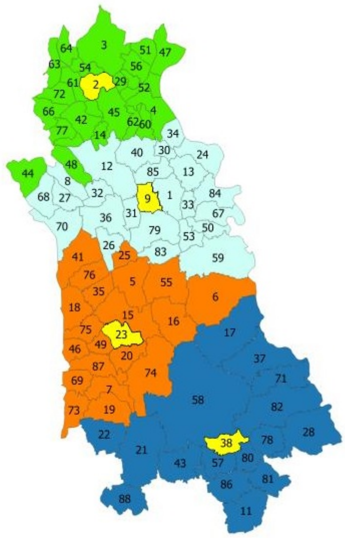
Finally, we note that the total expected reassignment cost is $\sum_{i\in I}\sum_{j\in I}\sum_{s\in S}\pi_{sk}r_{ijs}v_{ijs} = 163.82$. Since the total penalty cost is equal to zero (all variables ψ_{js} and φ_{js} are equal to zero) we obtain $3,299.38 + 163.82 = 3463.2$ as the optimal value to the stochastic problem.

Not surprisingly, we observe that the second stage (re)allocation decisions lead to less quality solutions both in terms of contiguity and compactness. The latter is testified by the expected total (re)assignment costs $\sum_{i\in I}\sum_{j\in I}\sum_{s\in S}\pi_s r_{ijs}w_{ijs} = 3,396.09$ (see Remark 6).

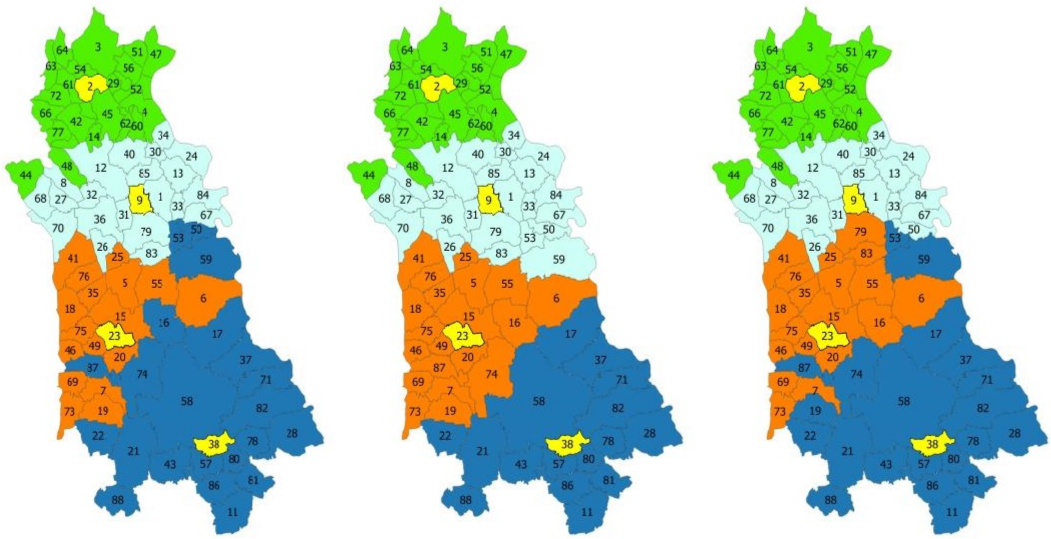
Overall, in this illustrative instance we observe what we would expect (and wish) in practice: a major plan is initially devised (the first stage solution) that suffers several minor changes according to how uncertainty is revealed. In fact, a major change in the first stage solution would indicate that it was not hedging appropriately against uncertainty.

In a second phase, we want to see the effect of changing the penalty factor ω and thus we considered the instance similar to the previous one but with $\omega = 2$. In Figure 4.3 we depict side by side the first stage solution for both instances. Figure 4.3a is the same as Figure 4.2a but we repeat it for the sake of an easier comparison.

When $\omega = 2$, the penalty for reassignment decisions becomes higher than before. In



(a) First stage



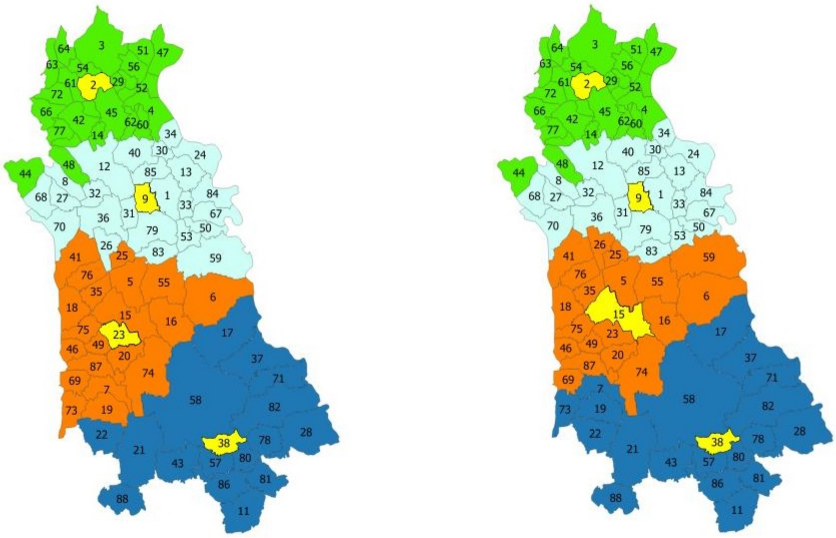
(b) Second stage: Scenario 1

(c) Second stage: Scenario 2

(d) Second stage: Scenario 3

Figure 4.2. Solutions for the instance defined by $p = 4$, $k = 3$, $\omega = 1$, $\alpha = 0.20$.

this case, the model suggests no second stage change independently from the demand observed. However, in order to better hedge against uncertainty, some changes are observed in the first stage solution: for $\omega = 2$, TUs 7, 19, 73 are assigned to the dark blue district, while units 26 and 59 are included in the orange cluster, whose representative is now TU 15. These changes are reflected in a lower compactness of the proposed solution that is reflected by a higher value of the objective function of the model, which increases up to 3,498.56.



(a) $\omega = 1$ **(b)** $\omega = 2$

Figure 4.3. First stage solutions for the instances with $p = 4$, $k = 3$, and $\alpha = 0.20$.

Finally, in addition to determining the solutions for the above two instances we computed %VSS ((EEV-SP)/SP*100) and the %EVPI ((SP-WS)/SP*100). As explained before, this is a means to quantify the relevance of considering the stochastic approach we have proposed.

We start by computing the solution for the single scenario (deterministic) problem that emerges from replacing the demands by their expectation. Since the difference between the two above instances regards the value of ω it is easy to conclude that the single scenario problem has the same optimal solution for both $\omega = 1$ and $\omega = 2$ since, when we consider one single scenario, no reassignment is part of an optimal solution. The resulting solution is depicted in Figure 4.4. Clearly, this solution differs from the optimal first stage solutions depicted in Figure 4.3.

The optimal solution to the expected value problem presents a higher compactness value than the two first stage solutions previously analyzed (the territory design costs, $\sum_{i \in I} \sum_{j \in I} c_{ij} x_{ij}$, have decreased to 3,249.18), it turns out to be very weak when implemented as a first stage solution in our stochastic problem. In fact, when scenarios 1 and 3 occur the model forces the reallocation of 10 TUs in order to meet the balancing

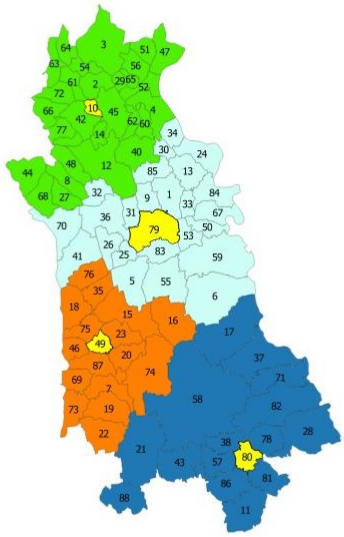


Figure 4.4. Optimal solution to the expected value problem ($p = 4$, $k = 3$, $\alpha = 0.20$).

constraints (Figure 4.5). Thus, we obtained %VSS equal to 3.87% ($\omega = 1$) and to 12.77% ($\omega = 2$). This shows that for the two instances considered the expected value problem provides a poor approximation to the stochastic problem, which indicates a high relevance of capturing uncertainty in these cases.

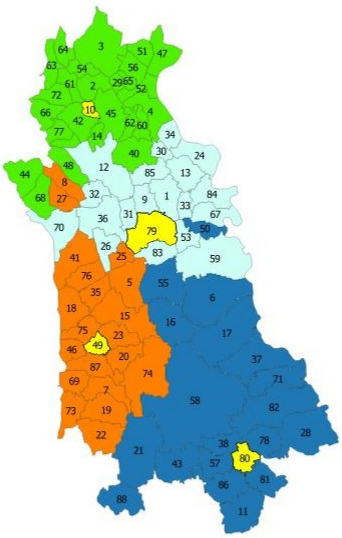
Finally, we computed the EVPI as explained in the previous section. This was accomplished by solving the single scenario problem for every possible scenario. We obtained 6.00% ($\omega = 1$) and to 17.64% ($\omega = 2$). Again, we conclude for a clear relevance of the introduced stochastic approach in the analyzed cases.

4.5.3 Computational results

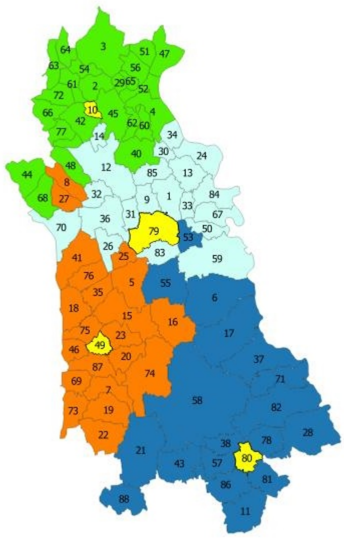
Having analyzed two instances in detail, we proceeded with results obtained by solving the mathematical model proposed for the SDPR for all the generated test instances. For obvious reasons we do not present the solutions obtained. Instead, we focus on the measures that help us to quantify the relevance of capturing stochasticity in our problem: the VSS and the EVPI. As before, we present percentage values.

The results obtained for the % VSS are depicted in Figure 4.6. The values used in these figure can be found in Appendix J and K - Tables .8 and .9.

The first aspect emerging from observing Figure 4.6 is that we obtain larger values for $\omega = 2$. This is not surprising. In fact, as explained above, the optimal solution to the expected value problem does not change with ω (in a single scenario, no relocation occurs because we can directly plan optimally for that scenario). Therefore, when that solution is considered as a first stage solution for the stochastic problem, possible relocation will cost more when $\omega = 2$ than when $\omega = 1$ which explains that the optimal solution to the expected value problem provides a better approximation for the stochastic problem when



(a) Scenario 1



(b) Scenario 3

Figure 4.5. Map of the EEV solutions ($p = 4, \alpha = 0.20, k = 3$)

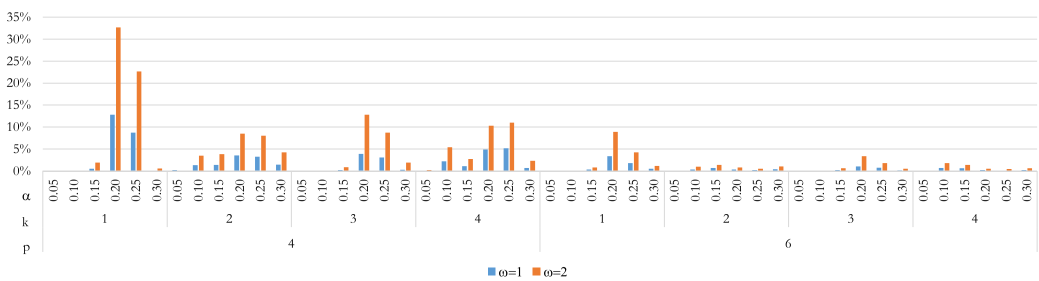


Figure 4.6. % VSS

$\omega = 1$. Overall, we observe an average % VSS equal to 1.29 for $\omega = 1$ and 3.60 for $\omega = 2$. Additionally, the %VSS decreases with an increase in p. This means that when more districts are considered, the expected value problem provides a better approximation to the stochastic problem. This indicates that with a larger number of districts we typically observe fewer reassignments required in the second stage.

