Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in
Ingegneria Informatica

Mathematical Programming for the Optimisation of Ambulances Deployment

Relatore: **Prof. Matteo Fischetti**

Laureando: **Andrea Bugin**

Anno Accademico 2019-2020
Data di Laurea 06-04-2020

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Corso di Laurea Magistrale in Ingegneria Informatica

MATHEMATICAL PROGRAMMING FOR THE OPTIMISATION OF AMBULANCES DEPLOYMENT

Laureando
ANDREA BUGIN

Relatore
CH.MO PROF.
MATTEO FISCHETTI

ANNO ACCADEMICO 2019/2020

Abstract

Emergency medical services (EMS), also known as ambulance services or paramedic services, are emergency services which treat illnesses and injuries that require an urgent medical response, providing out-of-hospital treatment and transport to definitive care. An example of the EMS can be seen when an incident or an injury occurs, which starts from the call for help and ends when the ambulance arrives at the hospital. But EMS are not only a ride to the hospital, because behind that, there is a complex system involving multiple people and agencies. A comprehensive EMS system is ready every day for every kind of emergency. It's easy to understand that improving these services is very important for the community.

In this thesis we address the problem of the optimization of ambulance supplying for the ULSS of Veneto Region through a mixed integer linear programming model and we solved it using an optimization software. We analyse the state of the art, we describe the model and all the important parameters; then we compare the current results obtained with an alternative simple formula, with those obtained integrating the formula in the model itself. Finally, we use a different approach based on the number of accesses in the emergency rooms. Tests, carried out focusing basically on the ULSS 8 "Berica", show how an optimisation model can help reducing the number of ambulances under specific assumptions.

Acknowledgments

First of all I would like to thank my mum and my dad, because the only reason I reached this goal is because of you. I owe you everything.
A special thanks to my girlfriend Maddalena, you are the person that everyone would like to have by his side, love you!
To my relatives, that are always present in case of need.
To my friends, Vitto, Nicho ($\times 2$), Fabio, Matteo, Richy, with whom I spent 25 years laughing, joking and playing soccer; Ilie, for all the projects done together and all the lines of code that I had to debug because of you; to all the other friends that I can't mention for reasons of space with which I shared at least one piece of this journey.
Finally I want to thank my supervisor Prof. Matteo Fischetti, a superb teacher and an even better person who gave me the opportunity to work in this thesis, who has always made himself available to give me useful advice and support me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Emergency medical services (EMS), also known as ambulance services, are emergency services which treat illnesses and injuries that require an urgent medical response, providing out-of-hospital treatment and transport to definitive care [33]. These services are part of each community and play a very important role for the safety and care of people in the event of an accident or illness. They have the goal of intervening quickly and taking care of people in order to reduce mortality, disability, and suffering.

Hence EMS are a complex system, which do not include only healthcare and medical activities, but include also prevention and public education. A summary of the main components of EMS can be found in [1]:

- agencies and organizations (both private and public),
- communications and transportation networks,
- trauma systems, hospitals, trauma centers, and specialty care centers,
- rehabilitation facilities,
- highly trained professionals,
- volunteer and career prehospital personnel,
- physicians, nurses, and therapists,
- administrators and government officials,
- an informed public that knows what to do in a medical emergency.

These concepts are briefly summarized in the Figure 1.1 presented as a cycle.

For these reasons improving an aspect of the EMS could mean improving the quality of life of a community. In this work the main goal is roughly to find the

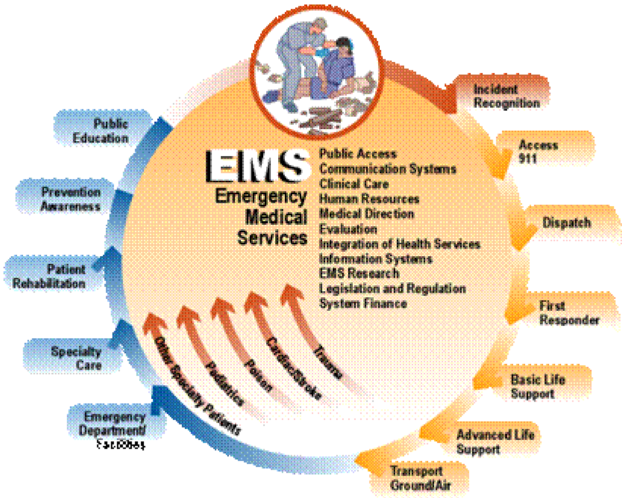


Figure 1.1: A representation of the components of the EMS [1].

minimum number of ambulances needed to provide an adequate service in a given region, but depending on country, area within country, or clinical need, emergency medical services may be provided by one or more different types of organization.

In this work we will focus basically in the Italy EMS and in particular in Veneto. EMS in Italy currently consist primarily of a combination of volunteer organizations providing ambulance service, supplemented by physicians and nurses who perform all Advanced Life Support (ALS) procedures. They are under Public Health Authorities control in each Italian region and the ambulance subsystem is provided by a variety of different sources [34].

The requirements and the constraints that define the problem we will address can be summarised more formally in the following way. As already stated, the problem is the minimisation of the number of ambulances needed in a given region and after careful analysis of the environment we had to study, we deduced that the main requirements are:

1. the solution must satisfied the greatest number of requests to reduce the probability of death,

2. the requests must be satisfied as fast as possible,

3. in any case the maximum response time is eight minutes for urban areas and twenty minutes for extra-urban areas; this response time is defined in the Press release relating to the DPR, March 27, 1992 [17],

4. the ambulances locations are fixed and they are the hospitals,

5. each ambulance must end its ride in the hospital prepared as a base of arrival.

Later on, we will discuss in depth these requirements and their mathematical formulation in the model, focusing also in the legislative part that regulates the system of health services in Italy and Veneto, analyzing in our case which laws and constraints must be respected.

In the next chapters we are going to present all the work done during these months, which includes mathematical formulations, implementation and testing phases. In particular, this report is structured as follows:

- in Chapter 2 we will present the main models developed in the literature for this problem and some variants, showing the pros and cons of each one,
- in Chapter 3 we will describe the Veneto healthcare system in detail and the differences with the literature and then we will present the model developed specifying also the choice of the parameters,
- in Chapter 4 we will show how we apply the model with a detailed example; then we will see a case of applicability, starting with the state of the art, and then using the model with different approaches,
- in Chapter 5 we report all the conclusions of our work and the possible future directions.

All the source code developed to build and solve the model is available at

https://github.com/AndreaB2604/MasterThesis[1]

[1]For more information to compile and run the code, see the README.md.

Chapter 2

Related Works

The problem, as described in the introduction, can have a considerable impact for what concern the security of people, but also for the budget of a Region or State, and it has been studied for over 50 years. For these reasons, one may think that the literature produced is almost complete: the models developed are stable and there are a lot of examples of applicability based on those models.

The reality is a little bit different, because there are surely a lot of examples, but the problem is changeable because it depends on various factors, especially when we want to apply one model in real cases. A practical example is the difference among countries: one for all the laws, that affect the constraints, the development and the resolution of a model. Another aspect to consider is the growth not only in computer technology, but also in modeling and algorithmic sophistication and in the performance of mathematical programming solvers: in this sense something that twenty years ago was not even tested because impossible to compute, today it could be solved in few seconds.

This latter aspect is crucial because, as we will see, the problem can be formulated with a Mixed Integer Linear Programming Model. The growth in computer technology led to a different approach when dealing with such problems and this can be seen reading the literature concerning a problem studied over decades, as the problem we are presenting. The first papers are focused mostly in the model formulation and little space is given to the implementation and testing phase, since finding the optimal solution was impossible even with small instances.

However, it is possible to identify some useful works in order to start approaching the problem and in this chapter we are going to present these works starting from the very beginning.

2.1 The evolution of the models

As already said the main goal is, in a non rigorous way, to minimise the number of ambulances maintaining an adequate level of service. This minimisation problem can be seen similarly as a maximisation problem with this formulation: maximise the level of service with a given number of ambulances; this clarification is needed because in the literature these two variants are both present, what changes is the focus of the objective function.

In the last years, three classes of models have been developed to solve the problem:

Deterministic (or static) models: they are used at the planning stage and ignore stochasticity and the availability of ambulances.

Probabilistic models: they consider the fact that ambulances operate as servers in a queuing system and cannot always answer a call.

Dynamic models: they were recently developed in order to relocate ambulances during the day to provide a better service.

In the next sections we will see the main models of all the three classes, focusing mostly on the deterministic models.

For the first class we will refer to [5, 28], where the authors formulate the first two models related to the problem and they were the basis from which literature developed. Then the next contribution comes from [26], where the authors developed a model to deal with multiple vehicles. Other notable improvements come from Daskin and Stern [7] and Hogan and ReVelle [12], where they introduced the concept of backup coverage.

For the second class, one of the first probabilistic models is the Maximum Expected Covering Location Problem formulation (MEXCLP) due to Daskin [6], where a probability of an ambulance to be unavailable was introduced. Another probabilistic model was proposed by ReVelle and Hogan in [25] in two variants and finally we will see the model developed by Ball and Lin in [2] called Rel-P.

For the third class we will see briefly the formulation of one model: the Dynamic Double Standard Model (DDSM) by Gendreau et al. [11]. In the literature other recent models were developed, for instance the Dynamic Available Coverage Location model (DACL) by Rajagopalan et al. [18], but sophisticated elements were used and, for most of them, the formulation is non linear; moreover the mathematical meaning of those elements is an advanced topic and it is beyond the scope of this work.

2.2 Deterministic models

Before starting the analysis of the literature mentioned in Section 2.1, we have to provide the main elements and the notation that will be used. The natural way to model the ambulance location problems is to consider a graph $G = (V \cup W, E)$ where the set of nodes of the graph is divided in two subset: the set of demand points, denoted with V, and the set of the potential ambulance location sites, which are commonly called service points (e.g., hospitals), denoted with W. Finally, the set of edges is $E \subseteq V \times W$. Note that, in general, $V \cap W \neq \emptyset$.

Another important element is the (shortest) travel time from vertex $i \in V$ to vertex $j \in W$, commonly denoted as t_{ij}. Now we can give the following definition:

Definition 2.2.1 (Covered point)**.** A demand point $i \in V$ is said to be *covered* by a site $j \in W$ if

$$t_{ij} \leq r_i$$

where r_i is a preset coverage standard related to node i. We can define the set of location sites covering demand point i as

$$W_i = \{j \in W : t_{ij} \leq r_i\}$$

The first deterministic model, formulated in [28], was called Location Set Covering Model (LSCM), which is based on the Set Covering Problem (SCP). Given a $m \times n$ matrix A with elements $a_{ij} \in \{0, 1\}$ and a vector $\boldsymbol{c} \in \mathbb{R}^n$, the Set Covering Model is defined:

$$\begin{aligned} \min \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & A\boldsymbol{x} \geq \mathbf{1} \\ & \mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1} \qquad \text{integer} \end{aligned}$$

where $\mathbf{0}$ and $\mathbf{1}$ represent a vector with all components equal to 0 and 1 respectively. An interpretation of the set covering model is the following. There are m objects $1, \ldots, m$ related to the rows of A and n subset of objects. Each subset is defined as

$$I_j = \big\{i \in \{1, \ldots, m\} : a_{ij} = 1\big\}$$

namely it contains (*cover*) the objects related to the row i with $a_{ij} = 1$, and it has a given cost c_j. The SCP required to choose a family of subsets I_j with minimum cost, in order to cover all the m objects at least once [9]. This is a well known NP-Complete problem [36].

Now we can present the LSCM. In this model the goal is to minimize the number of ambulances needed to provide a coverage of all demand points. In the original paper they set $r_i = r$, $\forall\, i \in V$. Let

$$x_j = \begin{cases} 1 & \text{if an ambulance is located in the vertex } j, \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\min \quad \sum_{j \in W} x_j \tag{2.2.1}$$

$$\text{s.t.} \quad \sum_{j \in W_i} x_j \geq 1 \qquad \forall\, i \in V \tag{2.2.2}$$

$$x_j \in \{0, 1\} \qquad \forall\, j \in W \tag{2.2.3}$$

The formulation reflects the set covering model where $\boldsymbol{c} = \mathbf{1}$ and defining $A = [a_{ij}]$ then $a_{ij} = 1$ if $t_{ij} \leq r$, 0 otherwise. The inequality (2.2.2) requires that there must be at least one ambulance which covers each demand point i.

This model is very simple and one may notice immediately that when an ambulance is dispatched in a demand point, other demand points related to that site can no longer be covered. Furthermore, with the constraint (2.2.3), the maximum number of ambulance for each site can be one and the situation would not change if the constraint was $x_j \in \mathbb{N}$, $\forall\, j \in W$, because it can be proved quite easily that if an optimal solution has $x_j \geq 2$ for some j, then also $x_j = 1$ is a feasible solution and it has a lower cost in the objective function, which proves that the initial solution with $x_j \geq 2$ was not optimal. Another assumption of the model is that up to $|W|$ ambulances are available, because if all the demand points are covered by only one site and each site covers at most one demand point, the solution requires exactly $|V| = |W|$ ambulances. Anyway, this model provides a lower bound to the number of ambulances needed to ensure full coverage.

As already said, an alternative approach is to maximise the level of service, namely the demand successfully served, with a limited number of ambulances. The first model to solve this problem was formulated in [5] and it is called Maximal Covering Location Problem (MCLP). Let x_j as defined in the LSCM and

$$d_i = \text{the demand of the node } i \in V$$

$$p = \text{maximum number of ambulances available}$$

$$y_i = \begin{cases} 1 & \text{if the vertex } i \text{ is covered by at least one ambulance} \\ 0 & \text{otherwise} \end{cases}$$

Then

$$
\begin{aligned}
\max \quad & \sum_{i \in V} d_i y_i & & \qquad (2.2.4)\\
\text{s.t.} \quad & \sum_{j \in W_i} x_j \geq y_i & \forall\, i \in V & \qquad (2.2.5)\\
& \sum_{j \in W} x_j = p & & \qquad (2.2.6)\\
& x_j \in \{0,1\} & \forall\, j \in W & \qquad (2.2.7)\\
& y_i \in \{0,1\} & \forall\, i \in V & \qquad (2.2.8)
\end{aligned}
$$

The constraints are basically the same as before with the difference of (2.2.5), where the right hand side indicates that if a node i is covered by one ambulance than there must be at least one ambulance located in a site j that covers i. The relation (2.2.6) is the constraint related to the number of ambulances that can be dispatched. The goal of this model is to use in the best way the limited resources available.

From a practical point of view the MCLP has been used to plan the reorganisation of the EMS in Austin, Texas, as described in [8]. The model was applied with a minor change, adding weights which multiply the objective function. In the paper they stated that the benefits of the EMS plan have been: the average response time had been reduced despite an increase in calls for service and from a financial point of view: \$3.4 million saved in construction costs and \$1.2 million in operating cost in 1984.