When we focus our analysis to the pairs (ω, p), we observe that probability 1 seems to "dominate" the others in terms of % VSS. This is an indication that a stochastic approach is less effective when one of the possible scenarios is more likely.

When we focus our attention in the pairs (k, p), we can see that the value of the stochastic solution is typically higher for $\alpha = 0.20, 0.25$. In fact, on average, % VSS is equal to 14.32% when $\alpha = 0.20$ and $\omega = 2$. Given the generated demand scenarios, on one hand, the high penalties paid for demand shortages or surplus make the two approaches substantially equivalent when $\alpha \leq 0.15$. On the other, no significant differences are detected when balancing constraints are highly relaxed ($\alpha = 0.30$). Not by chance, all the instances characterized by a null VSS are found in correspondence of these values of α.

The above comments suggest a deeper look into the solutions we are obtaining. In our objective function we are considering the (re)assignment costs as well as the penalty costs for shortage/surplus demands w.r.t. the reference values. Since the penalty costs are high, they may blur a solution feature of great relevance to us: the compactness. Therefore, we decided to take the instances for which penalties are observed and focus only on the (re)assignment costs which, as we have explained before, are a reliable measure of compactness of the solution. In particular, we computed a measure that we call the compactness value of the stochastic solution (CVSS) and that is computed as the VSS (i.e. for the same solutions) but ignoring the penalty costs (i.e. setting them to 0). The results can be observed in Figure 4.7. Overall, these results show that although the values % VSS are rather small, in terms of compactness, the expected value solution provides a poor approximation to the stochastic problem.

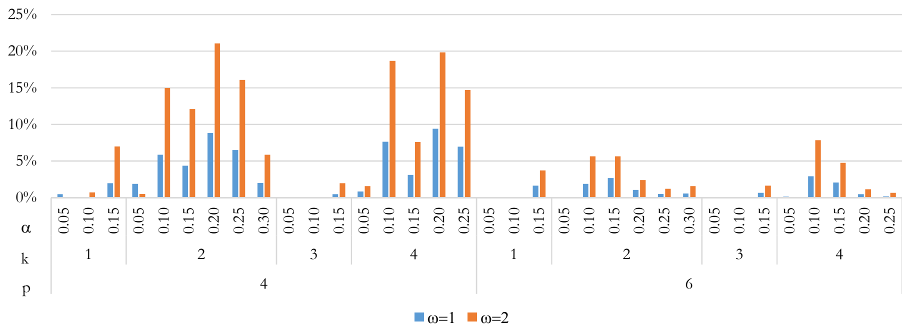


Figure 4.7. % *CVSS*

In Figure 4.8 we present the results obtained for % EVPI. The detailed results can be found in Appendix J and K - Tables .8 and .9. In this case, our computational experience reveals quite significant % EVPI values. On average, the % EVPI is equal to

50.00 % for $\omega = 1$ and 52.36 % for $\omega = 2$. These high values give an indication that capturing uncertainty in our districting problem is of great relevance. Moreover, the behavior of the EVPI is clear: it is rather insensitive to the adopted value of p and it shows a decreasing trend w.r.t. α. Not surprisingly, the lower the parameter α the higher the risk of observing demand surplus or shortages. Therefore, a decision maker would be willing to pay a higher price to know perfect information about the future.

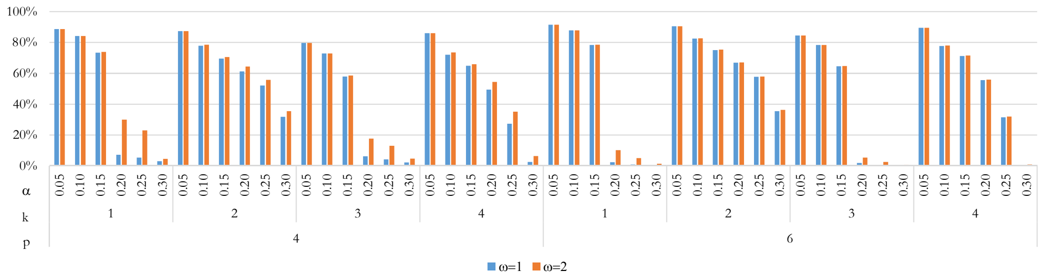


Figure 4.8. % EVPI

Lastly, it is important to report on the CPU time required to solve our stochastic model. This information is depicted in Figure 4.9 and detailed in in Appendix J and K - Tables .8 and .9.

Overall, we conclude that the model could be solved for all the instances within a reasonable CPU time. Most of the instances were solved in less than 1000 seconds and often much below that. The extreme cases are specified in Figure 4.9. The observed values are of relevance since they show that a stochastic districting problem such as the one that we are investigating in this chapter can be solved to optimality using tools that are available to most practitioners. We also note that the instances with $p = 6$ seem "easier" to solve than those with $p = 4$. We can also observe a tendency for higher CPU times with $\omega = 1$ than with $\omega = 2$. In fact, for the former we observed an overall average of 342 seconds while for the latter we observed 264 seconds. The averages per value of p can be found in Appendix J and K - Tables .8 and .9.

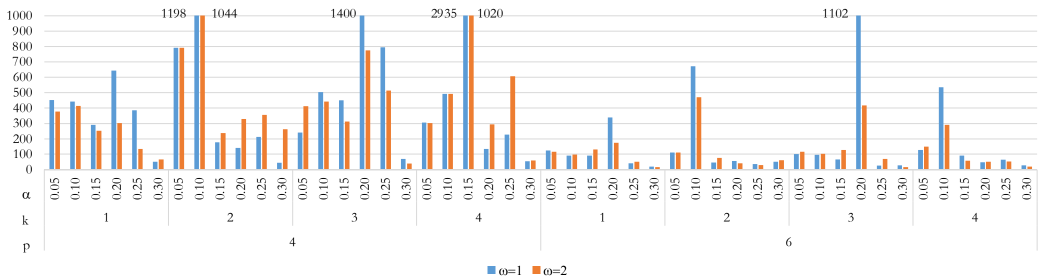


Figure 4.9. CPU time (seconds).

4.6 Extensions of the proposed model

In this section, we discuss some possible extensions to the problem investigated in the previous sections. These extensions are motivated by the need to include features of practical relevance that were not considered in our formulation. Examples of such features include territory dispersion and similarity w.r.t. to an existing districting plan. Furthermore, it may be relevant to manage some practical issues related to the second stage decisions. In fact, the reassignment recourse action that we have considered can lead to disadvantageous allocations for TUs as we discuss below.

The additional aspects discussed in this section will be modeled mathematically by the introduction of new sets of constraints in the optimization model presented in Section 4.3. In particular, we adopt all the assumptions and modeling considerations presented in that section.

4.6.1 Dispersion

As we mentioned directly in Section 4.1, territory dispersion indicates the maximum distance between any TU and the center of the district it is assigned to (Ríos-Mercado, 2016; Ríos-Mercado and Escalante, 2016). In the context of the stochastic districting problem we are studying, this aspect can be dealt with in different ways.

Let us suppose that there is a maximum desirable dispersion, say $l_{\max}$, for the districting to be obtained. This can be ensured for the first stage solution by considering the following constraints:

$$\ell_{ij}x_{ij} \leq l_{\max}, \qquad i,j \in I, \tag{4.24}$$

Moreover, if a maximum value for the dispersion should hold for the solution emerging in the second stage we must consider the following inequalities:

$$\ell_{ij}w_{ijs} \leq l_{\max}, \qquad i,j \in I, s \in S, \tag{4.25}$$

An alternative in our case is to consider the dispersion and the minimum difference between the dispersions at the first and second stages, as outcomes from the model. Denote by u_1 the first stage dispersions and by u_{2s} the second stage dispersion under scenario $s \in S$. The new setting can be expressed mathematically using the following constraints to be included in our stochastic model (M_2):

$$\ell_{ij}x_{ij} \leq u_1, \qquad i,j \in I, \tag{4.26}$$
$$\ell_{ij}w_{ijs} \leq u_{2s}, \qquad i,j \in I, s \in S, \tag{4.27}$$
$$z_s \geq u_{2s} - u_1, \qquad s \in S, \tag{4.28}$$
$$z_s \geq 0, \qquad s \in S, \tag{4.29}$$
$$u_1 \geq 0, \tag{4.30}$$
$$u_{2s} \geq 0, \qquad s \in S. \tag{4.31}$$

In the above relations, variables z_s ($s \in S$) are auxiliary variables that help us to quantify the difference between the first and second stage dispersions. In fact, Constraints (4.28) and (4.29) are simply the linearization of

$$z_s \geq \max\{u_{2s} - u_1\}, \qquad s \in S.$$

Considering a unitary penalty cost f_s and adding to the objective function (4.17) the term $\sum_{s \in S} \pi_s f_s z_s$, we know that in an optimal solution we will observe $z_s = \max\{u_{2s} - u_1\}$, $s \in S$. Therefore, we will incur in a penalty cost only in case of an effective increase in the maximum territory dispersion. Playing with the value of f_s we may discourage such increase. In this setting we can also add to the objective function a term accounting for the minimization of u_1.

4.6.2 Similarity with respect to an initial plan

Another relevant aspect often considered in DPs, regards the need to guarantee a certain degree of "similarity" w.r.t. to an original districting plan (Bozkaya et al., 2003; Caro et al., 2004; Bruno et al., 2017a). Let us denote by I_j the subset of TUs "similar" to j. This property can be ensured by means of the below proposed inequalities:

$$\sum_{i \in I_j} x_{ij} \geq \gamma |I_j| x_{jj}, \qquad j \in I, \tag{4.32}$$

$$\sum_{i \in I_j} w_{ijs} \geq \gamma |I_j| x_{jj}, \qquad j \in I, s \in S, \tag{4.33}$$

where $\gamma \in [0, 1]$ defines the minimum percentage of TUs to be kept together in the same district according to the initial plan. Two extreme cases can be identified: when $\gamma = 0$, districts are created from scratch; when $\gamma = 1$, the existing plan is fully preserved.

The stochastic setting that we introduced in section 4.3 for districting problems opens other modeling possibilities. For example, we may consider similarity constraints only associated with the second stage in order to ensure that in that stage we redesign districts keeping a certain degree of similarity w.r.t. the first stage districting plan. In this case, $I_j = \{i \in I : x_{ij} = 1\}$, $j \in I$. Constraints (4.32)–(4.33) can be specified for this case:

$$\sum_{i \in I} x_{ij} w_{ijs} \geq \gamma \sum_{i \in I} x_{ij}, \quad j \in I, s \in S, \tag{4.34}$$

The above expression can be linearized by introducing the auxiliary binary decision variables a_{ijs}, such that:

$$a_{ijs} = \begin{cases} 1, & \text{if TU } i \text{ is not reassigned under scenario } s; \\ 0, & \text{otherwise.} \end{cases} \qquad (i, j \in I,\ s \in S),$$

The following linear constraints can now be considered:

$$a_{ijs} \leq x_{ij}, \qquad i,j \in I, s \in S, \tag{4.35}$$
$$a_{ijs} \leq w_{ijs}, \qquad i,j \in I, s \in S, \tag{4.36}$$
$$a_{ijs} \geq x_{ij} + w_{ijs} - 1, \qquad i,j \in I, s \in S, \tag{4.37}$$
$$a_{ijs} \in \{0,1\}, \qquad i,j \in I, s \in S, \tag{4.38}$$

Finally, Constraints (4.34) can be rewritten as follows:

$$\sum_{i \in I} a_{ijs} \geq \gamma \sum_{i \in I} x_{ij}, \qquad j \in I, s \in S, \tag{4.39}$$

4.6.3 Reallocation constraints

In the SDPR that we are studying, second stage decisions ensure TUs reallocations to avoid demand shortages or surplus in the created districts. However, as shown by our computational experiments, adapting a solution to the occurring scenarios may disfavor the compactness (and thus contiguity) of the districts. Hence, a natural extension to our problem consists of limiting the number of reassignments in the second stage. This can be accomplished mathematically by adding the following constraints to our formulation:

$$\sum_{i \in I} \sum_{j \in I} x_{ij} w_{ijs} \geq |I| - m, \quad s \in S. \tag{4.40}$$

Constraints (4.40) ensure that in each scenario, at least $|I| - m$ TUs, assigned to j in the first stage, will be not reassigned in the second stage ($x_{ij}w_{ijs} = 1$, $i, j \in$, $s \in S$), where m is the maximum number of allowed reassignment in the second stage.

The above expression can be linearized by making use of the a-variables introduced above together with Constraints (4.35)–(4.38) and also

$$\sum_{i \in I} \sum_{j \in I} a_{ijs} \geq |I| - m, \qquad s \in S. \tag{4.41}$$

Moreover, specific reallocation rules can be also formulated to better drive the reassignment mechanism applied in the second stage. Given that the second stage decisions can force TUs to be assigned to farther districts centers, it may be reasonable to limit the increasing of TUs' (re)allocation distances. The expected effect is an improvement in the compactness of the second stage solutions.

$$\sum_{j \in I} \ell_{ij} w_{ijs} \leq (1+\delta) \sum_{j \in I} \ell_{ij} x_{ij}, \quad i \in I, s \in S \tag{4.42}$$

Constraints (4.42) ensure that in every scenario $s \in S$, TUs $i \in I$ can be reassigned only to district centers located at a distance not exceeding a δ % deviation from the first stage allocation one.

Finally, another alternative setting is the one in which we restrict the (second stage) reassignments only to a subset of TUs. In particular, one could allow reallocations only for those units located within a certain distance, say l', from their district center. This kind of constraints are particularly meaningful in those contexts in which equity and accessibility issues emerge. This is motivated by the need to discourage the reassignment of those TUs that are far from their district center in the first stage solution and that would become even farther by the reassignment. Mathematically we have:

$$\ell_{ij} x_{ij} - l' \leq M(1 - v_{ijs}), \qquad i, j \in I, s \in S, \tag{4.43}$$

The above constraints ensure, in fact, that if a unit i is located farther than distance l' from center j in the first stage solution (i.e. $\ell_{ij} x_{ij} - l' > 0$), it will be still assigned to j in the second stage solution (i.e. $v_{ijs} = 0, \quad i, j \in I, s \in S$). M denoted an arbitrary large value. In this case, it would make sense to add the following set of constraints to ensure that remaining units could be reassigned only within l' distance:

$$\sum_{j \in I} \ell_{ij} w_{ijs} \leq l', \qquad i \in I, s \in S. \tag{4.44}$$

Remark 7. *Constraints (4.41) are effectively active if, w.r.t to a certain setting of the parameters, $k \leq k^*$, where k^* is the number of TUs reassigned in an optimal solution to the SDPR.* ◇

Remark 8. *For lower values of δ, Constraints (4.42) may favour only the reallocation of those TUs whose distance from districts' representative are currently high enough.* ◇

4.7 Conclusions

In this chapter we investigated a stochastic districting problem triggered by uncertainty in the demand vector. By assuming that uncertainty to be captured by a finite number of scenarios each of which occurring with a given probability, a mixed-integer two-stage stochastic programming framework was developed and tested computationally. A large number of instances built using real geographical data—and thus of realistic size—was tested.

The results show that all the instances could be solved to optimality using a general purpose solver. This aspect is of particular relevance for a practitioner interested in this type of problem but not mastering sophisticated stochastic programming tools. Moreover, by making use of appropriate measures the computational tests also highlighted that capturing uncertainty can be of great relevance in the districting problems studied.

The work presented suggests several future research directions. First, in addition to jointly considering several aspects discussed in Section 4.6, the discussion provided in Section 4.6.1 highlights the clear multicriteria nature that a districting problem may exhibit. This calls for multicriteria stochastic variants of districting problems. Another aspect that may be interesting to investigate concerns the possibility of capturing uncertainty using chance constraints. This means that instead of modeling the balancing requirements as hard constraints, a (high) probability would be imposed to satisfaction of these constraints. This possibility leads to a totally different modeling framework which, nonetheless, may be worth investigating. Finally, we note that we have dealt with uncertainty by adopting stochastic programming models. This implies that uncertainty can be captured by some joint cumulative distribution function. If this is not the case, then we may need to resort to other possibilities such as robust optimization. This may also be a promising research direction.

General Conclusions

Location decisions play a crucial role in both private and public sectors as they can have long term impact on operational performances and on service levels. In the public sector, in particular, the optimal positioning of schools, hospitals and emergency services (i.e. fire and police stations, ambulances, etc.) is fundamental to provide users with good (ideally equitable) accessibility conditions and timely response to intervention requests. The same applies to private firms offering services of general interest (postal offices, banks, pharmacies, etc.) in sectors partially regulated by regional or national bodies. The main goals in public services contexts typical refer to social cost minimization, universality of services, and equity, in terms of users' accessibility, as well as to economic efficiency objectives, given the general political background characterized by the reduction of public expenditures and by the revision of the welfare state.

In literature, the two classes of problems mainly used to address location decisions are *facility location* and *districting problems*.

In this thesis, we proposed mathematical models to address some applications emerging in the *spatial organization of public services*.

The aim of this thesis was to show how these models can be employed to address real-world applications thus acting as decision support systems for decision and policy makers. To this end, after an introduction to the above classes of problems, given in Chapter 1, we explored two different applications, in the context of health-care and postal sector, in Chapters 2 and 3 respectively. Finally, in Chapter 4, a stochastic programming modeling framework for districting problems has been introduced.

Based on the analysis of the extant literature, various contributions have been highlighted.

A first contribution reside in the novelties of the proposed mathematical models.

The model proposed in Chapter 2 designs a of two-level capacitated blood supply chain, that considers, as distinguishing features, minimum and maximum capacity for processing activities, modular capacities and the self-sufficiency goal.

The model proposed in Chapter 3 integrates facility location and districting by introducing location decisions concerning the points (i.e. postboxes) to be clustered. The model is also embedded in a constructive heuristic procedure to include workshift duration and workload balance requirements.

The two-stage stochastic program presented in Chapter 4 seems to represent the first attempt to address a districting problem with uncertain demand ignoring routing deci-

sions. The extensions therein proposed also aim at offering a comprehensive modeling framework for stochastic districting problems.
A second contribution reside in the novelties of the targeted applications which have been poorly explored in the extant literature. Nevertheless, in comparison with previous works, the case studies analyzed in Chapters 2 and 3 involve medium-large scale instances. Therefore, their applicability is not limited to local management problems.
Lastly, the possibility to solve the introduced models using a commercial solver (CPLEX) within relatively limited computational times, can also represent a valuable element for all the practitioners interested in this kind of problems.
The present thesis opens to different research directions. Effort should be put in order to develop further versions of the proposed models adopting a multi-period setting. The introduction of the dimension ”time” in the models would enhance more realistic decision processes in which location and districting decisions are taken along a given planning horizon. Clearly, it would be also interesting to investigate models characterized by an integrated stochastic and multi-period setting.
The stochastic modeling framework proposed in Chapter 4 may be applied to tackle larger instances in order to test its effectiveness in addressing real-world case studies, supplied by real data.
Other promising research directions may be oriented to introduce and compare different modeling frameworks for districting based on single stage models using chance-constraints programming or robust optimization.
Finally, the multi-objective nature of the considered problems may also call for multi-criteria stochastic variants of districting problems.

Appendix

.1 Appendix to Chapter 2

In this Appendix we show the test data used to realize our computational experiments and we present the detailed results that were reported in Section 2.5. In particular, in Table .2 we report the data used to value cost parameters c_s, $c_{processing}$ and Δc^1. In Table .3 we report the coordinates of BCs and Hospitals currently located in the Campania Region. Table .4 reports the computational results obtained in terms of number of activated facilities and production levels. Finally, Table .5 reports the estimation of distribution costs presented in Section 2.6.4.

Appendix A

Here we report information used for the evaluation of the costs considered in the model. As suggested by the Regional Authority, such costs have to be estimated with reference to the "standard operational conditions", corresponding to the maximum collection and transformation capacity (30,000 and 50,000 blood units per year, respectively). The number of professionals needed to perform both the activities in standard operational conditions is reported in the following table. It is worth noticing that, in providing these estimations, the Regional Authority took into account data regarding staff rostering and holidays and considered that both the activities are performed within a single work shift per day, 200 days a year. Therefore, given the ratio between the total costs of personnel involved in both the activities, and also considering an aliquot of avoidable costs borne by equipment in the production phase, we set $c_s = 1$ and $c_{processing} = 2$. Accordingly, we set $\Delta c^1 = 1.6$.

Table .2. Number of professionals by activity and yearly gross salary (source: Regional Authority)

Professional	Required Number		Yearly gross salary [€]
	Collection	Production	
Doctors	3	5	65,000
Nurses	5	/	26,500
Administrative	2	/	25,500
Technicians	/	10	36,200

Appendix B

Table .3. Coordinates of BCs and Hospitals (Coordinates Reference System: ED50 / UTM Zone 32N EPSG:23032)

ID	East	North
1	4538891.6130	941636.4899
2	4536635.0260	940149.8653
3	4535065.8240	942383.4021
4	4536933.7270	940484.6093
5	4538058.2400	944765.7085
6	4533590.2980	941013.8773
7	4532522.6880	937126.3988
8	4535490.8600	942757.5913
9	4520214.8860	963825.4787
10	4537382.6560	939188.4607
11	4545299.6090	966643.6528
12	4562140.3430	948133.6905
13	4547215.1940	938436.8608
14	4568048.7730	986682.5204
15	4546720.0090	988479.6270
16	4573068.3330	1010646.9440
17	4525865.3790	977160.3325
18	4505577.4580	1050856.4880
19	4513113.1730	1005448.3890
20	4516263.2430	992837.7827
21	4514050.2980	1011881.4790
22	4471674.5600	1032976.7180
23	4546396.1250	943732.8760
24	4543262.4320	937918.9115
25	4522426.3230	913213.5432
26	4531308.3430	938643.3785
27	4541727.9180	945327.7060
28	4523223.1690	923494.3680
29	4533098.2540	930014.7884
30	4536190.3010	948719.7493
31	4531214.8220	954386.2088
32	4511114.1350	954583.4596
33	4560673.8330	942207.9348
34	4577938.2830	913554.9707
35	4563168.7360	962349.3488
36	4557516.4510	952376.0092
37	4555232.2540	945253.9324
38	4554791.3450	961363.1404
39	4591876.5920	948472.0830
40	4569876.5460	984636.7963
41	4578077.7340	963989.3984
42	4537148.9750	992311.8721
43	4559828.4580	1036242.2850
44	4549200.9070	1020408.0560
45	4523261.2040	1027152.7400
46	4456706.6230	1064771.8830
47	4533026.1620	972244.3393
48	4526175.1100	966523.8212
49	4493481.7250	1025336.9510
50	4483858.4270	1009023.0590

Appendix C

Table .4. Computational results: Number of activated facilities, production levels and CPU times