The MCLP model has been extended to deal with multiple vehicles. In the context of EMS, not only ambulances are used because there are also nontransporting EMS vehicles, which are vehicles that responds to emergencies, but are not designed to transport a patient [35]. In Italy there are three types of standard emergency vehicles in force in almost all regions [32]:

- Basic Life Support (BLS) vehicles: they can be summarised as ambulances owned by associations, social cooperatives or voluntary bodies, with only technical rescue personnel on board, certified and qualified for emergency services, with specific training course, and with "basic" health equipment on board;

- Intermediate Life Support (ILS) vehicles: ambulances that normally include a nurse trained and authorized to apply specific and advanced intervention;

- Advanced Life Support (ALS) vehicles: the crew has one or two rescuers, one of whom is a driver, a nurse and a doctor, often anesthesiologist-resuscitator, from the emergency room or directly from the operations center.

For these reasons Shilling et al. in [26] extended the MCLP, formulating the Tandem Equipment Allocation Model (TEAM) to manage several vehicle types. Let's suppose for example, we have two types of vehicles to dispatch: type A and B. In this case let p_A and p_B the number of available vehicles of type A and B; r_A and r_B the coverage standard for vehicles of type A and B. Let

$$W_{A,i} = \{j \in W : t_{ij} \leq r_A\}$$

$$W_{B,i} = \{j \in W : t_{ij} \leq r_B\}$$

$$x_{A,j} = \begin{cases} 1 & \text{if a vehicle of type A is located at node } j \\ 0 & \text{otherwise} \end{cases}$$

with $x_{B,i}$ similarly for vehicles of type B. Let

$$y_i = \begin{cases} 1 & \text{if the vertex } i \text{ is covered by two types of vehicle} \\ 0 & \text{otherwise} \end{cases}$$

Then the TEAM model follows:

$$\max \quad \sum_{i \in V} d_i y_i \tag{2.2.9}$$

$$\text{s.t.} \quad \sum_{j \in W_{A,i}} x_{A,j} \geq y_i \qquad \forall\, i \in V \tag{2.2.10}$$

$$\sum_{j \in W_{B,i}} x_{B,j} \geq y_i \qquad \forall\, i \in V \tag{2.2.11}$$

$$\sum_{j \in W} x_{A,j} = p_A \tag{2.2.12}$$

$$\sum_{j \in W} x_{B,j} = p_B \tag{2.2.13}$$

$$x_{A,j} \leq x_{B,j} \qquad \forall\, j \in W \tag{2.2.14}$$

$$x_{A,j},\, x_{B,j} \in \{0,1\} \qquad \forall\, j \in W \tag{2.2.15}$$

$$y_i \in \{0,1\} \qquad \forall\, i \in V \tag{2.2.16}$$

This formulation has the same constraints of the MCLP, except for the (2.2.14), which imposes a hierarchy between the two types of vehicles. It has been added in case there was the need, but it can be removed if not. Furthermore, we considered only two types of vehicles, but the model can be easily generalised with n types of vehicles.

As already noted, the previous models do not address the problem of multiple coverages, namely every time an ambulance is dispatched there could be demand points the coverage of which is no longer ensured. Hence few models were designed to satisfy this requirement. Hogan and Revelle presented two models in [12] called BACOP1 and BACOP2. For the first one, let's define

$$u_i = \begin{cases} 1 & \text{if the demand node } i \text{ is covered twice} \\ 0 & \text{otherwise} \end{cases}$$

Then the BACOP1 is

$$\max \quad \sum_{i \in V} d_i u_i \tag{2.2.17}$$

$$\text{s.t.} \quad \sum_{j \in W_i} x_j \geq 1 + u_i \qquad \forall\, i \in V \tag{2.2.18}$$

$$\sum_{j \in W} x_j = p \tag{2.2.19}$$

$$x_j \in \{0, 1\} \qquad \forall\, j \in W \tag{2.2.20}$$

$$u_i \in \{0, 1\} \qquad \forall\, i \in V \tag{2.2.21}$$

We can note that this formulation is identical to the MCLP with a small difference in the constraint (2.2.18) where if $u_i = 1$ than we want the number of sites that cover the node i to be greater than or equal to two.

In the BACOP2, the idea behind MCLP and the BACOP1 are combined and the objective function becomes multiobjective in the levels of coverage. Let y_i as in MCLP and u_i as in BACOP1, then the model is:

$$\max \quad w \sum_{i \in V} d_i y_i + (1 - w) \sum_{i \in V} d_i u_i \tag{2.2.22}$$

$$\text{s.t.} \quad \sum_{j \in W_i} x_j \geq y_i + u_i \qquad \forall\, i \in V \tag{2.2.23}$$

$$u_i \leq y_i \qquad \forall\, i \in V \tag{2.2.24}$$

$$\sum_{j \in W} x_j = p \tag{2.2.25}$$

$$x_j \in \{0, 1\} \qquad \forall\, j \in W \tag{2.2.26}$$

$$y_i \in \{0, 1\} \qquad \forall\, i \in V \tag{2.2.27}$$

$$u_i \in \{0, 1\} \qquad \forall\, i \in V \tag{2.2.28}$$

where w is a weight chosen in $[0, 1]$. In the objective function we want to maximise both the first coverage and the backup coverage with a weight as trade-off; then the constraints (2.2.23) and (2.2.24) work in tandem to determine which nodes receive backup coverage. The first constraint determines the number of facilities within the coverage standard of a node. If the total is one or more, $y_i = 1$. If the total is only one, $u_i = 0$. If the total is two, u_i also takes the value of one. The second constraint ensures that backup coverage can only be provided if first coverage is already in place. Finally, this model can be extended to deal with greater levels of coverage simply adding $|V|$ variables for each level and the appropriate constraints.

The last deterministic model we want to analyse, which addresses the problem of backup coverage, was formulated by Daskin and Stern in [7] and is called Hierarchical Objective Set Covering (HOSC) model. It was proposed some years before the BACOP models, but we present it after them because they are not related and there are some similarities between HOSC and the model we developed. They started from the LSCM and they modify it to incorporate two objectives:

- minimize the number of ambulances that are required to cover each of the demand points in a preset coverage standard;

- given the minimum number of ambulances, maximise the sum over all demand points of the number of ambulances in addition to the one required by the point above, that can respond to calls in demand point i.

This last objective has been introduced to maximize the amount of multiple coverage in the system. Let

$s_i =$ the number of additional EMS unit capable of responding to a call in zone i in the preset coverage standard r

$w =$ some positive weight

Then the HOSC model is:

$$\min \quad w\sum_{j\in W} x_j - \sum_{i\in V} s_i \tag{2.2.29}$$

$$\text{s.t.} \quad \sum_{j\in W_i} x_j \geq 1 + s_i \qquad \forall\, i \in V \tag{2.2.30}$$

$$x_j \in \{0,1\} \qquad \forall\, j \in W \tag{2.2.31}$$

$$s_i \in \mathbb{N} \qquad \forall\, i \in V \tag{2.2.32}$$

$$\tag{2.2.33}$$

In the objective function they want to minimize the number of ambulances needed to provide a full coverage, but also minimize the opposite of the number of additional ambulances able to respond in all the demands points, which is equivalent to maximize the opposite quantity, everything balanced by an appropriate weight which depends from the nature of the problem.

Note that the objective function weights equally all ambulances that can respond to a call in zone i. From a practical point of view, one might like to use a decreasing set of weights for additional ambulances.

2.3 Probabilistic models

Probabilistic models start from the intuition that any given ambulance may be busy when it is called. This uncertainty can be integrating within the mathematical formulation or using a queueing framework [4]. One of the first probabilistic model has been presented by Daskin in [6] and it is called Maximum Expected Covering Location Problem (MEXCLP). In this model the unavailability of an ambulance was developed assuming that each vehicle has a probability of being busy, called busy fraction and denoted as $q \in [0,1]$, and each ambulance is independent, with the probabilistic meaning, from the others. So, if the node $i \in V$ is covered by k ambulances, then the probability that k ambulances are busy at the same time is q^k. Let $H_{i,k}$ be a random variable equal to the demand of node i, given that k ambulances cover that node; hence

$$H_{i,k} = \begin{cases} 0 & \text{with probability } q^k \\ d_i & \text{with probability } (1-q^k) \end{cases}$$

The expected covered demand is

$$\begin{aligned} E[H_{i,k}] &= 0 \cdot q^k + d_i(1-q^k) \\ &= d_i(1-q^k) \end{aligned}$$

and the marginal contribution of the k-th ambulance in the expected coverage is

$$\begin{aligned} E[H_{i,k}] - E[H_{i,k-1}] &= d_i(1-q^k) - d_i(1-q^{k-1}) \\ &= d_i(1-q^k-1+q^{k-1}) \\ &= d_i(1-q)q^{k-1} \end{aligned}$$

Now let

$$x_j = \text{the number of ambulances located at node } j$$

$$y_{ik} = \begin{cases} 1 & \text{if the demand node } i \text{ is covered by at least } k \text{ ambulances} \\ 0 & \text{otherwise} \end{cases}$$

Then the MEXCLP is

$$\max \quad \sum_{i \in V} \sum_{k=1}^{p} d_i(1-q)q^{k-1} y_{ik} \tag{2.3.1}$$

$$\text{s.t.} \quad \sum_{j \in W_i} x_j \geq \sum_{k=1}^{p} y_{ik} \qquad \forall\, i \in V \tag{2.3.2}$$

$$\sum_{j \in W} x_j \leq p \tag{2.3.3}$$

$$x_j \in \mathbb{N} \qquad \forall\, j \in W \tag{2.3.4}$$

$$y_{ik} \geq 0 \qquad \forall\, i \in V,\ k \in \{1, \dots, p\} \tag{2.3.5}$$

where in the objective function we want to maximize the total coverage weighted with the probability that the ambulances are busy and the inequality (2.3.2) states that the sum of ambulances in servers j, which cover i, must be greater than the number of ambulances that cover i.

Of course, one the crucial parameters of this model is the busy fraction. This parameter has to be estimated time by time according to the nature of the problem: for example, it could be the total estimated duration of calls for all demand points divided by the total number of ambulances. The validity of the model derives from the fact that the (2.3.1) is concave in k, which means that if $y_{ik} = 1$ than $y_{ih} = 1$, $\forall h \leq k$, otherwise one should add that constraint explicitly.

The next important probabilistic model comes in two variants. Developed by ReVelle and Hogan in [25], it is called Maximum Availability Location Problem (MALP). This model seeks to position p ambulances in such a way that the maximum demand has a server available within the preset coverage standard r with a reliability level α. The two versions of MALP both have this same problem statement. They differ, however, in the manner in which the busy fraction of the servers is calculated. MALP I assumes all servers has the same probability of being busy, while the second version relaxed this assumption. We discuss only about MALP I since the MALP II is a generalisation that can be easily derived from the first one. Let x_j as defined in LSCM and y_{ik} as in MEXCLP, namely

$$x_j = \begin{cases} 1 & \text{if an ambulance is located in the vertex } j, \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ik} = \begin{cases} 1 & \text{if the demand node } i \text{ is covered by at least } k \text{ ambulances} \\ 0 & \text{otherwise} \end{cases}$$

$$q = \text{the busy fraction}$$

Hence the probability of the ambulance in node j of being busy is q^{x_j}. Note that if $x_j = 0$, which means that an ambulance is not present, then $q^{x_j} = 1$, which means that such ambulance is always busy with probability 1. Then the probability that all servers j, which cover node i, are busy is

$$\Pr[\text{all servers } j, \text{ which cover node } i, \text{ are busy}] = q^{\sum_{j \in W_i} x_j}$$

and the complementary is the probability that at least one server, which cover i, is not busy

$$\Pr[\text{at least one server, which covers } i, \text{ is not busy}] = 1 - q^{\sum_{j \in W_i} x_j}$$

In the model we want the last quantity to be greater than a reliability level α, namely

$$1 - q^{\sum_{j \in W_i} x_j} \geq \alpha$$

Recalling that $q \in [0, 1]$ and x_j is binary, we can linearize on x_j obtaining

$$q^{\sum_{j \in W_i} x_j} \leq 1 - \alpha$$

$$\sum_{j \in W_i} x_j \geq \lceil \log_q(1 - \alpha) \rceil = \left\lceil \frac{\log(1 - \alpha)}{\log q} \right\rceil =: b \qquad \forall\, i \in V$$

Now we can use this bound in the model, that can be written as

$$\max \quad \sum_{i\in V} d_i y_{ib} \tag{2.3.6}$$

$$\text{s.t.} \quad \sum_{j\in W_i} x_j \geq \sum_{k=1}^{b} y_{ik} \qquad \forall\, i \in V \tag{2.3.7}$$

$$y_{i,k} \leq y_{i,k-1} \qquad \forall\, i \in V,\ k \in \{2,\dots,b\} \tag{2.3.8}$$

$$\sum_{j\in W} x_j = p \tag{2.3.9}$$

$$x_j \in \{0,1\} \qquad \forall\, j \in W \tag{2.3.10}$$

$$y_{ik} \in \{0,1\} \qquad \forall\, i \in V,\ k \in \{1,\dots,p\} \tag{2.3.11}$$

Note that in this case the objective function is no longer concave in k, so we have to add (2.3.8) explicitly.

For completeness let's see also the constraint when the busy fraction is different for all ambulances. The probability of an ambulance of being busy is $q_j^{x_j}$, then the probability that all servers j, which cover node i, are busy is

$$\Pr[\text{all servers } j\text{, which cover node } i\text{, are busy is}] = \prod_{j\in W_i} q_j^{x_j}$$

and we want

$$\begin{aligned}
1 - \prod_{j\in W_i} q_j^{x_j} \geq \alpha \qquad & \iff \\
\prod_{j\in W_i} q_j^{x_j} \leq 1-\alpha \qquad & \iff \\
\sum_{j\in W_i} \log q_j^{x_j} \leq \log(1-\alpha) \qquad & \iff \\
\sum_{j\in W_i} (\log q_j) x_j \leq \log(1-\alpha) \qquad &
\end{aligned}$$

which is the constraint to add to the model.

The last probabilistic model was formulated in [2] as an extension of the LSCM and it is called Rel-P. The idea behind this model is similar to the MALP, but instead of computing server busy probabilities, they computed the probability that a given demand point will not find an available server. They defined a cost c_{jk} as the cost of locating k ambulances at node j and a quantity p_j, which is the maximum number of ambulances located at node j; then let

$$x_{jk} = \begin{cases} 1 & \text{if } k \text{ ambulances are located in the vertex } j, \\ 0 & \text{otherwise} \end{cases}$$

The model is

$$\min \quad \sum_{j \in W} \sum_{k=1}^{p_j} c_{jk} x_{jk} \tag{2.3.12}$$

$$\text{s.t.} \quad \sum_{k=1}^{p_j} x_{jk} \leq 1 \qquad \forall\, j \in W \tag{2.3.13}$$

$$\sum_{j \in W_i} \sum_{k=1}^{p_j} a_{jk} x_{jk} \geq b_i \qquad \forall\, i \in V \tag{2.3.14}$$

$$x_{jk} \in \{0,1\} \qquad \forall\, j \in V,\ k \in \{1, \dots, p_j\} \tag{2.3.15}$$

The constraint (2.3.14) is linear in the number of vehicles required to achieve a given reliability level. The constant a_{jk} and b_i have to be computed to ensure that given the number of ambulances covering demand point i, the probability of being unable to answer a call does not exceed a certain value. The mathematical definition of these two parameters is explained in the original paper, so we refer to that for more details. In short: a_{jk} is proportional to the logarithm of the probability that the number of calls in a given area is greater than k, while b_i is proportional to the logarithm of the reliability level.