					Number of activated facilities		Production levels per BC				
ID Scenario					BSs	BCc	1	2	3	4	CPU (seconds)
1	20	150000	0,035	40	6	3	68791	41099	40159		1295
2	25	150000	0,035	40	5	2	75558	75878			1178
3	30	150000	0,035	40	4	2	78536	71772			1463
4	20	150000	0,0375	40	6	2	73962	78948			1225
5	25	150000	0,0375	40	5	2	77352	79951			1195
6	30	150000	0,0375	40	4	2	77239	79445			1082
7	20	150000	0,04	40	5	2	76665	79277			1163
8	25	150000	0,04	40	5	2	79618	79279			832
9	30	150000	0,04	40	4	2	77601	75295			738
10	30	165000	0,035	40	6	3	45964	42720	78417		1194
11	20	165000	0,0375	40	7	3	43027	75041	47834		676
12	25	165000	0,0375	40	5	3	47842	49812	69016		844
13	30	165000	0,0375	40	4	3	77239	49224	40231		734
14	20	165000	0,04	40	6	3	44643	48677	77776		791
15	25	165000	0,04	40	4	3	48604	72447	44812		1096
16	30	165000	0,04	40	4	3	79616	43850	48961		619
17	25	172500	0,0375	40	6	4	40288	43470	41653	48844	1081
18	30	172500	0,0375	40	5	3	47352	77239	49224		1109
19	20	172500	0,04	40	6	3	77476	45896	49494		1204
20	25	172500	0,04	40	5	3	71950	54744	46620		1500
21	30	172500	0,04	40	5	3	49393	49780	78206		975
22	20	180000	0,04	40	6	4	42856	45896	49983	43649	628
23	25	180000	0,04	40	5	4	49671	45089	43250	42028	1042
24	30	180000	0,04	40	4	4	48728	42158	43488	47065	1268
25	20	150000	0,035	50	6	2	79533	71251			1232
26	25	150000	0,035	50	5	2	75558	75878			904
27	30	150000	0,035	50	5	2	78205	75615			1220
28	20	150000	0,0375	50	5	2	76478	73942			689
29	25	150000	0,0375	50	4	2	78107	71897			1150
30	30	150000	0,0375	50	4	2	74642	75371			1079
31	20	150000	0,04	50	5	2	77126	73077			1041
32	25	150000	0,04	50	4	2	76237	74374			638
33	30	150000	0,04	50	4	2	73454	76614			1500
34	25	165000	0,035	50	6	3	79863	43259	41926		924
35	30	165000	0,035	50	5	3	79994	40808	45379		1443
36	20	165000	0,0375	50	6	3	43613	45780	76584		1264
37	25	165000	0,0375	50	5	3	46745	72541	45883		1405
38	30	165000	0,0375	50	4	3	79484	41434	44657		862
39	20	165000	0,04	50	5	3	44262	79869	40899		702
40	25	165000	0,04	50	4	3	77665	44812	42640		1492
41	30	165000	0,04	50	4	3	42158	78068	44926		1002
42	30	172500	0,035	50	7	3	74896	42361	55273		1064
43	20	172500	0,0375	50	7	3	79788	48879	44341		656
44	25	172500	0,0375	50	6	3	77584	49510	45621		722
45	30	172500	0,0375	50	4	3	59817	71410	41348		1399
46	20	172500	0,04	50	6	3	45896	78050	48919		862
47	25	172500	0,04	50	5	3	49519	76568	46596		1134
48	30	172500	0,04	50	4	3	73570	49900	49895		906
49	30	180000	0,0375	50	5	4	40134	46857	49224	44222	996
50	20	180000	0,04	50	6	4	41067	45896	43369	49968	1293
51	25	180000	0,04	50	5	3	79314	44209	56767		1293
52	30	180000	0,04	50	4	3	78789	53588	47634		1427

Appendix D

Table .5. Computational results - Estimation of distribution costs

ID Scenario	r	D	α	d_{max}	Average allocation distance [km]	
					p-median	Model solution
1	20	150,000	0.035	40	32.6	38.6
2	25	150,000	0.035	40	32.6	45.1
3	30	150,000	0.035	40	30.4	46.1
4	20	150,000	0.0375	40	32.6	46.1
5	25	150,000	0.0375	40	32.6	45.4
6	30	150,000	0.0375	40	30.4	47.4
7	20	150,000	0.04	40	32.6	51.7
8	25	150,000	0.04	40	32.6	47.4
9	30	150,000	0.04	40	30.4	52.2
10	30	165,000	0.035	40	26.5	37.6
11	20	165,000	0.0375	40	30.6	38
12	25	165,000	0.0375	40	28.4	44.9
13	30	165,000	0.0375	40	26.5	45.6
14	20	165,000	0.04	40	30.6	44.8
15	25	165,000	0.04	40	28.4	43.9
16	30	165,000	0.04	40	26.5	45.6
17	25	172,500	0.0375	40	28.4	37.2
18	30	172,500	0.0375	40	26.5	40.1
19	20	172,500	0.04	40	30.6	40.1
20	25	172,500	0.04	40	28.4	40.1
21	30	172,500	0.04	40	26.5	45
22	20	180,000	0.04	40	30.6	36.2
23	25	180,000	0.04	40	28.4	43.3
24	30	180,000	0.04	40	26.5	39.7
25	20	150,000	0.035	50	30.4	42.3
26	25	150,000	0.035	50	30.4	45.1
27	30	150,000	0.035	50	30.4	45.5
28	20	150,000	0.0375	50	30.4	47.7
29	25	150,000	0.0375	50	30.4	47.4
30	30	150,000	0.0375	50	30.4	51.9
31	20	150,000	0.04	50	30.4	40.7
32	25	150,000	0.04	50	30.4	50.3
33	30	150,000	0.04	50	30.4	47
34	25	165,000	0.035	50	26.5	35.9
35	30	165,000	0.035	50	26.5	35
36	20	165,000	0.0375	50	26.5	33.3
37	25	165,000	0.0375	50	26.5	38.4
38	30	165,000	0.0375	50	26.5	39.9
39	20	165,000	0.04	50	26.5	40.7
40	25	165,000	0.04	50	26.5	43.8
41	30	165,000	0.04	50	26.5	38.4
42	30	172,500	0.035	50	26.5	39.2
43	20	172,500	0.0375	50	26.5	38.9
44	25	172,500	0.0375	50	26.5	36.2
45	30	172,500	0.0375	50	26.5	41.1
46	20	172,500	0.04	50	26.5	38.9
47	25	172,500	0.04	50	26.5	38.9
48	30	172,500	0.04	50	26.5	44.1
49	30	180,000	0.0375	50	26.5	34.7
50	20	180,000	0.04	50	26.5	33.7
51	25	180,000	0.04	50	26.5	38.3
52	30	180,000	0.04	50	26.5	45.2

.2 Appendix to Chapter 3

In this Appendix we first discuss in more detail the discretization approach applied to the case study presented in Section 3.4.1. Afterwards, we show the correlation obtained between clusters' compactness and the collection times.

Appendix E

In this appendix, we discuss in more detail the discretization approach applied to our case study. We chose to use census tracts as basic units since they represent the lowest aggregation level adopted for statistical purposes by the Italian Statistical Institute (ISTAT). The analysis of the available census data reveals that, on average, almost 160 inhabitants populate each tract and there are fewer than 500 in 96% of the cases (ISTAT, 2011). This indicates that the proposed discretization does not yield a very coarse representation of the total demand since we avoid massive aggregation of users in the extracted centroids.

Clearly, considering the distribution of the real user locations (e.g. private residences, house numbers) would have been an alternative and more precise approach. However, this would have easily rendered the problem intractable because of the very large number of points to be considered (40,955).

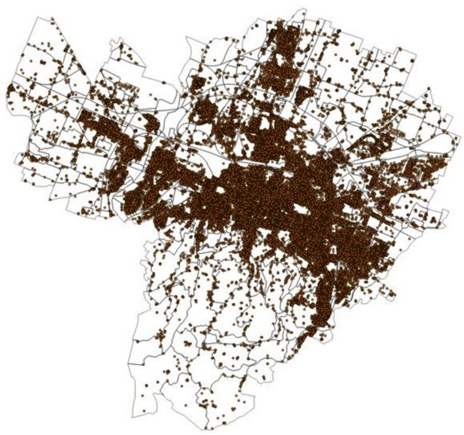


Figure .10. House numbers locations in Bologna

Of course, replacing Demand Points (DPs) with Aggregate Demand Points (ADPs) produces an intrinsic error that can be measured in different and alternative ways (Francis and Lowe, 2015). In order to evaluate this error, we refer to one of the indicators proposed by Francis and Lowe (2015), defined as the Distance Difference Error (DDE). To this aim, some extra notation is required. We denote by H the set of house numbers in Bologna ($|H| = 40{,}955$). In particular, we denote by H_i the set of house numbers located in a census tract whose centroid is $i \in I$ ($H = \cup_{i \in I} H_i$). We also indicate by l_{hj} the distance between nodes $h \in H$ and $j \in J$. Moreover, we define by $j_h^\star$ the closest postbox to node $h \in H$ ($j^\star_h = \{j \in J : l_{hj} = \min_{j \in J}\{l_{ij}\}$). A similar definition applies to the closest postbox to node $i \in I$, namely $j^\star_i$. Then the DDE is computed as DDE $= d_{ij^\star_i} - l_{hj^\star_h}$, $i \in I, h \in H_i$. A DDE equal to zero indicates that ADPs and DPs are identical. The results obtained (see Figure .11) show that in 95% of the cases (38,907 out of 40,955) the DDE is very low, ranging between -112.37 m and 145.00 m. The extreme cases, i.e. those with the highest absolute values of DDE, belong to the more peripheral tracts. However, their real impact on the solution of the model is actually rather limited. In fact, given the location mechanism we want to implement, the more peripheral census tracts always correspond to the most disadvantaged users. Therefore, since their closest postboxes will remain active, the real error we make in these cases is due to those house numbers whose closest postbox is different from that of the centroid ($h \in H_i : j^\star_h \neq j^\star_i, i \in I$). Figure .12 depicts the percentage of house numbers having a different closest postbox with respect to their census tract centroid. Our analysis shows that with some rare exceptions, these percentages are rather small, and often equal to zero, in the non-central areas. For instance, for $\alpha = 0.9$, only 190 out of 1,344 house numbers belonging to tracts $i \in I \setminus I'$ present a different closest postbox. This result implies that the magnitude of the DDE in the outer tracts is small in practice. Hence, based on the above findings, we consider the proposed discretization to be suitable for our problem.

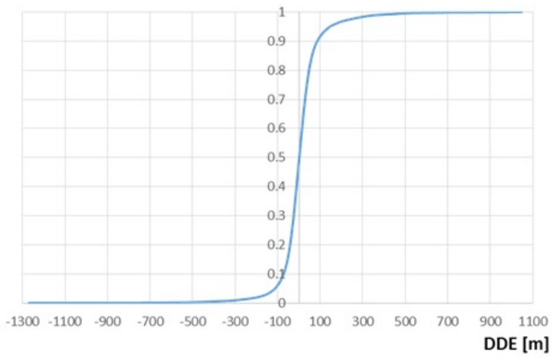


Figure .11. Distribution of DDE by the fraction of house numbers

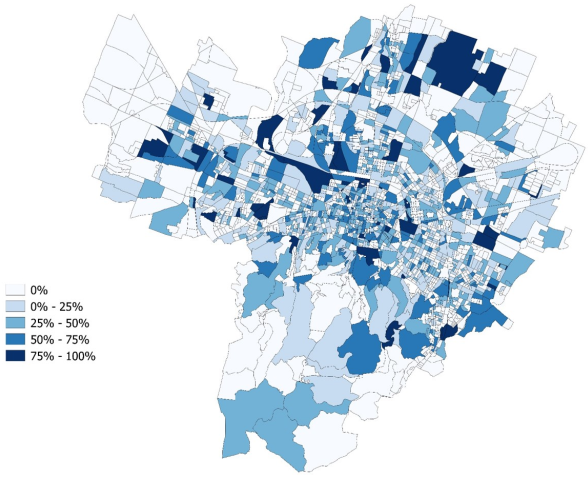


Figure .12. Percentage of house numbers presenting a different closest postbox

.

Appendix F. Accessibility considerations with real demand points

In this appendix we show that the accessibility considerations proposed still hold when considering the distribution of real demand points (i.e house numbers).

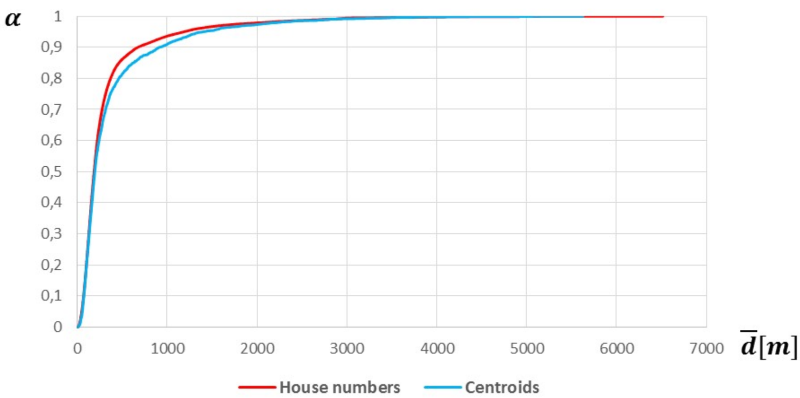


Figure .13. Distribution of the population, in terms of house numbers (red) and centroids (blue), by the distance from the closest postbox.

.

From the above figure it is possible to conclude that: i) the spatial criteria set by the regulatory authority are effectively respected for all the users in Bologna; ii) since the two curves are almost overlapping, accessibility considerations drawn upon the proposed aggregation are consistent with the actual distribution of demand points. Again, this allows us to consider census tracts' centroids a suitable representation of our problem.

Appendix G

In this appendix we show the correlation between the proxy measure of the travel times adopted in our model (i.e. clusters' compactness), and the collection times.

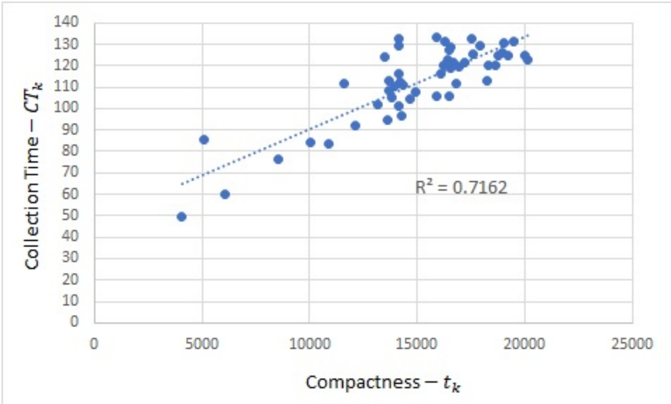


Figure .14. Correlation between clusters compactness and collection times

.

Appendix H

In this appendix we provide details about computational times for the considered instances.

Table .6. Computational times for the considered instances

α	β	p	UB	CPU (seconds)
1	1	2	22,000	/
	1	3	22,000	210
	0.2	3	22,000	720
	0.4	3	22,000	1,123
0.9	1	3	22,000	/
	1	4	22,000	/
	1	5	22,000	234
		5	20,831.18	341
		5	19,513.42	/
	1	6	22,000	101
	0.2	6	22,000	561
	0.4	6	22,000	1,320
0.8	1	5	22,000	/
	1	6	22,000	/
	1	7	22,000	923
		7	20,793.75	/
	1	8	22,000	1,015
		8	20,793.75	1,136
		8	19,662.62	1,245
	0.2	8	19,662.62	1,531
	0.4	8	19,662.62	1,723
0.7	1	6	22,000	/
	1	7	22,000	/
	1	8	22,000	/
	1	9	22,000	896
		9	20,778.42	1,073
		9	19,467.53	/
	1	10	22,000	480
		10	20,790.17	813
		10	19,592.17	1,021
		10	18,371.16	1,310
	0.2	10	18,371.16	1,549
	0.4	10	18,371.16	1,791
0.6	1	8	22,000	/
	1	9	22,000	/
	1	10	22,000	931
		10	20,708.56	1,090
		10	19,397.10	/
	1	11	22,000	511
		11	20,769.63	683
		11	19,559.55	791
		11	18,537.61	1,112
		11	17,569.20	1,254
		11	16,579.87	1,482
	0.2	11	16,579.87	1,756
	0.4	11	16,579.87	1,983
0	1	9	22,000	/
	1	10	22,000	/
	1	11	22,000	/
	1	12	22,000	632
		12	20,771.98	699
		12	19,661.82	834
		12	16,864.97	1,002
		12	16,012.69	/
	1	13	22,000	398
		13	19,779.87	478
		13	18,762.73	654
		13	17,789.15	801
	0.2	13	17,789.15	1,031
	0.4	13	17,789.15	1,383

.3 Appendix to Chapter 4

In this Appendix we show the test data used to realize our computational experiments and we present the detailed results that were reported in Section 4.5. For each TU, Table .7 reports the generated demand vector and the coordinates of the respective centroid. The original code assigned by ISTAT to uniquely identify TUs (i.e. PRO_COM) is also reported. For each tested instance, Tables .8 and .9 contain the % VSS, the % EVPI, and the CPU time in seconds required by the general purpose solver to solve the instance to optimality. Table .8 refers to the instances with 4 districts and Table .9 to instances with 6 districts.

Appendix I

Table .7. Demands and coordinates of the centroids.
(Coordinate Reference System: ED50 / UTM Zone 32N EPSG:23032.)

ID	PRO_COM	d_{is}			Coordinates	
		$s=1$	$s=2$	$s=3$	East	North
1	3001	6.75	8.44	10.13	466469.15	5057067.12
2	3002	5.89	7.36	8.84	457360.80	5070654.07
3	3006	1.46	1.82	2.18	458367.30	5075525.70
4	3008	3.36	4.20	5.04	464446.58	5067475.45
5	3012	2.49	3.11	3.73	461875.08	5046430.65
6	3016	7.89	9.87	11.84	472099.39	5044532.88
7	3018	7.69	9.62	11.54	458985.69	5033119.25
8	3019	7.19	8.99	10.79	453880.04	5058773.21
9	3021	1.06	1.33	1.60	463746.59	5056614.79
10	3022	2.17	2.71	3.25	457024.25	5067694.92
11	3023	5.52	6.91	8.29	475867.81	5018179.98
12	3024	7.57	9.47	11.36	458699.60	5060623.58
13	3025	7.86	9.82	11.78	468748.77	5059930.71
14	3026	2.48	3.10	3.72	457842.88	5064441.14
15	3027	7.04	8.80	10.56	461282.21	5042304.89
16	3030	1.76	2.21	2.65	466912.72	5041541.59
17	3032	4.50	5.62	6.75	473850.03	5040149.03
18	3036	7.25	9.06	10.87	454699.16	5043248.96
19	3037	4.98	6.22	7.46	458991.70	5030668.84
20	3039	7.10	8.87	10.64	461166.84	5037284.93
21	3040	6.94	8.67	10.41	463005.21	5025693.98
22	3041	1.63	2.04	2.44	458378.24	5027594.68
23	3042	6.77	8.46	10.15	460503.40	5039893.89
24	3043	2.56	3.21	3.85	470498.77	5062038.81
25	3044	7.03	8.79	10.55	460868.97	5049643.69
26	3045	7.87	9.83	11.80	458950.25	5050881.89
27	3047	2.61	3.26	3.91	453526.83	5056717.38
28	3049	5.17	6.46	7.75	483575.38	5027888.51
29	3051	5.45	6.81	8.17	460382.96	5071036.27
30	3052	3.91	4.89	5.87	465786.31	5062579.10
31	3055	6.07	7.58	9.10	461747.82	5054842.82
32	3058	5.42	6.77	8.13	457533.20	5057309.69
33	3060	6.13	7.67	9.20	468713.21	5055875.82
34	3062	0.86	1.08	1.29	466864.84	5064613.25
35	3065	1.94	2.43	2.92	457701.83	5045208.08
36	3066	7.86	9.83	11.79	458575.19	5054272.86
37	3068	1.51	1.89	2.27	477534.51	5036924.08
38	3069	3.11	3.89	4.67	473442.27	5026539.86
39	3070	5.22	6.52	7.83	454155.87	5063674.14
40	3071	2.06	2.57	3.09	462422.42	5062293.70
41	3073	1.68	2.10	2.52	455155.14	5049449.17
42	3076	2.00	2.50	3.00	455550.01	5066322.15
43	3077	3.51	4.38	5.26	467749.88	5024050.91
44	3079	6.21	7.76	9.31	448971.23	5059702.17
45	3082	1.68	2.10	2.52	459582.26	5067240.25
46	3083	1.03	1.28	1.54	454760.04	5038118.28
47	3084	5.33	6.67	8.00	465893.30	5074607.16
48	3088	7.05	8.82	10.58	454327.56	5060819.11
49	3090	3.36	4.20	5.03	457952.80	5038661.67
50	3091	2.26	2.83	3.39	471150.34	5053044.63

Continued on next page

Table .7 – *Continued from previous page*

ID	PRO_COM	d_{is} $s=1$	$s=2$	$s=3$	Coordinates East	North
51	3093	3.41	4.27	5.12	463470.03	5074904.46
52	3095	2.51	3.13	3.76	463306.10	5070284.71
53	3097	3.14	3.92	4.71	468822.99	5051981.15
54	3098	2.43	3.04	3.65	456024.66	5072731.63
55	3100	7.23	9.03	10.84	466041.49	5046326.82
56	3103	4.12	5.15	6.18	462321.93	5072918.07
57	3104	7.79	9.74	11.69	472386.07	5024115.73
58	3106	0.96	1.20	1.44	469940.81	5031855.83
59	3108	5.80	7.25	8.70	472407.20	5049331.80
60	3109	4.14	5.17	6.20	463432.25	5065757.38
61	3112	4.73	5.92	7.10	454556.78	5070880.61
62	3114	4.06	5.07	6.09	461944.95	5066077.72
63	3115	5.30	6.63	7.95	452274.44	5073187.92
64	3116	6.57	8.22	9.86	453719.37	5075116.23
65	3119	7.73	9.66	11.59	461721.07	5071260.50
66	3120	1.65	2.06	2.47	451532.47	5067352.78
67	3121	3.80	4.75	5.70	472451.22	5054797.69
68	3122	0.97	1.21	1.46	450911.26	5056848.38
69	3129	7.82	9.78	11.73	455062.48	5034226.27
70	3130	3.78	4.72	5.67	453178.69	5053187.48
71	3131	5.52	6.90	8.28	480191.40	5034504.94
72	3133	1.10	1.38	1.65	452854.75	5069497.90
73	3134	6.25	7.82	9.38	454650.76	5030566.17
74	3135	2.38	2.98	3.57	464109.73	5035165.96
75	3138	4.76	5.95	7.14	456097.22	5040500.81
76	3139	4.81	6.01	7.22	456553.87	5047228.17
77	3140	1.88	2.35	2.82	453174.04	5064999.29
78	3141	5.26	6.58	7.90	478459.44	5026931.12
79	3143	2.84	3.55	4.26	464592.97	5052731.03
80	3144	4.83	6.04	7.25	476071.81	5024724.15
81	3146	6.40	8.00	9.60	478582.56	5022122.45
82	3149	7.53	9.42	11.30	479715.01	5031009.29
83	3153	1.93	2.42	2.90	465341.02	5050075.12
84	3154	5.20	6.51	7.81	472051.38	5057268.59
85	3157	6.68	8.35	10.02	464398.06	5059931.68
86	3158	7.76	9.70	11.64	473466.51	5021464.46
87	3159	1.06	1.33	1.59	457661.23	5035969.14
88	3164	1.28	1.60	1.92	460929.97	5019673.68

Appendix J

Table .8. Computational results for instances with $p = 4$.

k	α	% VSS		% EVPI		CPU (seconds)	
		$\omega = 1$	$\omega = 2$	$\omega = 1$	$\omega = 2$	$\omega = 1$	$\omega = 2$
1	0.05	0.05	0.00	88.72	88.73	453	377
	0.10	0.01	0.12	84.28	84.30	442	415
	0.15	0.54	1.94	73.33	73.85	291	252
	0.20	12.77	32.66	7.13	29.99	643	303
	0.25	8.73	22.62	5.38	22.98	386	135
	0.30	0.00	0.56	2.95	4.39	50	66
2	0.05	0.23	0.06	87.40	87.42	791	792
	0.10	1.34	3.43	77.85	78.59	1198	1044
	0.15	1.38	3.83	69.49	70.63	177	238
	0.20	3.54	8.48	61.31	64.38	142	329
	0.25	3.24	8.06	52.07	55.82	212	355
	0.30	1.43	4.22	31.68	35.51	45	263
3	0.05	0.01	0.00	79.74	79.74	242	412
	0.10	0.00	0.02	72.83	72.86	505	441
	0.15	0.21	0.85	57.99	58.54	451	312
	0.20	3.87	12.77	6.18	17.64	1400	775
	0.25	3.05	8.73	4.10	12.98	795	514
	0.30	0.31	1.93	2.15	4.79	70	40
4	0.05	0.12	0.22	86.04	86.08	305	302
	0.10	2.21	5.40	72.06	73.49	492	492
	0.15	1.11	2.73	64.97	65.98	2935	1020
	0.20	4.85	10.26	49.52	54.41	135	295
	0.25	5.20	11.01	27.23	35.19	227	607
	0.30	0.70	2.32	2.58	6.30	54	60
Average		2.29	5.93	48.62	52.69	518	410

Appendix K

Table .9. Computational results for instances with $p = 6$.