All the probabilistic models we discussed about have an element in common: the probability of being busy must be estimated a priori and then used as a fixed input parameter; but in the literature there are also models based on spatially distributed queuing theory or simulation. Larson's hypercube model represents the most notable milestone for approaches using a queuing framework. The hypercube model and its various extensions have been found particularly useful in determining performance of EMS systems, but the tools used are advanced and their explanation is beyond the scope of this work. More details can be found in [14, 15, 37].

2.4 Dynamic Models

In both deterministic and probabilistic models, a long term perspective was taken and the demand in these models was naturally assumed to be the same for all periods, therefore static. However, in a real context, demand is not static, but fluctuates throughout the week, day of week, and even hour by hour within a given day. Dynamic redeployment models can aid managers make daily or even hourly plans to better respond to predictable demand fluctuations by time and space, but inevitably the ambulance relocation problem is more difficult to tackle since the model can become nonlinear and since it has to be solved more frequently.

Repede and Bernardo in [24], extended MEXCLP for multiple time intervals to capture the temporal variations in demand and unit busy probabilities, hence, called their model TIMEXCLP. Their application of TIMEXCLP to Louisville, Kentucky resulted in an increase of coverage while the average response time decreased by 36%. One of the most comprehensive dynamic relocation models was developed by Gendreau et al. [11]. The objective of their Dynamic Double Standard Model at time t (DDSM^t) is to maximize backup coverage within a radius r_1, while minimizing relocation costs. In addition to the standard coverage and site capacity constraints, the model takes into account a number of practical considerations inherent to the dynamic nature of the problem:

- double covering constraints,
- constraints on the number of ambulances at each site,
- constraints avoiding to move the same ambulances repeatedly,
- constraints avoiding round trips,
- constraints avoiding long trips.

Moreover, they consider two types of covering constraints. The absolute covering constraints require that all the demands be satisfied by an ambulance within r_2 minutes, and the relative covering constraints state that a proportion α of the total demand is also within r_1 minutes of an ambulance, with $r_2 > r_1$. Let

$$W_{i,1} = \{j \in W : t_{ij} \leq r_1\}$$

$$W_{i,2} = \{j \in W : t_{ij} \leq r_2\}$$

$$M^t_{j,l} = \begin{array}{l}\text{cost of repositioning, at time } t\text{, ambulance } l \\ \text{from its current site to site } j \in W\end{array}$$

$$x_{jl} = \begin{cases} 1 & \text{if ambulance } l \text{ is moved to site } j \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ik} = \begin{cases} 1 & \text{if node } i \text{ covered at least } k \text{ times} \\ 0 & \text{otherwise} \end{cases}$$

Then the redeployment model at time t is

$$\max \quad \sum_{i \in V} d_i y_{i,2} - \sum_{j \in W} \sum_{l=1}^{p} M^t_{jl} x_{jl} \tag{2.4.1}$$

$$\text{s.t.} \quad \sum_{j \in W_{i,2}} \sum_{l=1}^{p} x_{jl} \geq 1 \qquad \forall\, i \in V \tag{2.4.2}$$

$$\sum_{i \in V} d_i y_{i,1} \geq \alpha \sum_{i \in V} d_i \tag{2.4.3}$$

$$\sum_{j \in W_{i,1}} \sum_{l=1}^{p} x_{jl} \geq y_{i,1} + y_{i,2} \qquad \forall\, i \in V \tag{2.4.4}$$

$$y_{i,2} \leq y_{i,1} \qquad \forall\, i \in V \tag{2.4.5}$$

$$\sum_{j \in W} x_{jl} = 1 \qquad \forall\, l \in \{1, \ldots, p\} \tag{2.4.6}$$

$$\sum_{l=1}^{p} x_{jl} \leq p_j \qquad \forall\, j \in W \tag{2.4.7}$$

$$x_{jl} \in \{0, 1\} \qquad \forall\, j \in W,\, l \in \{1, \ldots, p\} \tag{2.4.8}$$

$$y_{ik} \in \{0, 1\} \qquad \forall\, i \in V\; k \in \{1, 2\} \tag{2.4.9}$$

In the model, relation (2.4.2) ensures the single coverage requirement, (2.4.3) imposes that a proportion α of all demand is covered, (2.4.4) ensures that the number of ambulances located within r_1 must be at least one if $y_{i,1} = 1$ or at least two if $y_{i,2} = 1$, (2.4.5) states that a demand point cannot be covered twice if it is not covered at least once, (2.4.6) specifies that each available ambulance must be assigned to a potential location site and (2.4.7) assigns an upper bound to the number of ambulances that each site can host. This last constraint can be deleted if arbitrary ambulances are allowed in each site.

To solve DDSM^t, Gendreau, Laporte and Semet have developed a fast tabu search heuristic implemented on parallel processors. This algorithm runs non-stop and continuously computes the best possible redeployment plans associated with the current positions of ambulances, in response to each potential anticipated ambulance request.

Chapter 3

Ambulance deployment as a Flow problem

In the previous chapter we have seen the state of the art for addressing the problem, but as already said, when there is the need to solve the problem of the ambulance minimisation in a Region or a State, the requirements change and this affects the formulation of the model. In the next sections we are going to explain exhaustively the requirements, the emergency medical services in Veneto, the reasons why we formulated a different model instead of applying one in the literature and also the reasons behind some assumptions.

3.1 The health system in Veneto

In this section, the aim is to provide an overview of the Veneto health system, touching on the main points that concern the problem we want to address.

The National Health System in Italy, called "Sistema Sanitario Nazionale" (SSN), is born in the 1978 with the Law n. 833 of 1978 [31]. Over the years the national health system has undergone numerous changes from an organizational point of view. Without going into detail and reporting only the main events, until 1992, the Local Health Units (USL) were designed as operational structures of the municipalities, individuals or associates, or of the Mountain Communities. The territory therefore coincided with the municipal one, while in the big cities with the areas of urban decentralization.

However, due to the ever increasing costs, in 1992 it was decided to radically change the organizational structure of the system. Legislative Decree n. 502 of 30 December 1992 [29] and the following amendments and additions made by Legislative Decree of 17 December 1993, n. 517 [30], reorganized the SSN operating on the nature of the USL, on the powers and responsibilities of the Regions. From this moment a process of aggregation and consolidation began; those Legislative Decrees establish that the Local Health Authorities (ASL in Italian), born from the old ASL, must coincide with the territory of the Provinces as a result of their conversion into Companies and therefore with recognition of public legal personality and instrumental bodies of the Region. The merger started in 1992 leads to a significant decrease in the number of local health authorities within the national territory with the aim of optimizing the management and containing the costs of healthcare expenditure.

This type of trend, after developing over the years, underwent a new acceleration at the regional level in 2015 with the implementation of a series of reforms that involved also the Veneto Region. Towards the end of June 2015, a draft law was delivered to the regional council, then converted into Regional Law n. 19 of 25 October 2016 [21]. In this law, a new governance system for the regional health system and a broader reorganization of the territory was imposed. The most important new features were:

- the merger of the previous ULSS companies ("Unità Locale Socio Sanitaria"); the new organisation includes nine ULSS company,
- the formation of a new regional instrumental body called "Azienda Zero" in which the planning, administration and management of regional functions are concentrated.

The establishment of Azienda Zero has the aim to bring together and concentrate in the hands of a single entity the functions of the health planning, as well as the organization and governance of the Regional Health System, attributing to it the operations of technical-administrative management on the regional territory. The centralization of the ULSS aims to allow the Region to save money and speed up procedures, thus allowing companies to be free to deal with the organization of services for citizens in the best possible way [16]. Hence, Azienda Zero is in charge for the management of the financial flows of regional health. In particular, it is in charge of the investments evaluation in the health sector and their monitoring and of the supply of ambulances in the public environment for the whole Region.

For what concern the new organisation, the nine ULSS are:

- Azienda ULSS 1 "Dolomiti",
- Azienda ULSS 2 "Marca Trevigiana".
- Azienda ULSS 3 "Serenissima",
- Azienda ULSS 4 "Veneto Orientale",
- Azienda ULSS 5 "Polesana",

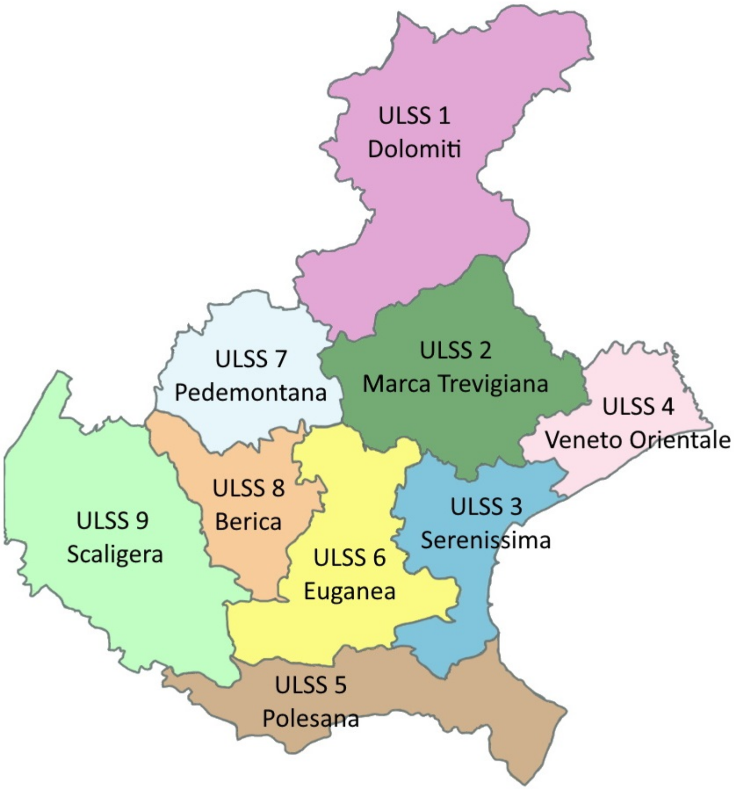


Figure 3.1: The Region of Veneto divided in the nine ULSS

- Azienda ULSS 6 “Euganea”,
- Azienda ULSS 7 “Pedemontana”,
- Azienda ULSS 8 “Berica”,
- Azienda ULSS 9 “Scaligera”.

A graphical representation of this division is shown in Figure 3.1.

As already mentioned in the introduction, emergency medical services are under Public Health Authorities control in each Italian Region and the ambulance subsystem is provided by a variety of different sources. The method of delivery can vary considerably from one location to another. In some locations, responsibility for the provision of EMS has been undertaken by the local hospital, while in others, services may be provided by a range of volunteer organizations, such as the Italian Red Cross (Croce Rossa Italiana), ANPAS (National Association for Public Assistance), other associations commonly known as “Cross” (Croce), usually followed by a colour (White Cross, Green Cross, Yellow Cross...), or by private companies [34].

In particular in Veneto, for ULSS 1 and 2, the service is characterized by the management of the urgent emergency service which involves the predominant use of vehicles owned by the ULSS with the support of volunteers. The service of ULSS 5, 6, 7 and 9 is managed by private entities, while for ULSS 3, 4 and 8, the service is carried out mainly with vehicles owned by the ULSS. In any case, this simple division between public and private is not so marked; particular conditions affect every ULSS: for example, in ULSS 5, an additional support of private entities is present especially in the summer due to a significant difference in requests dictated by seasonality.

Due to this organisation, we decided to consider and study each ULSS independently. Now recalling the main requirements of the problem, provided in Chapter 1, we cannot use any of the models whose primary goal is to maximise the level of service. Moreover, we decided to exclude also the LSCM for the reasons already explained in Section 2.2 and also the Rel-P because the focus is the minimisation of the costs, furthermore the parameters of the model were complex to estimate with the data available. The HOSC model, instead, is unable to provide the location of the additional vehicles and, similar to the LSCM, its primary goal is the coverage, so the model has no dependency on the amount of requests.

We want to conclude this section by also reporting other constraints that we encountered in the study of the problem, in the specific case of Veneto, from DGR n. 1515 of 29 October 2015 [19], but in the linear programming model we developed are not relevant:

- the institution must have at least two ambulances with minimum characteristics required for the accredited activity. Voluntary associations can have only one ambulance if they are based in a mountain municipality or in an island area with a population of less than 1500 inhabitants,

- ambulances must have a maximum age of 7 years from the first registration and a maximum number of 400000 kilometers,

- the emergency vehicles must be at most 12 years old from the first registration;

- the institution must have at least one reserve vehicle for every three vehicles in active service for companies up to nine vehicles in service,

- institutions with two vehicles may maintain only one vehicle in active service.

3.2 The model

In the previous sections we motivated the reasons why the models developed in the literature were not suitable in out context, so we decided to formulate a new model that could satisfy our requirements. The idea behind our model is that considering a specified range of time and a given region with demand points and service points there is a *flow* of requests from the demand points which ends in the service points and based on this incoming flow there must be an appropriate number of ambulances to satisfy that demand. This kind of problem can be described using directed graphs called, in literature, flow network.

A flow network is a directed graph $G = (V, A)$, in which a capacity $k_{ij} \geq 0$ is associated to every directed edge (i, j) and eventually a cost $c_{ij} : A \rightarrow \mathbb{R}^+$. There are also two vertices s and t, called respectively source and sink. The most common optimisation problem on flow networks aims to send the maximum flow from s to t. A more general problem aims to find a flow of minimum cost from s to t. This latter is defined from the following linear programming model:

$$\min \sum_{(i,j)\in A} c_{ij}x_{ij} \tag{3.2.1}$$

$$\text{s.t.} \sum_{(h,j)\in\delta^+(h)} x_{hj} = \sum_{(i,h)\in\delta^-(h)} x_{ih} \qquad \forall\, h \in V \setminus \{s,t\} \tag{3.2.2}$$

$$0 \leq x_{ij} \leq k_{ij} \qquad \forall\, (i,j) \in A \tag{3.2.3}$$

where $\delta^+(h) = \{(h, j) \in A : j \in V\}$ and $\delta^-(h) = \{(i, h) \in A : i \in V\}$. Hence in (3.2.2) we want that from each node the outgoing flow (the left hand side) must be equal to the incoming flow (the right hand side).

Now the requirements seen in Section 3.1 can be modeled using the tools just presented with some changes. We do not need the concept of capacity of a directed edge, but we need the capacity of a hospital, which is the maximum number of ambulances that a hospital can host and these points are also the sink; then we have a demand of a node, which quantity has to be defined appropriately, and these nodes are the sources.

Now we can present the model; first of all we decided to denote by t_{ij} the Euclidean distance, often denoted by $d(i, j)$, in order not to confuse it with the demand of a node d_i; this choice will be explained in Section 4.1. Hence, keeping a notation similar to that used in the literature, let

$$r_i = \text{preset coverage standard of node } i \in V$$

$$d_i = \text{demand of node } i \in V$$

$$
\begin{aligned}
x_{ij} &= \text{fraction of demand of } i \in V \text{ served by } j \in W \\
y_j &= \text{number of ambulances of } j \in W \\
k_j &= \text{capacity of hospital } j \\
W_i &= \{j \in W : t_{ij} \leq r_i\} \\
V_j &= \{i \in V : t_{ij} \leq r_i\} \\
w &= \text{a positive weight}
\end{aligned}
$$

First of all we can note that x_{ij}, $w \in \mathbb{R}$ and y_j, $k_j \in \mathbb{N}$. Then the model is:

$$
\min \quad \sum_{j \in W} y_j + w \sum_{i \in V} \sum_{j \in W} t_{ij} x_{ij} \tag{3.2.4}
$$

$$
\text{s.t.} \quad \sum_{j \in W_i} x_{ij} \geq d_i \qquad \forall\, i \in V \tag{3.2.5}
$$

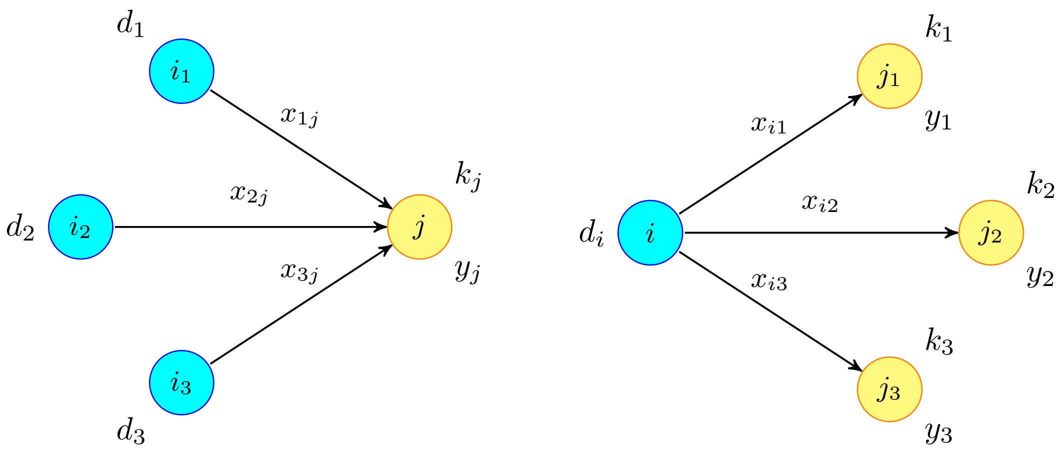
$$
\sum_{i \in V_j} x_{ij} \leq y_j \qquad \forall\, j \in W \tag{3.2.6}
$$

$$
x_{ij} \geq 0 \qquad \forall\, (i, j) \in E \tag{3.2.7}
$$

$$
y_j \in \{0, \ldots, k_j\} \qquad \forall\, j \in W \tag{3.2.8}
$$

where constraint (3.2.5) states that the fraction of demand served by hospitals around a node i, must be greater than the demand of such node and constraint (3.2.6) states that the demand served by a hospital j cannot exceed the number of ambulances present in j. Another important consideration is that in the objective function (3.2.4) there is an additional term with respect to the sum of the ambulances: we want to minimize also the sum of the fraction of demand multiplied by the distance and everything weighted with w, because in this way the model will be inclined to serve a point through its nearest hospital, which results in a faster response time of an ambulance. This concept is explained in depth in Section 4.1. Finally, we added a capacity to every hospital on the maximum number of ambulances it can host; in practice this constraint is usually useless because in most occasions there is no limit to the ambulances that a hospital can host.

A graphical representation of the concepts explained can be found in Figure 3.2. On the left there are three demand points in blue, linked to one hospital in yellow; the idea is that the hospital has to satisfy the fraction of demand of all this points, so the number of ambulances must be greater than the sum of all requests, which is the constraint (3.2.6) in the model. At the same time we can have multiple hospitals that can serve a demand point, figure on the right, and in this case we



(a) The number of ambulances must be greater than the fraction of demand of such points served by the hospital and less than the capacity

(b) The total demand of a node must be satisfied by the available hospitals

Figure 3.2: Graphical representation of the incoming and outgoing flow through demand and service points

want that the fraction of demand served by all hospitals must be greater than the demand of the node, which is the constraint (3.2.5).

This model has also several interesting properties:

Theorem 3.2.1. *In the optimal solution, if $t_{ij} > r_i$ then $x_{ij} = 0$, $\forall (i,j) \in V \times W$.*

Proof. Suppose by contradiction that there exists (i,j) such that $t_{ij} > r_i$ and $x_{ij} > 0$ and let z^* the optimal value of the objective function of the model under this assumption. Then $t_{ij} > r_i$ implies that $i \notin W_j$ and $j \notin V_i$, which means that for any value of $x_{ij} \geq 0$, the two constraints (3.2.5) and (3.2.6) hold; now we can consider the same objective function forcing $x_{ij} = 0$ and let $\hat{z}$ the value of this quantity, namely:

$$\begin{aligned}
\hat{z} &= \sum_{l \in W} y_l + w \sum_{k \in V} \sum_{l \in W} d(k,l) x_{kl} \\
&= \sum_{l \in W} y_l + w \sum_{\substack{k \in V \\ (k,l) \neq (i,j)}} \sum_{l \in W} d(k,l) x_{kl} + \underbrace{w d(i,j) x_{ij}}_{=0} \\
&= \sum_{l \in W} y_l + w \sum_{\substack{k \in V \\ (k,l) \neq (i,j)}} \sum_{l \in W} d(k,l) x_{kl} < z^*
\end{aligned}$$

which is impossible, because we assumed z^* the optimum. Note that the last inequality comes from the fact that since $w > 0$, $d(i,j) > 0$ then $wd(i,j)x_{ij} > 0$ in z^*. $\square$

This lemma means that we can formulate the model equivalently writing (3.2.5) and (3.2.6) respectively as

$$\sum_{j \in W} x_{ij} \geq d_i \qquad \forall\, i \in V$$

$$\sum_{i \in V} x_{ij} \leq y_j \qquad \forall\, j \in W$$

and imposing the constraint

$$x_{ij} = 0 \quad \forall\, (i,j) \in V \times W, \text{ if } \quad t_{ij} > r_i$$

This latter can be useful when you want to implement the model in practice.

Theorem 3.2.2. *The minimum number of ambulances deployed in the server points is always greater than the sum of the demand of the entire system.*

Proof. Let y the minimum number of ambulances deployed and d the total demand of the system, i.e.,

$$y = \sum_{j \in W} y_j \qquad\qquad d = \sum_{i \in V} d_i$$

By Theorem 3.2.1 and recalling (3.2.5), (3.2.6) and (3.2.7), we have that

$$\sum_{j \in W} x_{ij} = \sum_{j \in W_i} x_{ij} \geq d_i \qquad \forall\, i \in V \tag{3.2.9}$$

$$\sum_{i \in V} x_{ij} = \sum_{i \in V_j} x_{ij} \leq y_j \qquad \forall\, j \in W \tag{3.2.10}$$

Now we can sum (3.2.9) over $i \in V$ and sum (3.2.10) over $j \in W$ obtaining

$$d = \sum_{i \in V} d_i \leq \sum_{i \in V} \sum_{j \in W} x_{ij} \leq \sum_{j \in W} y_j = y$$

which concludes the proof. Moreover, since we require y_j to be integer, we can improve the bound:

$$y \geq \left\lceil \sum_{i \in V} \sum_{j \in W} x_{ij} \right\rceil \geq \left\lceil \sum_{i \in V} d_i \right\rceil$$

$\square$

Theorem 3.2.2 tells us that the total demand of the system is a lower bound of the number of ambulances we have to dispatch in the system.

Chapter 4

Implementation and results

In this chapter we are going to describe the data acquisition, the implementation choices and the results obtained. All the experiments were executed on a personal computer, the exact solver we used is IBM ILOG CPLEX version 12.10 [13] and for the data manipulation and generation we used Python 3.7 [10].

4.1 Implementation choices

In Chapter 3 we described the model without specifying in detail the parameters used in the implementation. Now suppose to apply the model in a given region and all the points inside are demands points: to implement such situation we decided to divide that region in squares of a given dimension and then consider the coordinates of the center of the square as a demand point $i \in V$ producing a grid of points. Of course, this approach does not consider the conformation of a real environment, for example the geographical conformation or the presence of impediments. In the same way we can define the service points as points whose coordinates are the center of a square of the grid. In this way the sets of demand and service points are defined. Figure 4.1 show an example of a 10×10 grid where demand points are represented in blue and service points in yellow.

As anticipated, another assumption is that the response time of an ambulance is the distance between the service point, where the rescue vehicle starts, and the demand point expressed in minutes. This because we assumed the average velocity of an ambulance to be 60 km/h and the path between a service and a demand point to be a straight line. Hence, considering two points $i \in V$ and $j \in W$, the time to reach the point i from j is the Euclidean distance, namely:

$$t_{ij} = d(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

where x_i is the x coordinate of point i and the other quantities follow. Note that

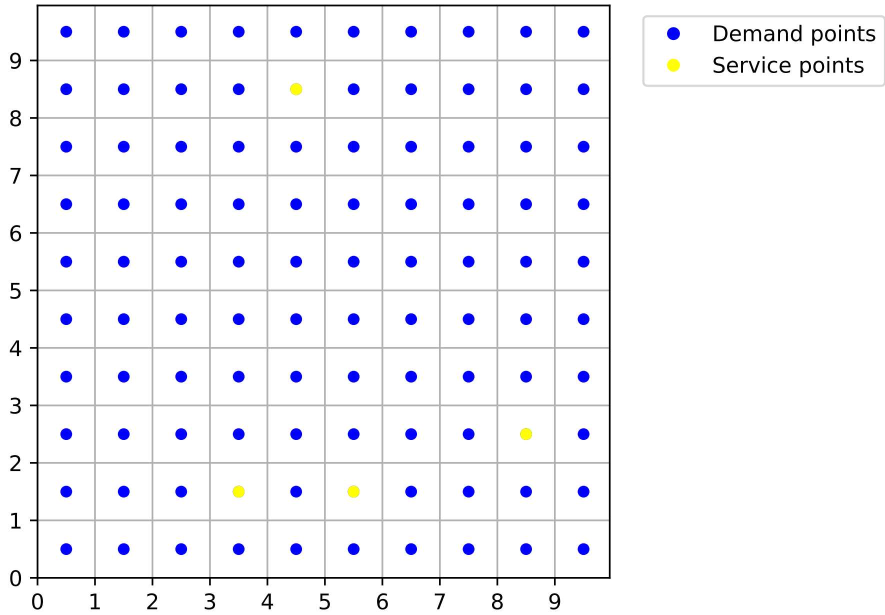


Figure 4.1: Example of a region divided in a 10×10 grid of demand points in blue; the yellow points represent the hospitals

t_{ij} is no longer a unit of time but one of distance; this abuse of notation is lawful provided that the preset coverage standards are also consistent. Similarly to the previous case, this is a strong assumption because the path between two points in the real life is not a straight line in general, and the average speed of an ambulance changes from area to area, but it can be easily modified based on the situation.

In Chapter 1 we stated that the maximum response time is eight minutes for urban areas and twenty minutes for extra-urban areas. First of all, using the grid approximation of a region, it is difficult identifying an urban area from an extra-urban; moreover, like in other countries, this standard is not currently being always met since there is a great disparity in geography and healthcare efficiency, depending on which region is being analyzed [34]. Hence in Veneto, the Deliberation of the Regional Council n. 128 of March 8, 2019 [20] decreed that the maximum response time must be 18 minutes for each area and, since the velocity of an ambulance we assumed to be 60 km/h = 1 km/min, then in the model the preset coverage standard is $r_i = r = 18\,\text{km}$, $\forall\, i \in V$.

Another important consideration regarding the model is that there is not any law that regulate the maximum number of ambulances that a hospital can host, which leads us to edit the constraint (3.2.8) in the following:

$$y_j \in \mathbb{N} \qquad \forall\, j \in W$$

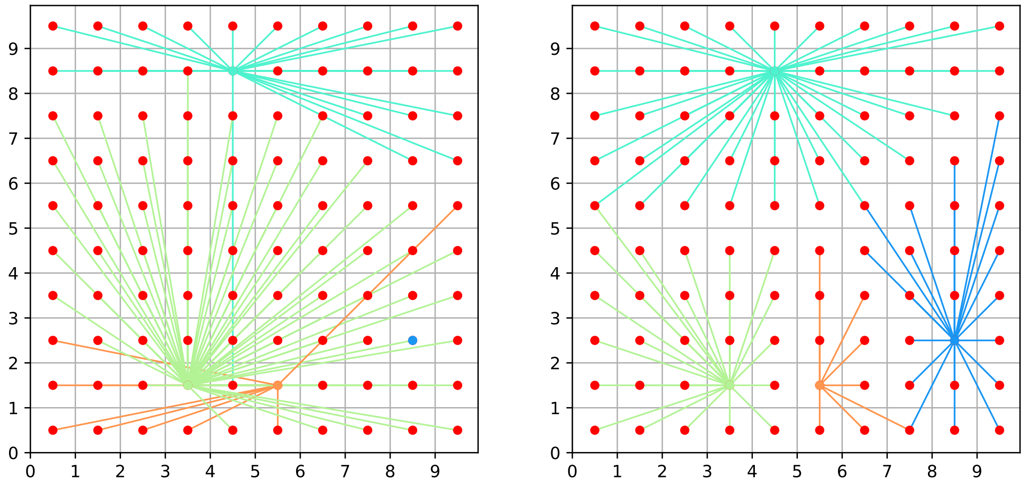
The next important parameter we have to describe is w: as said previously adding that additional term in the objective function weighted with a positive weight, the model will be inclined to serve a point through its nearest hospital, which results in a faster response time of an ambulance. Note that this consideration is not necessary because in this sense there is no legal constraint to satisfy but let's have a look at both situations.