		% VSS		% EVPI		CPU (seconds)	
k	α	$\omega = 1$	$\omega = 2$	$\omega = 1$	$\omega = 2$	$\omega = 1$	$\omega = 2$
1	0.05	0.00	0.00	91.56	91.56	124	116
	0.10	0.00	0.00	87.90	87.90	90	97
	0.15	0.36	0.81	78.46	78.64	90	131
	0.20	3.36	8.91	2.20	10.21	338	175
	0.25	1.77	4.24	0.79	4.92	41	51
	0.30	0.50	1.19	0.19	1.37	19	17
2	0.05	0.01	0.01	90.55	90.55	112	111
	0.10	0.34	1.00	82.58	82.75	671	471
	0.15	0.67	1.42	75.05	75.41	46	77
	0.20	0.35	0.80	66.91	67.18	56	41
	0.25	0.22	0.51	57.74	57.96	36	29
	0.30	0.36	1.02	35.56	36.26	51	62
3	0.05	0.00	0.00	84.44	84.44	101	117
	0.10	0.00	0.00	78.41	78.41	96	102
	0.15	0.24	0.59	64.58	64.80	67	128
	0.20	1.02	3.36	1.77	5.38	1102	418
	0.25	0.76	1.77	0.53	2.52	26	70
	0.30	0.15	0.50	0.19	0.69	28	17
4	0.05	0.02	0.00	89.47	89.48	128	149
	0.10	0.66	1.79	77.63	78.03	536	291
	0.15	0.61	1.38	71.11	71.51	91	57
	0.20	0.22	0.52	55.70	55.93	48	52
	0.25	0.13	0.45	31.49	31.88	65	52
	0.30	0.20	0.60	0.25	0.88	28	19
Average		0.50	1.29	51.04	52.02	166	119

Bibliography

Abbatecola, L., Fanti, M. P., Mangini, A. M., and Ukovich, W. (2016). A decision support approach for postal delivery and waste collection services. *IEEE Transactions on Automation Science and Engineering*, 13(4):1458–1470.

Aboolian, R., Berman, O., and Krass, D. (2007). Competitive facility location and design problem. *European Journal of Operational Research*, 182(1):40–62.

AGCOM (2014). Public consultation on the criteria for the distribution of access points to the public postal network. Italian Authority for Communication Guarantees. URL: www.agcom.int [accessed on June 10, 2018].

Ahmadi-Javid, A., Seyedi, P., and Syam, S. S. (2017). A survey of healthcare facility location. *Computers & Operations Research*, 79:223–263.

Akdoğan, M. A., Bayındır, Z. P., and Iyigun, C. (2018). Locating emergency vehicles with an approximate queuing model and a meta-heuristic solution approach. *Transportation Research Part C: Emerging Technologies*, 90:134–155.

Albareda-Sambola, M., Fernández, E., and Saldanha-da Gama, F. (2011). The facility location problem with bernoulli demands. *Omega*, 39(3):335–345.

Alexandris, G. and Giannikos, I. (2010). A new model for maximal coverage exploiting gis capabilities. *European Journal of Operational Research*, 202(2):328–338.

Ayodele, T., Ogunjuyigbe, A., Odigie, O., and Munda, J. (2018). A multi-criteria gis based model for wind farm site selection using interval type-2 fuzzy analytic hierarchy process: The case study of nigeria. *Applied Energy*, 228:1853–1869.

Balinski, M. L. (1965). Integer programming: methods, uses, computations. *Management science*, 12(3):253–313.

Barbati, M. and Bruno, G. (2017). Exploring similarities in discrete facility location models with equality measures. *Geographical Analysis*. Forthcoming.

Barbati, M. and Piccolo, C. (2016). Equality measures properties for location problems. *Optimization Letters*, 10(5):903–920.

Bard, J. F. (2004). Staff scheduling in high volume service facilities with downgrading. *IIE Transactions*, 36(10):985–997.

Bard, J. F., Binici, C., and deSilva, A. H. (2003). Staff scheduling at the united states postal service. *Computers & Operations Research*, 30(5):745–771.

Bard, J. F. and Jarrah, A. I. (2009). Large-scale constrained clustering for rationalizing pickup and delivery operations. *Transportation Research Part B: Methodological*, 43(5):542–561.

Bard, J. F. and Wan, L. (2008). Workforce design with movement restrictions between workstation groups. *Manufacturing & Service Operations Management*, 10(1):24–42.

Barrett, M., Davidson, E., Prabhu, J., and Vargo, S. L. (2015). Service innovation in the digital age: key contributions and future directions. *MIS Quarterly*, 39(1):135–154.

Beliën, J. and Forcé, H. (2012). Supply chain management of blood products: A literature review. *European Journal of Operational Research*, 217(1):1–16.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*, volume 28. Princeton University Press.

Bender, M., Kalcsics, J., Nickel, S., and Pouls, M. (2018). A branch-and-price algorithm for the scheduling of customer visits in the context of multi-period service territory design. *European Journal of Operational Research*, 269(1):382–396.

Bender, M., Meyer, A., Kalcsics, J., and Nickel, S. (2016). The multi-period service territory design problem–an introduction, a model and a heuristic approach. *Transportation Research Part E: Logistics and Transportation Review*, 96:135–157.

Benzarti, E., Sahin, E., and Dallery, Y. (2013). Operations management applied to home care services: Analysis of the districting problem. *Decision Support Systems*, 55(2):587–598.

Bergey, B. K., Ragsdale, C. T., and Hoskote, M. (2003). A simulated annealing genetic algorithm for the electrical power districting problem. *Annals of Operations Research*, 121(1-4):33–55.

Berman, O., Drezner, Z., and Krass, D. (2010). Generalized coverage: New developments in covering location models. *Computers & Operations Research*, 37(10):1675–1687.

Berman, O., Drezner, Z., and Krass, D. (2011). Discrete cooperative covering problems. *Journal of the Operational Research Society*, 62(11):2002–2012.

Berman, O., Drezner, Z., Tamir, A., and Wesolowsky, G. O. (2009). Optimal location with equitable loads. *Annals of Operations Research*, 167(1):307–325.

Bianchi, G., Bruni, R., Reale, A., and Sforzi, F. (2016). A min-cut approach to functional regionalization, with a case study of the italian local labour market areas. *Optimization Letters*, 10(5):955–973.

Bieniek, M. (2015). A note on the facility location problem with stochastic demands. *Omega*, 55:53–60.

Birge, J. R. and Louveaux, F. V. (2011). *Introduction to Stochastic Programming.* Springer, New York Dordrecht Heidelberg London. 2nd edition.

Blagojević, M., Šelmić, M., Macura, D., and Šarac, D. (2013). Determining the number of postal units in the network–fuzzy approach, serbia case study. *Expert Systems with Applications*, 40(10):4090–4095.

Blais, M., Lapierre, S. D., and Laporte, G. (2003). Solving a home-care districting problem in an urban setting. *Journal of the Operational Research Society*, 54(11):1141–1147.

Bosnes, V., Aldrin, M., and Heier, H. E. (2005). Predicting blood donor arrival. *Transfusion*, 45(2):162–170.

Boysen, N., Fedtke, S., and Weidinger, F. (2017). Truck scheduling in the postal service industry. *Transportation Science*, 51(2):723–736.

Bozkaya, B., Erkut, E., and Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, 144(1):12–26.

Bruno, G., Diglio, A., Melisi, A., and Piccolo, C. (2017a). A districting model to support the redesign process of Italian provinces. In Sforza, A. and Sterle, C., editors, *International Conference on Optimization and Decision Science*, pages 245–256. Springer, Cham.

Bruno, G., Esposito, E., Genovese, A., and Piccolo, C. (2016a). Institutions and facility mergers in the italian education system: models and case studies. *Socio-Economic Planning Sciences*, 53:23–32.

Bruno, G., Genovese, A., and Piccolo, C. (2016b). Capacity management in public service facility networks: a model, computational tests and a case study. *Optimization Letters*, 10(5):975–995.

Bruno, G., Genovese, A., and Piccolo, C. (2017b). Territorial amalgamation decisions in local government: Models and a case study from italy. *Socio-Economic Planning Sciences*, 57:61–72.

Bruno, G. and Giannikos, I. (2015). Location and gis. In *Location Science*, pages 509–536. Springer.

Bruno, G. and Laporte, G. (2002). An interactive decision support system for the design of rapid public transit networks. *INFOR: Information Systems and Operational Research*, 40(2):111–118.

Butch, S. H. (2002). Computerization in the transfusion service. *Vox sanguinis*, 83:105–110.

Calik, H., Labbé, M., and Yaman, H. (2015). p-center problems. In *Location Science*, pages 79–92. Springer.

Camacho-Collados, M. and Liberatore, F. (2015). A decision support system for predictive police patrolling. *Decision Support Systems*, 75:25–37.

Camacho-Collados, M., Liberatore, F., and Angulo, J. M. (2015). A multi-criteria police districting problem for the efficient and effective design of patrol sector. *European Journal of Operational Research*, 246(2):674–684.

Cardenas, I. D., Dewulf, W., Beckers, J., Smet, C., and Vanelslander, T. (2017). The e-commerce parcel delivery market and the implications of home B2C deliveries vs pick-up points. *International Journal of Transport Economics*, 44(2):235–256.

Carlsson, J. G. (2012). Dividing a territory among several vehicles. *INFORMS Journal on Computing*, 24(4):565–577.

Carlsson, J. G. and Delage, E. (2013). Robust partitioning for stochastic multivehicle routing. *Operations Research*, 61(3):727–744.

Caro, F., Shirabe, T., Guignard, M., and Weintraub, A. (2004). School redistricting: Embedding gis tools with integer programming. *Journal of the Operational Research Society*, 55(8):836–849.

Carreras, M. and Serra, D. (1999). On optimal location with threshold requirements. *Socio-Economic Planning Sciences*, 33(2):91–103.

Cetin, E. and Sarul, L. S. (2009). A blood bank location model: A multiobjective approach. *European Journal of Pure and Applied Mathematics*, 2(1):112–124.

Church, R. and ReVelle, C. (1974). The maximal covering location problem. In *Papers of the Regional Science Association*, volume 32, pages 101–118. Springer.

Cooper, L. (1963). Location-allocation problems. *Operations research*, 11(3):331–343.

Correia, I. and Saldanha-da-Gama, F. (2015). Facility location under uncertainty. In Laporte, G., Nickel, S., and Saldanha-da-Gama, F., editors, *Location Science*, chapter 5, pages 177–205. Springer.

Custer, B., Johnson, E. S., Sullivan, S. D., Hazlet, T. K., Ramsey, S. D., Murphy, E. L., and Busch, M. P. (2005). Community blood supply model: development of a new model to assess the safety, sufficiency, and cost of the blood supply. *Medical decision making*, 25(5):571–582.

D'Amico, S. J., Wang, S. J., Batta, R., and Rump, C. M. (2002). A simulated annealing approach to police district design. *Computers & Operations Research*, 29(6):667–684.

Daskin, M. S. and Dean, L. K. (2005). Location of health care facilities. In *Operations research and health care*, pages 43–76. Springer.

Daskin, M. S. and Maass, K. L. (2015). The p-median problem. In *Location science*, pages 21–45. Springer.

Davis, R., Geiger, B., Gutierrez, A., Heaser, J., and Veeramani, D. (2009). Tracking blood products in blood centres using radio frequency identification: a comprehensive assessment. *Vox sanguinis*, 97(1):50–60.

De Assis, L. S., Franca, P. M., and Usberti, F. L. (2014). A redistricting problem applied to meter reading in power distribution networks. *Computers & Operations Research*, 41:65–75.

De Fréminville, P. D. L. P., Desaulniers, G., Rousseau, L. M., and Perron, S. (2015). A column generation heuristic for districting the price of a financial product. *Journal of the Operational Research Society*, 66(6):965–978.

Delen, D., Erraguntla, M., Mayer, R. J., and Wu, C.-N. (2011). Better management of blood supply-chain with gis-based analytics. *Annals of Operations Research*, 185(1):181–193.

Difrancesco, R. M., Huchzermeier, A., and Schröder, D. (2017). Optimizing the return window for online fashion retailers with closed-loop refurbishment. *Omega*, 78:205–221.