Let's consider the case $w = 0$: in this case a hospital can serve all the points within a radius equal to r, hence in the optimal solution a demand point covered by more than one hospital may not be served by the closest. The solution of the model is correct and it respects all the laws, but it's not difficult to understand that if a demand point is served by a hospital which is not the nearest, then the respond time is greater and in the healthcare sector, the shorter the response time, the higher the probability of a patient to survive. This observation is expressed by one of the requirements in Chapter 1. For this reason, we decided to add a term which depends on the mutual distances between service and demand points. This term must be multiplied by an appropriate weight since the focus of the model must be to minimise the number of ambulances.

In our case, as we will see, the order of magnitude of y_j is about $10^0 \sim 10^1$, the order of magnitude of t_{ij} is $10^1 \sim 10^2$ and for x_{ij} it is $10^{-3} \sim 10^{-2}$ so we decided to let $w = 10^{-1}$ making the order of magnitude of the number of ambulances slightly higher than the other. In this way the model will choose to reduce as much as possible the number of ambulances and secondly to serve each point with the nearest hospital. In the implementation phase we decided to divide the objective function by w in order to avoid problems of numerical representation. The results do not change, they are only scaled by a factor w^{-1}.

In Figure 4.2 we can see an example of this argument, where we used the grid seen previously, we set $r = 7$ and we generate random number to simulate a demand for each point. In particular Figure 4.2a shows the plot of the optimal solution with $w = 0$ and we can see for example that the demand point in $(4.5, 1.5)$ is served by the hospital in $(4.5, 8.5)$, which is 7 unit distant, even if the hospital in $(3.5, 1.5)$ is only 1 unit distant; the same argument can be applied for the demand point in $(3.5, 8.5)$. On the other hand, Figure 4.2b, where $w > 0$ we can see that the model is inclined to serve a point through its nearest hospital. In practice this turns into a faster response time and hopefully a greater chance of saving the patient.

Note that if $w = 0$ the Theorem 3.2.1 does not hold anymore, because there can be a $x_{ij} > 0$ such that $t_{ij} > r$ and the value of the objective function stays the same. In particular all the constraints continue to be valid for any value of x_{ij}, which is conceptually wrong. If one would like to implement the model with $w = 0$, then the alternative formulation seen in Section 3.2 is useful.



(a) $w = 0$: in this case a hospital can serve a point which is much more distance with respect to another hospital

(b) $w = 10^{-1}$: the model is inclined to serve a point through its nearest hospital

Figure 4.2: Example for the two different values of w with $r = 7$

Other important considerations that come out from Figure 4.2a and 4.2b are the following: in the second figure there are demand points which are partially served by a hospital and partially by another one; this situation raises a question related to the hospital from which an ambulance must be deployed in case of a request. This argument will be discussed later because it seldom occurs and can be solved using unrelated approaches. The second consideration involves the first figure, because there is a hospital, the blue circle, which does not host any ambulance. This means that such structure is useless in our context and this is allowed in the optimal solution and is not formally regulated by any law. However, in the health sector, there are other costs and even much greater than those involving the procurement of ambulances, for example hospital maintenance, that have led us to modify again the constraint (3.2.8) such that $y_j \in \mathbb{N}^*$; in this way each hospital hosts at least one ambulance.

Thanks to this last modification we can also improve the bound given in the Theorem (3.2.2).

Lemma 4.1.1. If $y_j \in \mathbb{N}^*$, $\forall\, j \in W$ then

$$y = \sum_{j \in W} y_j \geq \max\left\{ |W|, \left\lceil \sum_{i \in V} d_i \right\rceil \right\}$$

Proof. From Theorem 3.2.2 we know that

$$y = \sum_{j \in W} y_j \geq \left\lceil \sum_{i \in V} d_i \right\rceil$$

Now by hypothesis $y_j \in \mathbb{N}^*$, $\forall\, j \in W$ we have that $y_j \geq 1\, \forall\, j \in W$, hence

$$y = \sum_{j \in W} y_j \geq \sum_{j \in W} 1 = |W|$$

and the thesis follows. □

4.2 Data acquisition and manipulation

In the previous section we explained how we set some parameters showing some cases of applicability and assuming that we have all the data necessary to solve the model. In this section we want to explain how those data were generated and what regions we have considered.

4.2.1 Defining the sets of demand and service points

As already explained, in Italy each region manages the health services independently through the local healthcare companies. In Veneto they are called ULSS and are nine. Hence, we decided to take the ULSS independently and at the first approach we focused on two of them:

- Azienda ULSS 6 "Euganea",
- Azienda ULSS 8 "Berica".

Recall that, from a given region, we want to create the grid of squares of a given size. To accomplish this goal, we first fixed the dimension of the square edge to 1 km, then we decided to consider an image representing the region and, using OpenCV [3], interpolate the image extracting the pixels of interest and, from them, derive the demand points. To make this procedure uniform, the starting image has to be in black and white with the interesting region in black. This avoid problems of segmentation and the black and white image can be easily obtained with common programs for image processing.

In Figure 4.3 we can see the process of segmentation in details for ULSS 6 "Euganea": Figure 4.3a shows the starting black and white image and Figure 4.3b shows the interpolated image. Note that for the edges we decided to insert a point if at least a half of the pixels in that square are black.

(a) The starting black and white image: this image has been created using GIMP [27]

(b) The resulting grid: each square is represented by the coordinates of its center in the model

Figure 4.3: The black-and-white and the interpolated image of ULSS 6

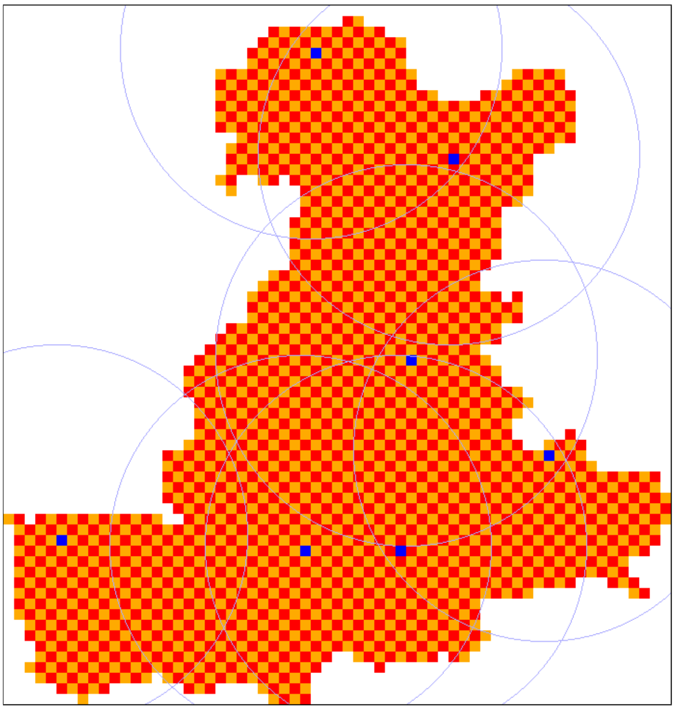


Figure 4.4: The coverage of the hospitals in ULSS 6 with $r = 18\,\mathrm{km}$

```
IMG_SRC : ulss6/ULSSEuganeaBW.png
IMG_PX_PER_KM : 11
NUMBER_HOSPITALS : 7
MAX_DISTANCE : 18 km
HOSPITALS_COORD_SECTION
1 319 682
2 462 572
3 418 363
4 561 264
5 55 176
6 308 165
7 407 165
```

Figure 4.5: An example of the input instance

For what concern the service points, there is no source from which it is possible to obtain the hospital coordinates automatically; for this reason, it was necessary to manually specify the coordinates within the image. Another element which must be manually specified is the number of pixels corresponding to one kilometer, to interpolate the image with squares of $1\,\mathrm{km}^2$. All this information is related to a particular image, therefore it was decided to create a starting instance where the list of parameters was inserted specified as a textual document. A first example of the starting instance is shown in the Figure 4.5 where:

- IMG_SRC is the path of the black and white image,
- IMG_PX_PER_KM is the number of pixels corresponding to one kilometers,
- NUMBER_HOSPITALS is the number of the service points within the region,
- MAX_DISTANCE is the preset coverage standard, whose value has been already discussed,
- HOSPITALS_COORD_SECTION is a keyword to notify that in the next NUMBER_HOSPITALS rows the coordinates of the service points are specified; we added an incremental index at the beginning of each row for control purpose.

From this data another important information can be derived: we said that the preset coverage standard in our case is $r = 18\,\mathrm{km}$, but the distribution of the hospitals in a region does not assure that all the service points are covered within that r. In fact, we can easily answer this question drawing a circle with radius r and check if all the points are covered. Figure 4.4 shows in blue the points where the hospitals are located and for each of them a circle shows what points are covered from that hospital. It is easy to see that few demand points in the center left of the figure are not covered so we can immediately conclude that the problem is impossible with such r. This means that with our assumptions there is no way of solving the problem with the constraint derived from Deliberation [20]. The problem could be overcome by increasing r to the minimum number of kilometers such that all points are covered, in the case of ULSS 6 it is $r = 20\,\text{km}$, but this violets such law because all demand points will have that r. This led us to modify the definition of r_i in this way: if for a point there exist at least a hospital within $r = 18\,\text{km}$ or less, than its preset coverage standard remains the same; otherwise we set the standard to the distance from the nearest hospital. In mathematical terms:

$$r_i = \max\big\{18, \min\{d(i,j) : j \in W\}\big\}$$

4.2.2 Extracting the demand of ambulances

As one can notice, in the previous sections we basically specified every element of the model but the demand of each node, which we recall is denoted with d_i. Since this parameter is the more complex to specify, we started with a pre-existing formula, which has been used as the reference value in the past. This formula has been proposed in [22] by AGENAS (the Italian National Agency for Regional Healthcare Services), which is a non-economic public body funded in 1993 and subject to oversight by the Ministry of Health. Its tasks are identified by the Standing Conference on the Relations between the State, the Regions and the Autonomous Provinces, and it also carries out the tasks laid down by the existing legislation [23].

Three factors were considered essential when the problem was studied in [22] and they were the following:

- the use of a homogeneous and rational criterion established at regional level by the Department of Health to ensure coverage of the territory with medical means, around which the network of basic ambulances is then placed,
- the standard requirement thus defined for each territory governed by the Operations Centers, must be reasoned and rationalized by the Heads of the Operations Centers, formalized by the Coordination and approved by the Department of Health,
- information sharing with the institutions that govern the territory is necessary, in order to avoid parochial pushes that unbalance the system (that's why the proposal must be technically defensible and homogeneous).

The formula proposed is the following:

$$\text{Number of ALS vehicles} = \frac{1}{2}\left(\frac{PLL}{60000} + \frac{PLM}{40000} + \frac{LA}{350\,\mathrm{km}^2} + \frac{MA}{300\,\mathrm{km}^2}\right) \qquad (4.2.1)$$

where:

$$PLL = \text{Population living in lowland}$$
$$PLM = \text{Population living in mountain}$$
$$LA = \text{Lowland area (in km}^2\text{)}$$
$$MA = \text{Mountain area (in km}^2\text{)}$$

Two elements from Equation (4.2.1) catch the attention from our point of view: the first one is the lowland and mountain areas because at the beginning of Section 4.1 we explained that the approximation derived from the grid construction we used, do not consider the geographical conformation, which means that we are not able to distinguish between lowland and mountain and this holds also for the population living in mountainous areas. This fact led us to simplify the formula setting $PLM = 0$ and $MS = 0$. The updated formula follows:

$$\text{Number of ALS vehicles} = \frac{1}{2}\left(\frac{PLL}{60000} + \frac{LS}{350\,\mathrm{km}^2}\right) \qquad (4.2.2)$$

The second important elements of (4.2.1) is the population itself, which we have never considered until now, so the formula can't be used if we don't have the population distribution inside the region. Hence, we decided to develop a population generator since this parameter can be useful for testing purposes and also for various considerations.

For this purpose, we considered bivariate normal distributions to generate our population and we decided to consider a restricted number of cities for convenience. In practice we considered the top-n most populated cities of an ULSS, we extracted the coordinates of those cities in the black and white image and finally we generated as much points as the total population in the ULSS, using n bivariate normal distribution with means the coordinates of each city and the probability of being called proportional to its own population.

Let's see an example to clarify using ULSS 6 as instance: we firstly extract the top-n most populated cities, in our case the top-15 cities, which are shown in the first two columns of Table 4.1; then we associated to each city its own coordinates in the black and white image, which we recall it's shown in Figure 4.3a. This data is reported in the third column of the table. The next step is to generate the population using this information: hence we iterated for 935000 times, which is

City	**Population**	**Coord. in the BW image (px)**	**Coord. in the interpolated image**
Padova	213696	(429, 381)	(39, 34)
Albignasego	26562	(411, 301)	(37, 27)
Selvazzano Dentro	23502	(330, 367)	(30, 33)
Vigonza	23462	(502, 424)	(45, 38)
Abano Terme	20455	(337, 321)	(30, 29)
Cittadella	20333	(324, 682)	(29, 62)
Piove di Sacco	20265	(544, 255)	(49, 23)
Monselice	17482	(305, 182)	(27, 16)
Rubano	16747	(334, 407)	(30, 37)
Cadoneghe	16460	(458, 427)	(41, 38)
Este	16367	(221, 172)	(20, 15)
Campodarsego	15025	(438, 491)	(39, 44)
Ponte San Niccolò	13663	(460, 348)	(41, 31)
Vigodarzere	13288	(419, 443)	(38, 40)
San Martino di Lupari	13208	(396, 682)	(36, 62)
Top-15 total population	470515		
Total Population	935000		

Table 4.1: The top-15 most populated cities of ULSS 6

the total population of ULSS 6, generating time by time one of the coordinates of the top-15 cities with a probability distribution proportional to the population of that city with respect to the other 14 cities. For example, Padova has a population of 213696, which is the 45% of the top-15 cities, hence the multivariate normal distribution with mean (429, 381) has the 45% of probability of generate a point. Finally we fixed the covariance matrix of each bivariate normal distribution; we set it empirically after multiple tests:

$$\Sigma = \begin{bmatrix} K_x & 0 \\ 0 & K_y \end{bmatrix}$$

where

$$K_x = \text{length of the region in kilometers}$$

$$K_y = \text{height of the region in kilometers}$$

To accomplish this goal, we have to add some additional information to the instance file in input; these information are the population of an ULSS, the dimension of the surface considered, in our case we have squares of 1 km of edge and we have to specify the population density of the cities, which we encoded in a CSV file.

This process is only an approximation of the population living in a certain area, as one can guess, but it has the advantage that it is scalable, because one can change the distribution at will or consider different cities to make the distribution more realistic or, even better, we can add as much cities as we want to improve the simulation. An example of the random generation for ULSS 6 is shown in Figure 4.6, where on the left we can see the horizontal plane and on the right the perspective. From these figures we can see that in the plot a single bivariate normal distribution is represented; this because there is a great imbalance between the population of Padova and that of all the other cities; moreover, some of those cities are located near Padova.

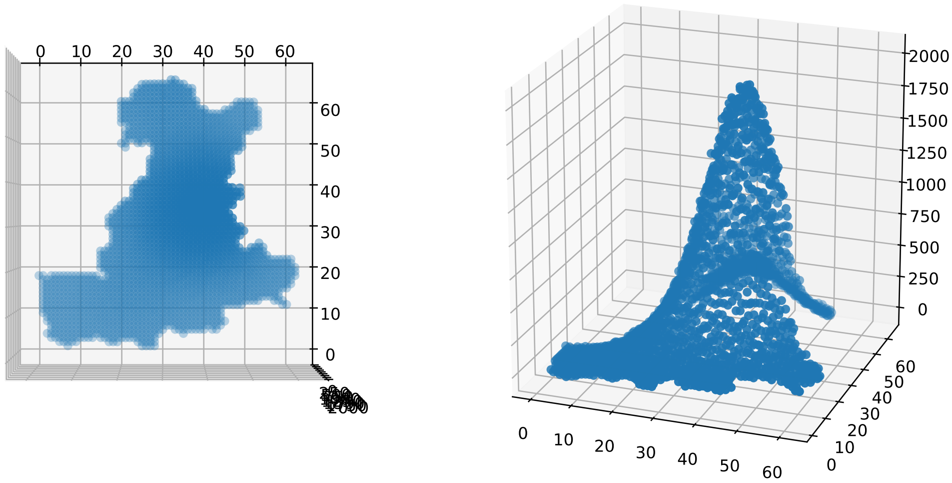
4.2.3 Solving the model for ULSS 6

Now that we have the population living in a service point and we set the area of a square, we can solve the model using the results of Equation (4.2.2) as the demand of each point. The interpretation of the output solution related to the x_{ij} variables will be the fraction of ambulances needed to serve i according to (4.2.2).

Before looking briefly at the solutions, we can do some preliminary considerations. First of all, we have that in ULSS 6 there aren't mountainous areas, hence the results from (4.2.1) and (4.2.2) are the same. Secondly, Lemma 4.1.1 gives us a lower bound of the total number of ambulances, hence, considering a population of 935000, a total area of $2125\,\text{km}^2$ and applying the AGENAS formula, we have that the minimum number of ambulances needed is 10.83, hence from the lemma the minimum number of ambulances is 11.

Since the population generation is a random process we should report a multitude of results but from time to time the value of the y variables stayed the same, while the x_{ij} vary and in particular we noted that also the points served by two hospital can change. Nevertheless, we report explicitly only the y variables.

In Table 4.2 we can see the solution in terms of number of ambulances for each hospital. In Figure 4.7 we can see the complete solution of the model where each edge, in shades of blue, represents a x_{ij} variable and its color represents the value according to the colorbar on the right. As expected, the number of ambulances of y_3 is much greater than the others because there is the center of the city of Padova; moreover, there are some points that are shared between two hospitals, but they are few, as already stated.



(a) The horizontal plane of ULSS 6: we can see the more density in the center right region

(b) 3D-plot of the population distribution: we can see basically a single bivariate normal distribution

Figure 4.6: The random population distribution of ULSS 6

Hospital	**Ambulances**	**Coord. in the interpolated image**
y_1	1	(29, 62)
y_2	1	(42, 52)
y_3	5	(38, 33)
y_4	1	(51, 24)
y_5	1	(5, 16)
y_6	1	(28, 15)
y_7	1	(37, 15)
Total	11	

Table 4.2: The minimum number of ambulances for ULSS 6

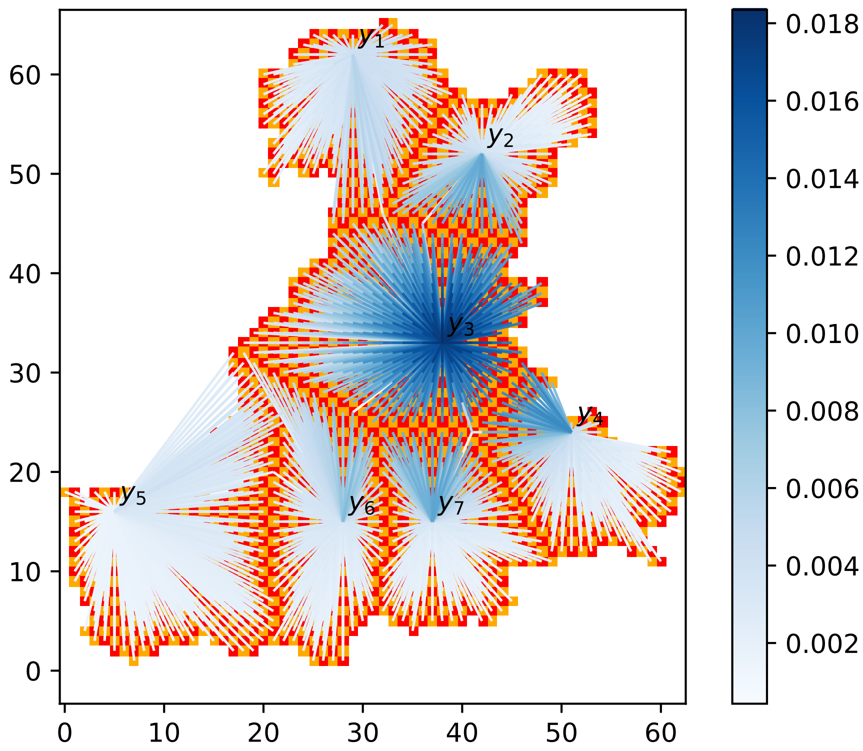


Figure 4.7: The plot of the solution for ULSS 6: the color of an edge corresponds to the value of the respective x_{ij} variable according to the colorbar on the right

4.3 A case study: ULSS 8 “Berica”

In Sections 4.1 and 4.2 we explained the tools used to solve the model showing time by time examples and analyzing the results having applied a formula developed independently from the optimisation model. In this section the goal is to test another approach and compare the results with the ones obtained using Equations 4.2.1 and 4.2.2. The ULSS considered is ULSS 8 “Berica”, as already mentioned, it is one of the public agencies of Veneto and for this reason, the territory and the management is under the control of Azienda Zero.

4.3.1 The solutions using AGENAS formula

Following the same path as in the previous sections, let’s have a look at the solution derived from the known formula. First of all, we can have a look to the number of ambulances provided by Equation 4.2.2 without applying the model; this data is reported in Table 4.3. Two things to notice: the first one is that in this case

Hospital	**PLL**	**PLM**	**LA**	**MA**	**Number of ALS vehicles**	**Ceiling value**
Arzignano	97775	2797	195	13	1.15	2
Vicenza	267681	0	469	0	2.90	3
Noventa Vicentina	34805	0	147	0	0.50	1
Lonigo	48045	0	237	0	0.74	1
Valdagno	12080	36247	24	155	0.85	1
Total	460386	39044	1071	169	6.14	8

Table 4.3: The requirement of ALS vehicles according to 4.2.1

we have nonzero values for the mountainous area and the population living there, a piece of information that will be lost in the model; the second one is that the formula returns generally a floating point, but when we have to make it integer we don't have any information that the rounding function returns the best value because, if the floor function can be applied, we do not have any assurance that number is sufficient to satisfy the demand, so we are forced to use the ceiling function (column 7 of the table).

Let's see now the results obtained with the tools developed. From the previous table we can see that the total number of vehicles is 6.14, so Lemma 4.1.1 tells us that the lower bound of the number of ambulances is 7. In Figure 4.8 we can see the grid of demand points and the hospitals. We can notice that with a preset coverage standard $r = 18\,\text{km}$, not all the demand points are covered, in particular few points in the upper right area are not covered, in fact the minimum preset coverage standard should be $r = 21\,\text{km}$. For the population, we considered the 10 most populated cities, specified in the Table 4.4 and an example of the random generation of the population can be seen in Figure 4.9. In 4.9a, 4.9b and 4.9c we decided to report the orthogonal projections because, unlike ULSS 6 example, here the contribution of the other cities, which are not the most populated, is not negligible and we can see a distribution which we cannot approximate basically with a single bivariate normal distribution.

The solution of the model applying the Equation 4.2.2 are shown in Table 4.5 and Figure 4.10. Anyway, a comparison between the two solutions found with and without the model is quite difficult, because in Table 4.3 there is the non negligible contribution coming from the population and the area of the mountain. However if from these data, we sum the population living in lowland and mountain and the relative areas and then we apply Equation 4.2.2, the number of ambulance

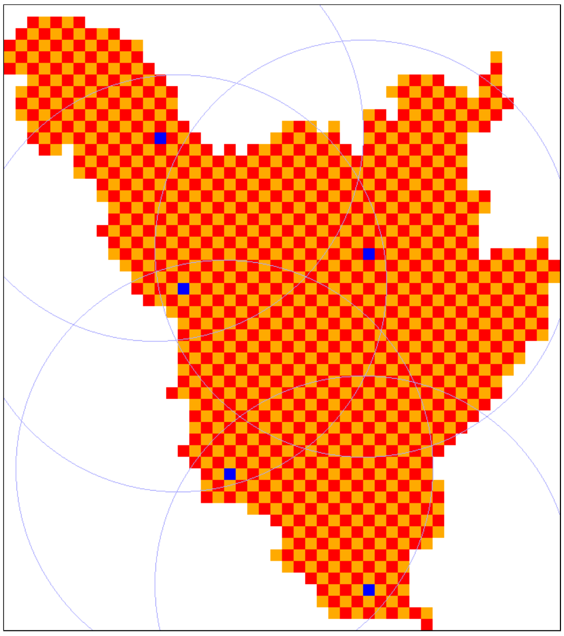


Figure 4.8: The coverage of the hospitals in ULSS 8 with $r = 18\,\mathrm{km}$

resulting is exactly the same, hence with the model seven ambulances are sufficient with respect to the eight of the standalone formula.

4.3.2 The solution using another approach

So far we have relied to an equation for the demand of the points and we know the principle which led to that formulation, but the details of how this formula was derived are not known, hence we don't know what specific situations and assumptions were made; moreover it comes from a study proposed in 2011, therefore in light of all this, the formula may be obsolete or inappropriate for the situation we are considering. These reasons brought us to take another path with other data and in particular the idea is focused in the emergency calls.

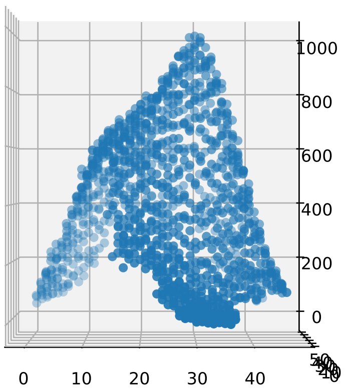
From the database of Azienda Zero, we extracted the accesses made with ambulances to emergency rooms for each month from each city of ULSS 8 in 2018; these data are shown in Table 4.7 on page 54. These are not the data relative to the number of ambulance interventions; hence we know only how many people were brought to emergency rooms and the city where they live. This means that if an ambulance was called but it returned empty because the patient did not need to be transported to the hospital, that call was not added to the data. Note that we don't know the hospital where the patients were brought; for some cities the data of the accesses of some months were not available, moreover the database

City	**Population**	**Coord. in the interpolated image**
Vicenza	112443	(31, 33)
Valdagno	26425	(11, 43)
Arzignano	26176	(15, 30)
Montecchio Maggiore	24004	(21, 28)
Lonigo	16722	(19, 14)
Dueville	13922	(32, 41)
Chiampo	12938	(11, 31)
Altavilla Vicentina	11976	(24, 27)
Cornedo Vicentino	11968	(14, 40)
Torri di Quartesolo	11831	(38, 30)
Top-10 total population	268405	
Total Population of ULSS 8	50000	

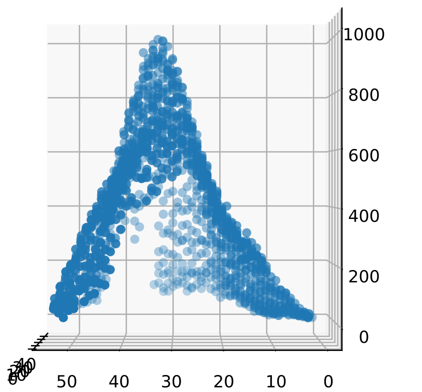
Table 4.4: The top-10 most populated cities of ULSS 8

Hospital	**Variable**	**Ambulances**	**Coord. in the interpolated image**
Valdagno	y_1	1	(13, 43)
Arzignano	y_2	1	(15, 30)
Vicenza	y_3	3	(31, 33)
Lonigo	y_4	1	(19, 14)
Noventa Vicentina	y_5	1	(31, 4)
Total		7	

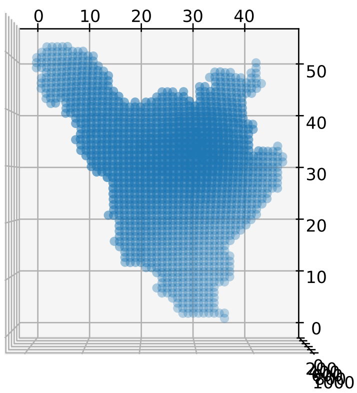
Table 4.5: The minimum number of ambulances of ULSS 8 applying (4.2.2)



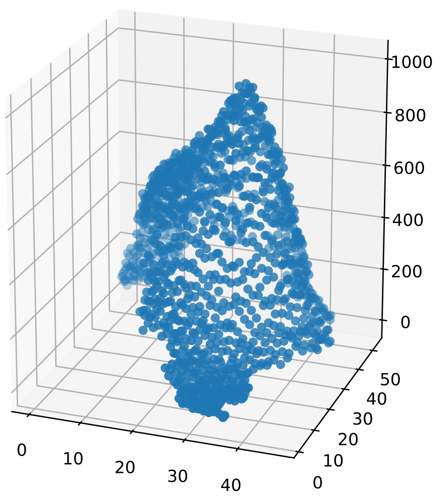
(a) The vertical plane



(b) The lateral plane



(c) The horizontal plane



(d) 3D-plot of the population distribution

Figure 4.9: The random population distribution of ULSS 8

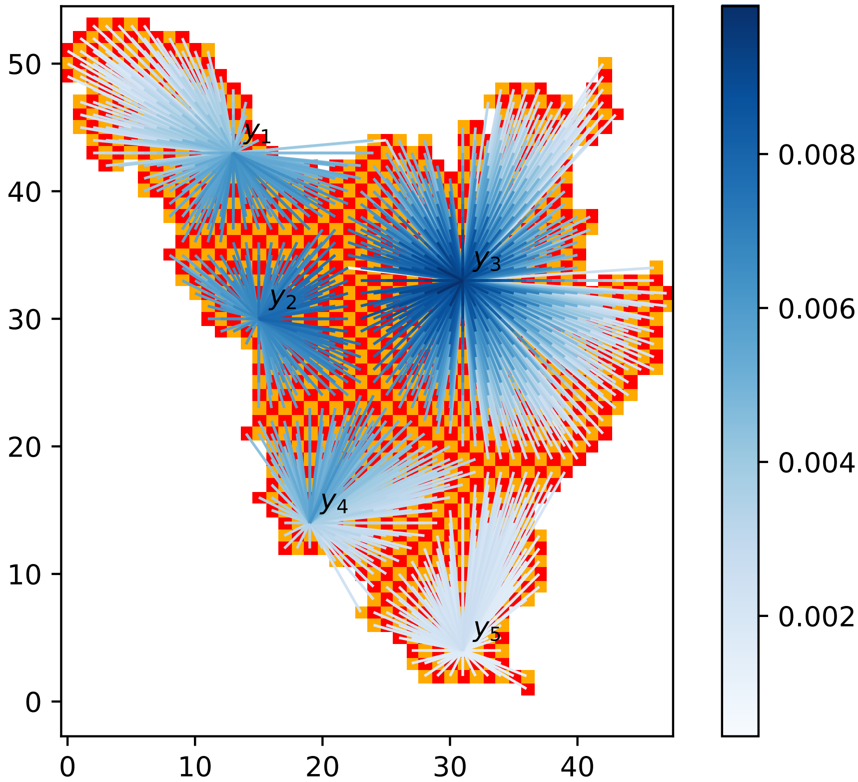


Figure 4.10: The plot of the solution for ULSS 8 applying (4.2.2)

underlined unreported accesses in emergency rooms, where the city was not specified, which are shown in the penultimate row of the table under the voice "other sources".