Ding, Y., Gao, X., Huang, C., Shu, J., and Yang, D. (2017). Service competition in an online duopoly market. *Omega*, 77:58–72.

Drezner, T. and Eiselt, H. A. (2002). Consumers in competitive location models. *Facility location: Applications and theory*, pages 151–178.

Eiselt, H. A. and Laporte, G. (1995). Objectives in location problems. In Drezner, Z., editor, *Facility location a survey of application and method*. Springer-Verlag, New York, USA.

Eiselt, H. A. and Marianov, V. (2011). *Foundations of location analysis*, volume 155. Springer Science & Business Media.

Elalouf, A., Hovav, S., Tsadikovich, D., and Yedidsion, L. (2015). Minimizing operational costs by restructuring the blood sample collection chain. *Operations Research for Health Care*, 7:81–93.

Erbaş, M., Kabak, M., Özceylan, E., and Çetinkaya, C. (2018). Optimal siting of electric vehicle charging stations: A gis-based fuzzy multi-criteria decision analysis. *Energy*, 163:1017–1031.

Erdoğan, G. (2017). An open source spreadsheet solver for vehicle routing problems. *Computers & Operations Research*, 84:62–72.

Erkut, E. (1990). The discrete p-dispersion problem. *European Journal of Operational Research*, 46(1):48–60.

Erkut, E., Karagiannidis, A., Perkoulidis, G., and Tjandra, S. A. (2008). A multicriteria facility location model for municipal solid waste management in north greece. *European Journal of Operational Research*, 187(3):1402–1421.

ESPAS (2015). Global Trends to 2030: Can the EU meet the challenges ahead? URL: www.ec.europa.eu [accessed in September 2018].

European Directorate for the Quality of Medicines (2015). Guide to the preparation, use and quality assurance of blood components. URL: www.edqm.eu [accessed on June 10, 2018].

European Parliament and Council of European Union (1998). Directive 97/67/EC of the European Parliament and of the Council of 15 December 1997 on common rules for the development of the internal market of Community postal services and the improvement of quality of service. URL: www.eur-lex.europa.eu [accessed on June 10, 2018].

European Parliament and Council of European Union (2002). Directive 2002/39/EC of the European Parliament and of the Council of 10 June 2002 amending Directive 97/67/ec with regard to the further opening to competition of Community postal services. URL: www.eur-lex.europa.eu [accessed on June 10, 2018].

European Parliament and Council of European Union (2003). Setting standards of quality and safety for the collection, testing, processing, storage and distribution of human blood and blood componentsand amending Directive 2001/83/EC. URL: www.eur-lex.europa.eu [accessed on June 10, 2018].

European Parliament and Council of European Union (2004). Certain technical requirement for blood and blood components. URL: www.eur-lex.europa.eu [accessed on June 10, 2018].

European Parliament and Council of European Union (2005a). Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards Community standards and specifications relating to a quality system for blood establishments. URL: www.eur-lex.europa.eu [accessed on June 10, 2018].

European Parliament and Council of European Union (2005b). Implementing Directive 2002/98/EC of the European Parliament and of the Council as regards traceability requirements and notification of serious adverse reactions and events. URL: www.eur-lex.europa.eu [accessed on June 10, 2018].

European Parliament and Council of European Union (2008). Directive 2008/6/ec of the European Parliament and of the Council of 20 February 2008 amending Directive 97/67/ec with regard to the full accomplishment of the internal market of Community postal services. URL: www.eur-lex.europa.eu [accessed on June 10, 2018].

Fahimnia, B., Jabbarzadeh, A., Ghavamifar, A., and Bell, M. (2017). Supply chain design for efficient and effective blood supply in disasters. *International Journal of Production Economics*, 183:700–709.

Farahani, R. Z., Asgari, N., Heidari, N., Hosseininia, M., and Goh, M. (2012). Covering problems in facility location: A review. *Computers & Industrial Engineering*, 62(1):368–407.

Farahani, R. Z., SteadieSeifi, M., and Asgari, N. (2010). Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*, 34(7):1689–1709.

Ferland, J. and Guénette, G. (1990). Decision support system for the school districting problem. *Operations Research*, 38(1):15–21.

Fernández, E. and Landete, M. (2015). Fixed-charge facility location problems. In *Location science*, pages 47–77. Springer.

Franceschetti, A., Jabali, O., and Laporte, G. (2017). Continuous approximation models in freight distribution management. *TOP*, 25(3):413–433.

Francis, R. L. and Lowe, T. J. (2015). Demand point aggregation for some basic location models. In Laporte, G., Nickel, S., and Saldanha da Gama, F., editors, *Location Science*, pages 487–506. Berlin: Springer.

Galvão, L. C., Novaes, A. G., De Cursi, J. S., and Souza, J. C. (2006). A multiplicatively-weighted Voronoi diagram approach to logistics districting. *Computers & Operations Research*, 33(1):93–114.

García, S., Labbé, M., and Marín, A. (2011). Solving large p-median problems with a radius formulation. *INFORMS Journal on Computing*, 23(4):546–556.

García, S. and Marín, A. (2015). Covering location problems. In Laporte, G., Nickel, S., and Saldanha da Gama, F., editors, *Location Science*, pages 93–114. Berlin: Springer.

García-Ayala, G., González-Velarde, J. L., Ríos-Mercado, R. Z., and Fernández, E. (2016). A novel model for arc territory design: promoting eulerian districts. *International Transactions in Operational Research*, 23(3):433–458.

Giannikos, I. (1998). A multiobjective programming model for locating treatment sites and routing hazardous wastes. *European Journal of Operational Research*, 104(2):333–342.

Glynn, S. A., Kleinman, S. H., Schreiber, G. B., Zuck, T., Mc Combs, S., Bethel, J., Garratty, G., Williams, A. E., and Study, R. E. D. (2002). Motivations to donate blood: demographic comparisons. *Transfusion*, 42(2):216–225.

Godin, G., Conner, M., Sheeran, P., Bélanger-Gravel, A., and Germain, M. (2007). Determinants of repeated blood donation among new and experienced blood donors. *Transfusion*, 47(9):1607–1615.

Grünert, T. and Sebastian, H.-J. (2000). Planning models for long-haul operations of postal and express shipment companies. *European Journal of Operational Research*, 122(2):289–309.

Güneş, E. D. and Nickel, S. (2015). Location problems in healthcare. In *Location science*, pages 555–579. Springer.

Gutiérrez, J., Condeço-Melhorado, A., and Martín, J. C. (2010). Using accessibility indicators and gis to assess spatial spillovers of transport infrastructure investment. *Journal of transport geography*, 18(1):141–152.

Gwak, J. H., Lee, B. K., Lee, W. K., and Sohn, S. Y. (2017). Optimal location selection for the installation of urban green roofs considering honeybee habitats along with socio-economic and environmental effects. *Journal of environmental management*, 189:125–133.

Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations research*, 12(3):450–459.

Hakimi, S. L. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations research*, 13(3):462–475.

Haugland, D., Ho, S. C., and Laporte, G. (2007). Designing delivery districts for the vehicle routing problem with stochastic demands. *European Journal of Operational Research*, 180(3):997–1010.

Hemmelmayr, V., Doerner, K. F., Hartl, R. F., and Savelsbergh, M. W. (2009). Delivery strategies for blood products supplies. *OR spectrum*, 31(4):707–725.

Hemmelmayr, V., Doerner, K. F., Hartl, R. F., and Savelsbergh, M. W. (2010). Vendor managed inventory for environments with stochastic product usage. *European Journal of Operational Research*, 202(3):686–695.

Hernandez, P., Alonso-Ayuso, A., Bravo, F., Escudero, L. F., Guignard, M., Marianov, V., and Weintraub, A. (2012). A branch-and-cluster coordination scheme for selecting prison facility sites under uncertainty. *Computers & Operations Research*, 39(9):2232–2241.

Hess, S. W., Weaver, J. B., Siegfeldt, H. J., Whelan, J. N., and Zitlau, P. A. (1965). Nonpartisan political redistricting by computer. *Operations Research*, 13(6):998–1006.

Higgs, G. (2006). Integrating multi-criteria techniques with geographical information systems in waste facility location to enhance public participation. *Waste management & research*, 24(2):105–117.

Higgs, G. and Langford, M. (2013). Investigating the validity of rural–urban distinctions in the impacts of changing service provision: The example of postal service reconfiguration in wales. *Geoforum*, 47:53–64.

Hinojosa, Y., Puerto, J., and Saldanha-da Gama, F. (2014). A two-stage stochastic transportation problem with fixed handling costs and a priori selection of the distribution channels. *Top*, 22(3):1123–1147.

Holmes, J., Williams, F. B., and Brown, L. A. (1972). Facility location under a maximum travel restriction: An example using day care facilities. *Geographical Analysis*, 4(3):258–266.

Hong, S.-H. and Wolak, F. A. (2008). Relative prices and electronic substitution: Changes in household-level demand for postal delivery services from 1986 to 2004. *Journal of Econometrics*, 145(1-2):226–242.

Irnich, S. (2008). Solution of real-world postman problems. *European Journal of Operational Research*, 190(1):52–67.

ISTAT (2011). Dati del censimento generale della popolazione italiana (italian national census data). URL: www.istat.it [accessed in September 2018].

Italian Ministry of Health (2016). Il sistema trasfusionale italiano in cifre. URL: www.salute.gov.it [accessed on June 10, 2018].

Italian State-Region Conference (2012). Guidelines for the accreditation of blood transfusion services and units blood collection and blood components. URL: www.gazzettaufficiale.it [accessed on June 10, 2018].

Jabbarzadeh, A., Fahimnia, B., and Seuring, S. (2014). Dynamic supply chain network design for the supply of blood in disasters: a robust model with real world application. *Transportation Research Part E: Logistics and Transportation Review*, 70:225–244.

Jarrah, A. I. and Bard, J. F. (2012). Large-scale pickup and delivery work area design. *Computers & Operations Research*, 39(12):3102–3118.

Jarrah, A. I., Bard, J. F., and deSilva, A. H. (1992). A heuristic for machine scheduling at general mail facilities. *European Journal of Operational Research*, 63(2):192–206.

Jarrah, A. I., Bard, J. F., and deSilva, A. H. (1994a). Equipment selection and machine scheduling in general mail facilities. *Management Science*, 40(8):1049–1068.

Jarrah, A. I., Bard, J. F., and deSilva, A. H. (1994b). Solving large-scale tour scheduling problems. *Management Science*, 40(9):1124–1144.

Jarrah, A. I., Qi, X., and Bard, J. F. (2014). The destination-loader-door assignment problem for automated package sorting centers. *Transportation Science*, 50(4):1314–1336.

Jung, H., Lee, K., and Chun, W. (2006). Integration of gis, gps, and optimization technologies for the effective control of parcel delivery service. *Computers & Industrial Engineering*, 51(1):154–162.

Kalcsics, J. (2015). Districting problems. In Laporte, G., Nickel, S., and Saldanha da Gama, F., editors, *Location Science*, pages 595–622. Berlin: Springer.

Kalcsics, J., Nickel, S., and Schröder, M. (2005). Towards a unified territorial design approach—applications, algorithms and GIS integration. *TOP*, 13:1–56.

Kariv, O. and Hakimi, S. L. (1979). An algorithmic approach to network location problems. i: The p-centers. *SIAM Journal on Applied Mathematics*, 37(3):513–538.

Katsaliaki, K. and Brailsford, S. C. (2007). Using simulation to improve the blood supply chain. *Journal of the Operational Research Society*, 58(2):219–227.

Khan, M. M.-U.-H., Vaezi, M., and Kumar, A. (2018). Optimal siting of solid waste-to-value-added facilities through a gis-based assessment. *Science of The Total Environment*, 610:1065–1075.

Kim, K., Chun, Y., and Kim, H. (2017). p-functional clusters location problem for detecting spatial clusters with covering approach. *Geographical Analysis*, 49(1):101–121.

Kınay, Ö. B., Kara, B. Y., Saldanha-da Gama, F., and Correia, I. (2018a). Modeling the shelter site location problem using chance constraints: A case study for istanbul. *European Journal of Operational Research*, 270(1):132–145.

Kınay, Ö. B., Saldanha-da Gama, F., and Kara, B. Y. (2018b). On multi-criteria chance-constrained capacitated single-source discrete facility location problems. *Omega*.