For the above reasons we decided to use as number of calls the sum of the maximum number of accesses from each city and then, since one assumption of our model is that the average travel time of an ambulance is one hour, divide this number for $24 \times 30 = 720$ that are the hours per month. This result is the hourly average of the maximum number of accesses in the emergency rooms and from now we consider that number the total number of requests from the demand points and the results are summarised in Table 4.8 on page 56. Now that we have the accesses, we could have drawn the boundaries for each city and distributed the requests, but it would have required an excessive effort. Hence, we decided to distribute uniformly the total requests proportional to the population of each node, namely the larger the population of a node, the more calls are made; so the demand of node i become:

$$d_i = \frac{P_i}{P_{tot}} \cdot d \tag{4.3.1}$$

where:

$$P_i = \text{population of node } i \in V$$

$$P_{tot} = \sum_{i \in V} P_i = \text{total population within the region}$$

$$d = \sum_{i \in V} d_i = \text{total demand of the region}$$

In the case of ULSS 8, $P_{tot} = 500000$ and $d = 4.249$. We know also, from Lemma 4.1.1, that this latter is the lower bound of the number of ambulances needed.

The solution of the model follows and it is shown in Table 4.6 and Figure 4.11. We tested this variant with a multitude of instances randomly generated, but as in the previous cases the number of ambulances did not change. Thus, considering all the situations seen until now, we can state empirically that the model is stable to little variations of the requests of the points. Comparing these results with the ones obtained before, we can note that the number of ambulances is decreased of one unit in the hospital of Vicenza. Another aspect that is visible from the Figure 4.10 and 4.11, is the points covered by each hospital since they stay the same apart from few points equally distant from the respective hospitals.

These results are quite impressive because in the beginning of the section, we started with a number of ambulance of eight applying the standalone AGENAS formula; then developing an optimisation model and using the same formula contextualizing it in the model, we were able to find a better solution and finally, with an independent argument based of the requests of the points we obtained an even better solution. Certainly, these results must be validated before applying them in a real context, because the assumptions made to simplify the writing of the model are strong. Unfortunately, we do not have detailed data to test the model in a real context and this can become a starting point for future studies. Finally, it must be said that the numbers obtained would not be definitive due to other constraints present in [19], as explained in Section 3.1.

4.3.3 A closer look: the choice of the weight and the points served by multiple hospitals

In the last part of this chapter we want to talk about two relevant topics that were not studied in depth:

- the choice of w,
- the demand points served by multiple hospitals in the optimal solution.

The first topic is quite crucial in the model and in Section 4.1 it was set to a fixed value and never edit. The considerations we are going to see holds for all the instances we have seen so far. Recalling what we have said: the weight w was inserted in the objective function to balance the contribution of the number of ambulances and the one from the multiplication between the x variables and the distances with hospitals. In this way the model will choose to minimize the number of ambulances as first target and secondly serve each point with the nearest hospital.

After solving some instances, we can see that in the objective function there is a gap, which order of magnitude is 10^2, between the y variables and the $wd(i,j)x_{ij}$ terms. This gap can be reduced to 10^1, namely set $w = 1$ and all the solutions will not change minimally, and for this reason, we do not report the graphs, but despite the potential choice, we decided to keep a greater gap if larger instances were considered.

A different situation comes if we change w in both direction, because if we increase its value, the model is not incline to minimise the number of ambulance at all, it minimises the second term and the number of ambulances becomes of the order of hundreds or more.

On the other hand, if we decrease the value the model will try to minimise the overall ambulance as much as possible. About that it is interesting to see the solution of ULSS 8 instance fixing $w = 10^{-2}$ or smaller, because the minimum number of ambulances needed to serve all points is five, applying the average

Hospital	**Variable**	**Ambulances**	**Coord. in the interpolated image**
Valdagno	y_1	1	(13, 43)
Arzignano	y_2	1	(15, 30)
Vicenza	y_3	2	(31, 33)
Lonigo	y_4	1	(19, 14)
Noventa Vicentina	y_5	1	(31, 4)
Total		6	

Table 4.6: The minimum number of ambulances of ULSS 8 applying (4.3.1)

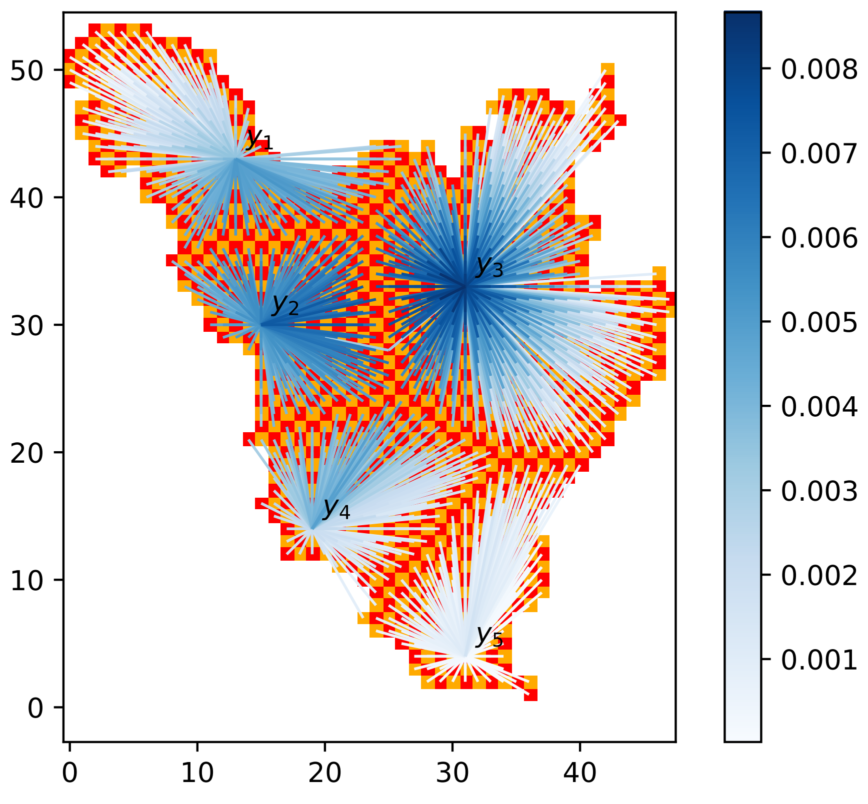


Figure 4.11: The plot of the solution for ULSS 8 applying (4.3.1)

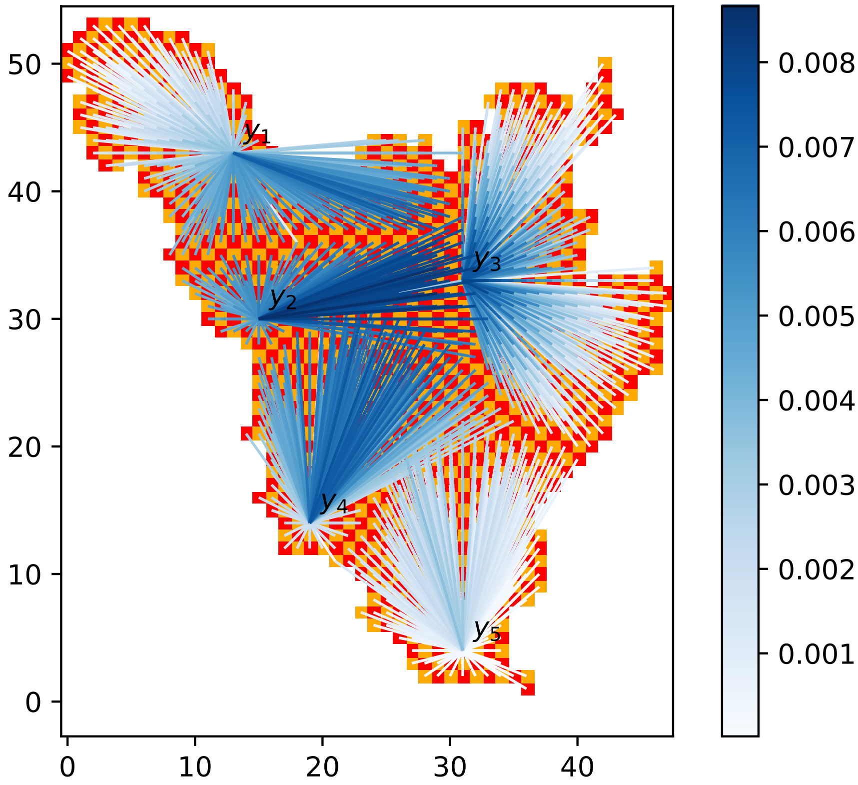


Figure 4.12: The plot of the solution for ULSS 8 applying (4.3.1) for small w

maximum number of accesses, one for each hospital, and the graph can be shown in Figure 4.12. This reflects the concepts explained in Section 4.1 and this solution, despite being better, violates the requirements of "satisfy the requests as fast as possible". Note also that the solution in Table 4.5, returned using AGENAS formula cannot be improved for Lemma 4.1.1. The main concept behind this retrospective is that, based on the instances and the requirements, one will choose a different value of w.

The second important topic concerns the demand points that are served by multiple hospitals in the optimal solution. This case is present in all the three optimal solutions seen in Sections 4.2 and 4.3. Conceptually it is correct and this is perfectly allowed by the requirements. The problem comes if one of these points generate a request, because we did not define any policy related to the hospital from which an ambulance must be deployed. In this context many different choices can be made: for example, a choice can be to pick up randomly a hospital defining a probability distribution, another one can be to send an ambulance from the hospital that hosts the larger number of vehicles and so forth and so on.

Now the question is which choice could be the best: even in this case one option can be better than another based on the requirements of the problem. In our case we have to recall that the main goals are the minimisation of the ambulances and the rapidity of response. A reasonable approach derived directly from the model is to send an ambulance according to a probability which is proportional to the value of the x_{ij} variables serving that point. Another choice in line with this policy is to send an ambulance from the nearer hospital and if the distance is the same, send from the hospital which host the highest number of vehicles. However, these options can fail because it is true that the response is the quickest, but we have to account that other requests can be generated. Hence the option we want to propose and analyse is the following: send the ambulance from the hospital that serves those points that have the lowest probability to generate a number of requests greater than the number of ambulances hosted by such hospital.

More formally, starting from an optimal solution returned by the model, suppose the point $k \in V$ to be served by more than one hospital; let's define the new quantities:

$$R_i = \text{the number of requests generated from point } i \in V$$

$$H_i = \text{the set of hospitals that serve } i \in V$$

$$C_j = \text{the set of demand points } i \in V \text{ served only by hospital } j$$

$$S_j = \sum_{i \in C_j} R_i$$

Hence $|H_k| \geq 2$, R_i and S_j are two random variables and clearly S_j is the sum of all requests generated by the points served only by hospital j. Then the choice of the hospital is:

$$\begin{aligned} k &= \arg\min_{j \in H_i} \left\{ \Pr\left[S_j < y_j \right] \right\} \\ &= \arg\min_{j \in H_i} \left\{ \Pr\left[\sum_{i \in C_j} R_i < y_j \right] \right\} \end{aligned}$$

Now this equation can be expanded defining the distribution functions and trying to derive the probability of the sum of random variables, but the definition of the probability functions is outside the scope of this work and we are not claiming that one the particular options we presented to address the problem is the best; with this insight we just want to discuss about an aspect coming from the output of the model and some tools to analyse it.

City	Number of accesses per month to hospitals												Total
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	
Agugliaro	2	2	8	4	4	3	3	4	5	3	3	4	45
Albettone	9	4	8	5	5	8	5	5	7	11	9	9	85
Alonte	13	12	7	14	8	9	1	6	1	4	8	10	93
Altavilla Vicentina	32	28	32	25	47	49	40	27	36	38	31	33	418
Altissimo	8	7	6	6	7	5	5	8	12	3	6	7	80
Arcugnano	20	24	21	17	34	29	23	23	22	18	20	28	279
Arzignano	127	114	109	102	102	108	111	106	110	115	100	99	1303
Asigliano Veneto	2	1	1	10	1	3	3	-	2	1	4	-	28
Barbarano Vicentino	26	15	19	20	-	-	-	-	-	-	-	-	80
Bolzano Vicentino	11	13	13	17	23	15	18	18	16	23	22	18	207
Brendola	21	33	20	28	19	36	29	30	30	27	38	38	349
Bressanvido	7	7	5	3	14	8	7	7	4	5	8	4	79
Brogliano	18	12	17	23	34	18	25	14	11	29	21	16	238
Caldogno	35	21	41	28	40	43	40	29	26	44	28	32	407
Camisano Vicentino	22	18	23	30	43	33	38	31	28	28	36	27	357
Campiglia Dei Berici	7	10	4	2	11	4	7	6	10	6	6	8	81
Castegnero	10	8	6	6	7	7	8	11	7	8	16	5	99
Castelgomberto	30	30	32	38	24	27	27	37	24	31	25	29	354
Chiampo	56	53	43	44	51	48	46	52	52	43	61	34	583
Cornedo Vicentino	72	67	73	60	58	72	80	55	76	81	60	73	827
Costabissara	25	17	17	22	13	20	13	12	19	15	24	15	212
Creazzo	37	45	35	32	43	35	50	41	41	35	39	43	476

Continued on next page

City	Number of accesses per month to hospitals												Total
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	
Crespadoro	3	5	3	7	6	3	6	4	5	3	3	3	51
Dueville	48	38	46	28	61	46	48	45	44	52	49	49	554
Gambellara	21	13	21	22	15	12	17	20	20	16	15	12	204
Gambugliano	1	5	7	1	3	3	5	5	7	2	4	7	50
Grisignano Di Zocco	12	17	15	10	15	22	14	19	15	15	12	16	182
Grumolo Delle Abbadesse	10	5	12	8	9	11	16	16	15	7	16	13	138
Isola Vicentina	27	31	26	37	33	32	42	36	29	31	24	38	386
Longare	26	18	14	15	26	24	22	23	16	25	28	18	255
Lonigo	109	76	107	110	101	82	90	103	103	114	80	101	1176
Montebello Vicentino	28	28	25	23	29	25	34	28	34	35	23	27	339
Montecchio Maggiore	109	96	101	106	102	109	102	89	113	114	116	114	1271
Montegalda	13	6	11	10	13	18	12	21	10	11	4	16	145
Montegaldella	5	10	2	3	2	6	1	4	3	2	-	7	45
Monteviale	12	4	9	5	6	6	11	9	6	8	8	8	92
Monticello Conte Otto	25	29	19	30	31	19	33	24	24	40	34	34	342
Montorso Vicentino	14	9	3	13	7	13	10	8	8	13	12	6	116
Mossano	15	6	13	4	-	-	-	-	-	-	-	-	38
Nanto	10	13	10	9	7	8	11	5	5	12	3	13	106
Nogarole Vicentino	3	-	1	3	4	8	8	5	2	2	3	3	42
Noventa Vicentina	46	28	46	25	39	42	43	38	44	39	30	42	462
Orgiano	22	15	13	8	8	9	15	14	19	16	16	14	169
Pojana Maggiore	21	18	22	17	25	21	25	28	20	16	21	25	259

Continued on next page

City	Number of accesses per month to hospitals												Total
	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	
Pozzoleone	12	1	5	3	3	10	5	6	7	9	3	3	67
Quinto Vicentino	22	15	18	8	21	19	16	4	19	16	17	14	189
Recoaro Terme	52	42	51	58	53	53	47	45	43	48	45	44	581
San Pietro Mussolino	3	3	7	4	4	6	2	2	5	2	5	8	51
Sandrigo	28	22	18	35	34	20	29	24	21	27	31	38	327
Sarego	55	39	34	40	39	37	32	36	34	50	27	31	454
Sossano	32	30	20	16	22	18	24	21	30	23	24	34	294
Sovizzo	19	16	18	14	25	19	28	22	13	19	14	27	234
Torri Di Quartesolo	38	44	31	39	36	31	38	44	45	42	45	44	477
Trissino	46	41	36	46	28	44	37	39	34	54	40	37	482
Val Liona	14	10	21	7	12	4	7	7	9	12	4	13	120
Valdagno	267	152	212	231	211	210	196	187	186	218	191	200	2461
Vicenza	568	458	494	405	530	522	532	525	515	573	510	553	6185
Villaga	6	6	6	5	3	5	6	4	7	8	8	7	71
Zermeghedo	9	5	4	5	1	7	4	4	5	6	8	3	61
Zovencedo	6	-	-	1	5	3	3	3	1	3	3	2	30
Other Sources	171	150	175	189	234	253	248	252	260	228	201	216	2577
Total	2518	2045	2216	2136	2391	2360	2398	2291	2315	2479	2242	2372	27645

Table 4.7: Accesses made with ambulances to emergency rooms for each month from each city of ULSS 8 in 2018

City	**Max. monthly accesses**	**Avg. hourly accesses**
Agugliaro	8	0.011
Albettone	11	0.015
Alonte	14	0.019
Altavilla Vicentina	49	0.068
Altissimo	12	0.017
Arcugnano	34	0.047
Arzignano	127	0.176
Asigliano Veneto	10	0.014
Barbarano Vicentino	26	0.036
Bolzano Vicentino	23	0.032
Brendola	38	0.053
Bressanvido	14	0.019
Brogliano	34	0.047
Caldogno	44	0.061
Camisano Vicentino	43	0.060
Campiglia Dei Berici	11	0.015
Castegnero	16	0.022
Castelgomberto	38	0.053
Chiampo	61	0.085
Cornedo Vicentino	81	0.113
Costabissara	25	0.035
Creazzo	50	0.069
Crespadoro	7	0.010
Dueville	61	0.085
Gambellara	22	0.031
Gambugliano	7	0.010
Grisignano Di Zocco	22	0.031
Grumolo Delle Abbadesse	16	0.022
Isola Vicentina	42	0.058
Longare	28	0.039
Lonigo	114	0.158
Montebello Vicentino	35	0.049
Montecchio Maggiore	116	0.161
Montegalda	21	0.029

Continued on next page

City	**Max. monthly accesses**	**Avg. hourly accesses**
Montegaldella	10	0.014
Monteviale	12	0.017
Monticello Conte Otto	40	0.056
Montorso Vicentino	14	0.019
Mossano	15	0.021
Nanto	13	0.018
Nogarole Vicentino	8	0.011
Noventa Vicentina	46	0.064
Orgiano	22	0.031
Pojana Maggiore	28	0.039
Pozzoleone	12	0.017
Quinto Vicentino	22	0.031
Recoaro Terme	58	0.081
San Pietro Mussolino	8	0.011
Sandrigo	38	0.053
Sarego	55	0.076
Sossano	34	0.047
Sovizzo	28	0.039
Torri Di Quartesolo	45	0.063
Trissino	54	0.075
Val Liona	21	0.029
Valdagno	267	0.371
Vicenza	573	0.796
Villaga	8	0.011
Zermeghedo	9	0.013
Zovencedo	6	0.008
Other Sources	353	0.490
Total	3059	4.249

Table 4.8: Maximum accesses per month and relative average hourly accesses for ULSS 8

Chapter 5

Conclusion and future works

In this thesis we introduced, to the best of our knowledge, a new model to solve the problem of the minimisation of the number of ambulances in a given region, maintaining a certain level of service. Moreover, from this model we derived important results that provide a lower bound of the number of vehicles to deploy in a given region, which can be used as a first estimate.

Later we applied our model to a real instance, ULSS 8. We first showed the number of ambulances required if an independent formula is used, then we used that formula within our model to extract the demand of the points, and finally we used real data, concerning the accesses in the emergency rooms. The results show that, in this situation, the overall number of ambulances can be reduced by two units. Unfortunately we were not able to test the results in a real environment for lack of significant data, mainly because in the development phase, we introduced some assumptions to simplify the writing of the model and a validation of the results is needed to verify if the simplifications made are consistent. A positive note is that some assumptions can be easily modified without changing the structure of the model.

In this aspect, in the future some improvements can be implemented: first of all testing the results on real environments with different sizes, but also we would like to obtain the data of the actual requests from each hospital (this aspect can be used to improve the demand parameter of the model). Also, the average speed and the time it takes for a vehicle to fulfill a call can improve the preset coverage standard. In this direction there are more sophisticated improvements that can be implemented: for example, map the travel of a vehicle in the road network adding different travel times for each stretch. This would remove the assumption of reaching the demand points in a straight line. For what concerns the distribution of the requests, an improvement can follow from the recognition of the cities in a given region using the boundaries which can lead to a better population distribution and consequently of the demand distribution. As we explained, in the Italian EMS there are different types of vehicles used whereas in the model we considered only the ambulances. Considering different type of vehicles with, for example, different types of requests and coverage standards, can be another integration to the model.

To conclude we want to point out that we have not spoken about the running time of the optimisation software, even though we model the problem using a mixed integer linear programming model. This is because the number and the order of magnitude of the integer variables of our instances is low and the optimisation software solves such instances in tens of a seconds, and it only takes a few seconds for more complex instances specially built.

Bibliography

[1] National Highway Traffic Safety Administration. *What is EMS?* `https://www.ems.gov/whatisems.html`. [Online; accessed 16-January-2020].

[2] Michael O Ball and Feng L Lin. "A reliability model applied to emergency service vehicle location". In: *Operations research* 41.1 (1993), pp. 18–36.

[3] G. Bradski. "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[4] Luce Brotcorne, Gilbert Laporte, and Frederic Semet. "Ambulance location and relocation models". In: *European journal of operational research* 147.3 (2003), pp. 451–463.

[5] Richard Church and Charles ReVelle. "The maximal covering location problem". In: *Papers of the regional science association*. Vol. 32. 1. Springer-Verlag. 1974, pp. 101–118.

[6] Mark S Daskin. "A maximum expected covering location model: formulation, properties and heuristic solution". In: *Transportation science* 17.1 (1983), pp. 48–70.

[7] Mark S Daskin and Edmund H Stern. "A hierarchical objective set covering model for emergency medical service vehicle deployment". In: *Transportation Science* 15.2 (1981), pp. 137–152.

[8] David J Eaton et al. "Determining emergency medical service vehicle deployment in Austin, Texas". In: *Interfaces* 15.1 (1985), pp. 96–108.

[9] Matteo Fischetti. *Introduction to Mathematical Optimization*. Independently Published, 2019. ISBN: 9781692792022. URL: `https://books.google.it/books?id=OsbhyQEACAAJ`.

[10] Python Software Foundation. *python*. `https://www.python.org/`. [Online; accessed 12-February-2020].

[11] Michel Gendreau, Gilbert Laporte, and Frédéric Semet. "A dynamic model and parallel tabu search heuristic for real-time ambulance relocation". In: *Parallel computing* 27.12 (2001), pp. 1641–1653.

[12] Kathleen Hogan and Charles ReVelle. "Concepts and applications of backup coverage". In: *Management science* 32.11 (1986), pp. 1434–1444.

[13] IBM. *IBM ILOG CPLEX Optimization Studio*. https://www.ibm.com/it-it/products/ilog-cplex-optimization-studio. [Online; accessed 12-February-2020].

[14] Richard C Larson. "A hypercube queuing model for facility location and redistricting in urban emergency services". In: *Computers & Operations Research* 1.1 (1974), pp. 67–95.

[15] Richard C Larson. "Approximating the performance of urban emergency service systems". In: *Operations research* 23.5 (1975), pp. 845–868.

[16] Mattia Micheletto. "The health care system of Veneto region: expenditure, funding and effects of the 2016 reform". MA thesis. University of Padova, 2018, pp. 21–22.

[17] Gazzetta Ufficiale n. 126, ed. *Comunicato relativo al Decreto del Presidente della Repubblica 27 marzo 1992*. https://www.gazzettaufficiale.it/eli/id/1992/05/30/092A2542/sg. [Online; accessed 17-January-2020]. May 1992.

[18] Hari K Rajagopalan, Cem Saydam, and Jing Xiao. "A multiperiod set covering location model for dynamic redeployment of ambulances". In: *Computers & Operations Research* 35.3 (2008), pp. 814–826.

[19] Giunta Regionale. *Deliberazione della Giunta Regionale n. 1515 del 29 ottobre 2015*. https://bur.regione.veneto.it/BurvServices/pubblica/DettaglioDgr.aspx?id=310187. [Online; accessed 10-March-2020]. Oct. 2015.

[20] Giunta Regionale. *Deliberazione della Giunta Regionale n. 248 del 08 marzo 2019, Allegato A*. https://bur.regione.veneto.it/BurvServices/pubblica/Download.aspx?name=248_AllegatoA_389900.pdf&type=9&storico=False. [Online; accessed 13-February-2020]. Mar. 2019.

[21] Giunta Regionale. *Legge Regionale n. 19 del 25 ottobre 2016*. https://bur.regione.veneto.it/BurvServices/Pubblica/DettaglioLegge.aspx?id=332095. [Online; accessed 9-March-2020]. Oct. 2016.

[22] Agenzia Nazionale per i servizi sanitari Regionali. *Monitor n. 27 - Agenas*. https://www.agenas.gov.it/archivio-monitor-2011/monitor-n-27. [Online; accessed 15-February-2020]. Mar. 2011.

[23] Agenzia Nazionale per i servizi sanitari Regionali. *Who we are - Agenas*. https://www.agenas.gov.it/contenuto-inglese. [Online; accessed 15-February-2020].

[24] John F Repede and John J Bernardo. “Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky”. In: *European journal of operational research* 75.3 (1994), pp. 567–581.

[25] Charles ReVelle and Kathleen Hogan. “The maximum availability location problem”. In: *Transportation science* 23.3 (1989), pp. 192–200.

[26] David Schilling et al. “The TEAM/FLEET models for simultaneous facility and equipment siting”. In: *Transportation science* 13.2 (1979), pp. 163–175.

[27] The GIMP Development Team. *GIMP*. Version 2.10.12. June 12, 2019. URL: `https://www.gimp.org`.

[28] Constantine Toregas et al. “The location of emergency service facilities”. In: *Operations research* 19.6 (1971), pp. 1363–1373.

[29] Gazzetta Ufficiale, ed. *Decreto legislativo 30 dicembre 1992, n. 502*. `https://www.gazzettaufficiale.it/eli/id/1994/01/07/094A0049/sg`. [Online; accessed 10-March-2020]. Dec. 1992.

[30] Gazzetta Ufficiale, ed. *Decreto legislativo 7 dicembre 1993, n. 517*. `https://www.gazzettaufficiale.it/eli/id/1993/12/15/093G0599/sg`. [Online; accessed 10-March-2020]. Dec. 1993.

[31] Gazzetta Ufficiale, ed. *Legge 23 dicembre 1978, n. 833*. `https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=1978-12-28&atto.codiceRedazionale=078U0833&elenco30giorni=false`. [Online; accessed 10-March-2020]. Dec. 1978.

[32] Wikipedia. *Servizio Sanitario di Urgenza ed Emergenza in Italia — Wikipedia, L'enciclopedia libera*. [Online; accessed 23-January-2020]. 2020. URL: `%5Curl%7B//it.wikipedia.org/w/index.php?title=Servizio_Sanitario_di_Urgenza_ed_Emergenza_in_Italia&oldid=110294867%7D`.

[33] Wikipedia contributors. *Emergency medical services — Wikipedia, The Free Encyclopedia*. `https://en.wikipedia.org/w/index.php?title=Emergency_medical_services&oldid=935866422`. [Online; accessed 16-January-2020]. 2020.

[34] Wikipedia contributors. *Emergency medical services in Italy — Wikipedia, The Free Encyclopedia*. `https://en.wikipedia.org/w/index.php?title=Emergency_medical_services_in_Italy&oldid=917374923`. [Online; accessed 16-January-2020]. 2019.

[35] Wikipedia contributors. *Nontransporting EMS vehicle — Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Nontransporting_EMS_vehicle&oldid=935104959. [Online; accessed 23-January-2020]. 2020.

[36] Wikipedia contributors. *Set cover problem — Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Set_cover_problem&oldid=925501244. [Online; accessed 19-January-2020]. 2019.

[37] Ahmed S Zaki, Hsing Kenneth Cheng, and Barnett R Parker. "A simulation model for the analysis and management of an emergency service system". In: *Socio-Economic Planning Sciences* 31.3 (1997), pp. 173–189.