Labbé, M. and Laporte, G. (1986). Maximizing user convenience and postal service efficiency in post box location. *Belgian Journal of Operations Research, Statistics and Computer Science*, 26:21–35.

Laporte, G., Nickel, S., and da Gama, F. S. (2015). *Location science*, volume 528. Springer.

Lei, H., Laporte, G., and Guo, B. (2012). Districting for routing with stochastic customers. *EURO Journal on Transportation and Logistics*, 1(1-2):67–85.

Lei, H., Laporte, G., Liu, Y., and Zhang, T. (2015). Dynamic design of sales territories. *Computers & Operations Research*, 56:84–92.

Lei, H., Wang, R., and Laporte, G. (2016). Solving a multi-objective dynamic stochastic districting and routing problem with a co-evolutionary algorithm. *Computers & Operations Research*, 67:12–24.

Li, B. N., Dong, M. C., and Chao, S. (2008). On decision making support in blood bank information systems. *Expert Systems with Applications*, 34(2):1522–1532.

Lin, C. (2009). Stochastic single-source capacitated facility location model with service level requirements. *International Journal of Production Economics*, 117(2):439–451.

Lin, M., Chin, K. S., Fu, C., and Tsui, K. L. (2017). An effective greedy method for the meals-on-wheels service districting problem. *Computers & Industrial Engineering*, 106:1–19.

Louveaux, F. V. (1986). Discrete stochastic location models. *Annals of Operations research*, 6(2):21–34.

Malczewski, J. (2004). Gis-based land-use suitability analysis: a critical overview. *Progress in planning*, 62(1):3–65.

Marianov, V. and Serra, D. (2001). Hierarchical location–allocation models for congested systems. *European Journal of Operational Research*, 135(1):195–208.

Marianov, V., Serra, D., et al. (2002). Location problems in the public sector. *Facility location: Applications and theory*, 1:119–150.

Marsh, M. T. and Schilling, D. A. (1994). Equity measurement in facility location analysis: A review and framework. *European Journal of Operational Research*, 74(1):1–17.

McWilliams, D. L. and McBride, M. E. (2012). A beam search heuristics to solve the parcel hub scheduling problem. *Computers & Industrial Engineering*, 62(4):1080–1092.

Melo, M. T., Nickel, S., and Saldanha-Da-Gama, F. (2009). Facility location and supply chain management–a review. *European journal of operational research*, 196(2):401–412.

Mitrović-Minić, S., Krishnamurti, R., and Laporte, G. (2004). Double-horizon based heuristics for the dynamic pickup and delivery problem with time windows. *Transportation Research Part B: Methodological*, 38(8):669–685.

Morganti, E., Seidel, S., Blanquart, C., Dablanc, L., and Lenz, B. (2014). The impact of e-commerce on final deliveries: Alternative parcel delivery services in France and Germany. *Transportation Research Procedia*, 4:178–190.

Mourão, M. C., Nunes, A. C., and Prins, C. (2009). Heuristic methods for the sectoring arc routing problem. *European Journal of Operational Research*, 196(3):856–868.

Murray, A. T. (2010). Advances in location modeling: Gis linkages and contributions. *Journal of geographical systems*, 12(3):335–354.

Nagurney, A., Masoumi, A. H., and Yu, M. (2012). Supply chain network operations management of a blood banking system with cost and risk minimization. *Computational Management Science*, 9(2):205–231.

Niblick, B., Monnell, J. D., Zhao, X., and Landis, A. E. (2013). Using geographic information systems to assess potential biofuel crop production on urban marginal lands. *Applied Energy*, 103:234–242.

Nickel, S., Saldanha-da Gama, F., and Ziegler, H.-P. (2012). A multi-stage stochastic supply network design problem with financial decisions and risk management. *Omega*, 40(5):511–524.

Nikali, H. (2008). Substitution of letter mail for different sender-receiver segments. In Crew, M. A. and Kleindorfer, P. R., editors, *Competition and Regulation in the Postal and Delivery Sector*, pages 89–106. Cheltenham: Edward Elgar Publishing.

Novaes, A. G., de Cursi, J. E. S., and Graciolli, O. D. (2000). A continuous approach to the design of physical distribution systems. *Computers & Operations Research*, 27(9):877–893.

Novaes, A. G. and Graciolli, O. D. (1999). Designing multi-vehicle delivery tours in a grid-cell format. *European Journal of Operational Research*, 119(3):613–634.

Ortiz-Astorquiza, C., Contrerasn, I., and Laporte, G. (2017). Multi-level facility location problems. *European Journal of Operational Research*.

Osorio, A. F., Brailsford, S. C., and Smith, H. K. (2015). A structured review of quantitative models in the blood supply chain: a taxonomic framework for decision-making. *International Journal of Production Research*, 53(24):7191–7212.

Özsoy, F. A. and Pınar, M. Ç. (2006). An exact algorithm for the capacitated vertex p-center problem. *Computers & Operations Research*, 33(5):1420–1436.

Phillips, P. A. and Wright, C. (2009). E-business's impact on organizational flexibility. *Journal of Business Research*, 62(11):1071–1080.

Pierskalla, W. P. (2005). Supply chain management of blood banks. In *Operations research and health care*, pages 103–145. Springer.

Pirkul, H. and Jayaraman, V. (1998). A multi-commodity, multi-plant, capacitated facility location problem: formulation and efficient heuristic solution. *Computers & Operations Research*, 25(10):869–878.

Prastacos, G. P. (1984). Blood inventory management: an overview of theory and practice. *Management Science*, 30(7):777–800.

Qu, Y. and Bard, J. F. (2012). A grasp with adaptive large neighborhood search for pickup and delivery problems with transshipment. *Computers & Operations Research*, 39(10):2439–2456.

ReVelle, C., Murray, A. T., and Serra, D. (2007). Location models for ceding market share and shrinking services. *Omega*, 35(5):533–540.

ReVelle, C. S. and Eiselt, H. A. (2005). Location analysis: A synthesis and survey. *European journal of operational research*, 165(1):1–19.

Revelle, C. S., Eiselt, H. A., and Daskin, M. S. (2008). A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, 184(3):817–848.

Ricca, F., Scozzari, A., and Simeone, B. (2013). Political districting: from classical models to recent approaches. *Annals of Operations Research*, 204(1):271–299.

Ricca, F. and Simeone, B. (2008). Local search algorithms for political districting. *European Journal of Operational Research*, 189(3):1049–1426.

Ríos-Mercado, R. Z. (2016). Assessing a metaheuristic for large-scale commercial districting. *Cybernetics and Systems*, 47(4):321–338.

Ríos-Mercado, R. Z. and Escalante, H. J. (2016). GRASP with path relinking for commercial districting. *Expert Systems with Applications*, 44:102–113.

Ríos-Mercado, R. Z. and López-Pérez, J. F. (2013). Commercial territory design planning with realignment and disjoint assignment requirements. *Omega*, 41(3):525–535.

Rosenfield, D. B., Engelstein, I., and Feigenbaum, D. (1992). An application of sizing service territories. *European Journal of Operational Research*, 63(2):164–172.

Şahin, G. and Süral, H. (2007). A review of hierarchical facility location models. *Computers & Operations Research*, 34(8):2310–2331.

Şahin, G., Süral, H., and Meral, S. (2007). Locational analysis for regionalization of turkish red crescent blood services. *Computers & Operations Research*, 34(3):692–704.

Santini, V., Truschi, F., Bertelli, A., and Lazzaro, C. (2013). Cost of red blood cell transfusion: an activity-based cost description. *Drugs and Cell Therapies in Hematology*, 2(2):157–167.

Schoepfle, O. and Church, R. (1991). A new network representation of a "classic" school districting problem. *Socio-Economic Planning Sciences*, 25(3):189–197.

Schreiber, G. B., Schlumpf, K. S., Glynn, S. A., Wright, D. J., Tu, Y., King, M. R., Higgins, M. J., Kessler, D., Gilcher, R., Nass, C. C., et al. (2006). Convenience, the bane of our existence, and other barriers to donating. *Transfusion*, 46(4):545–553.

Serra, D. (1996). The coherent covering location problem. *Papers in Regional Science*, 75(1):79–101.

Sindhu, S., Nehra, V., and Luthra, S. (2017). Investigation of feasibility study of solar farms deployment using hybrid ahp-topsis analysis: Case study of india. *Renewable and Sustainable Energy Reviews*, 73:496–511.

Sonmez, A. D. and Lim, G. J. (2012). A decomposition approach for facility location and relocation problem with uncertain number of future facilities. *European Journal of Operational Research*, 218(2):327–338.

Talen, E. and Anselin, L. (1998). Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. *Environment and planning A*, 30(4):595–613.

Tavares, P., Figueira, F., Mousseau, J., and Roy, V. (2007). Multiple criteria districting problems. *The public transportation network pricing system of the Paris region. Annals of Operations Research*, 154:69–9.

Telser, L. G. (1978). Economic theory and the core. Technical report.

Telser, L. G. (2005). *A theory of efficient cooperation and competition*. Cambridge University Press.

Terh, S. H. and Cao, K. (2018). Gis-mcda based cycling paths planning: a case study in singapore. *Applied Geography*, 94:107–118.

Toregas, C., Swain, R., ReVelle, C., and Bergman, L. (1971). The location of emergency service facilities. *Operations research*, 19(6):1363–1373.

Universal Postal Union (2015). Development of postal services in 2015. URL: www.upu.int [accessed on June 10, 2018].

Van Boening, M. V. and Wilcox, N. T. (1996). Avoidable cost: Ride a double auction roller coaster. *The American Economic Review*, pages 461–477.

Wang, Q., Batta, R., Bhadury, J., and Rump, C. M. (2003). Budget constrained location problem with opening and closing of facilities. *Computers & Operations Research*, 30(13):2047–2069.

Wang, Q., Batta, R., and Szczerba, R. J. (2005). Sequencing the processing of incoming mail to match an outbound truck delivery schedule. *Computers & Operations Research*, 32(7):1777–1791.

Weber, A. (1929). *Theory of the Location of Industries*. University of Chicago Press.

WHO Expert Group (2012). Expert consensus statement on achieving self-sufficiency in safe blood and blood products, based on voluntary non-remunerated blood donation (vnrbd). *Vox sanguinis*, 103(4):337–342.

Williamson, L. M. and Devine, D. V. (2013). Challenges in the management of the blood supply. *the Lancet*, 381(9880):1866–1875.

Wu, H., Wang, J., Duan, H., Ouyang, L., Huang, W., and Zuo, J. (2016). An innovative approach to managing demolition waste via gis (geographic information system): a case study in shenzhen city, china. *Journal of Cleaner Production*, 112:494–503.

Xie, S. and Ouyang, Y. (2016). Railroad caller districting with reliability, contiguity, balance, and compactness considerations. *Transportation Research Part C: Emerging Technologies*, 63:65–76.

Yanık, S., Sürer, Ö., and Öztayşi, B. (2016). Designing sustainable energy regions using genetic algorithms and location-allocation approach. *Energy*, 97:161–172.

Zahiri, B., Torabi, S., Mousazadeh, M., and Mansouri, S. (2015). Blood collection management: Methodology and application. *Applied Mathematical Modelling*, 39(23-24):7680–7696.

Zenker, M. and Boysen, N. (2018). Dock sharing in cross-docking facilities of the postal service industry. *Journal of the Operational Research Society*, 69:1–16.

Zhang, J., Wang, X., and Huang, K. (2017). On-line scheduling of order picking and delivery with multiple zones and limited vehicle capacity. *Omega*, 79:104–115.

Zhang, X. and Bard, J. F. (2005). Equipment scheduling at mail processing and distribution centers. *IIE Transactions*, 37(2):175–187.

Zhang, X., Chakravarthy, A., and Gu, Q. (2009). Equipment scheduling problem under disruptions in mail processing and distribution centres. *Journal of the Operational Research Society*, 60(5):598–610.

Zhong, H., Hall, R. W., and Dessouky, M. (2007). Territory planning and vehicle dispatching with driver learning. *Transportation Science*, 41(1):74–89.

Zoltners, A. A. and Sinha, P. (2005). Sales territory design: thirty years of modeling and implementation. *Marketing Science*, 24(3):313–331.