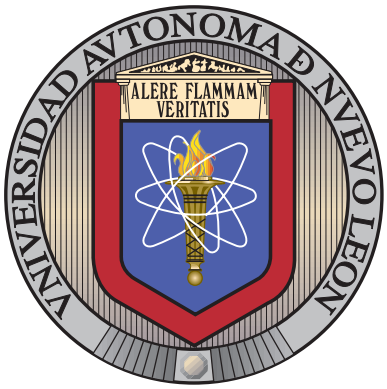
Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado



Stochastic methodologies for locating and dispatching two types of ambulances with partial coverage

por

Beatriz Alejandra García Ramos

como requisito parcial para obtener el grado de

Doctorado en Ciencias en Ingeniería de Sistemas

2024

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado

Stochastic methodologies for locating and dispatching two types of ambulances with partial coverage

por

Beatriz Alejandra García Ramos

como requisito parcial para obtener el grado de

DOCTORADO EN CIENCIAS EN INGENIERÍA DE SISTEMAS

2024

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica
Subdirección de Estudios de Posgrado

Los miembros del Comité de Tesis recomendamos que la Tesis *Stochastic methodologies for locating and dispatching two types of ambulances with partial coverage*, realizada por el alumno Beatriz Alejandra García Ramos, con número de matrícula 1550385, sea aceptada para su defensa como requisito parcial para obtener el grado de Doctorado en Ciencias en Ingeniería de Sistemas.

El Comité de Tesis

Dr. Roger Z. Ríos Mercado
Asesor

Dra. Yasmín Á. Ríos Solis
Co-Asesor

Dra. Iris Abril Martínez Salazar
Revisor

Dra. María Angélica Salazar Aguilar
Revisor

Dr. Vincent André Lionel Boyer
Revisor

Dra. Yajaira Cardona Valdés
Revisor

Dra. Irma Delia García Calvillo
Revisor

Vo. Bo.

Subdirector de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León, 2024

Contents

List of Figures

List of Tables

Acknowledgment

Agradezco a Dios principalmente porque durante mi vida ha puesto en mi camino a las personas correctas y me ha regalado los momentos correctos para mejorar personal y académicamente. Agradezco que me permita tener salud y que me demuestre su amor directamente y a través de las personas que están a mi alrededor.

Agradezco a mis padres Beatriz Ramos y Jesús García que en cada etapa de mi vida me han apoyado para seguir creciendo como mujer y como profesionista. Gracias por su acompañamiento en mi camino, por sus consejos, su amor y su protección, y por siempre preocuparse por el bienestar de mi hermana y mío para que tengamos una gran calidad de vida y valoremos cada instante.

Agradezco a mi hermana Karina García que ha sido mi compañera de desvelos y distracciones y que me ha impulsado a recordar que la vida no es cien por ciento el trabajo si no que también hay que disfrutar de los buenos momentos en compañía de la familia y los amigos.

Agradezco a mis amigos desde maestría Alberto, Delavy, Gabriela, Mayra, Alan, Astrid, Citlali, Pablo y Yessica, quienes han sido parte importante de mi vida. Ellos han sido de gran apoyo cuando necesito hablar sobre mis problemas académicos porque son quienes más me entienden, y además de eso siempre están para mí en las buenas y en las malas. Agradezco también a mis amigos Obed, Arturo, Paola y Airy, que me han acompañado en muchos años de mi vida y quienes son para mí como una familia.

Agradezco a la Facultad de Ingeniería Mecánica y Eléctrica (FIME) y a la Universidad Autónoma de Nuevo León (UANL) por ofrecer un gran programa académico en el Posgrado en Ciencias en Ingeniería de Sistemas en el cual pude concluir una etapa académica que me ayudará a superarme a mí misma y por la beca que se me proporcionó para llevar a cabo mis estudios. Agradezco también las instalaciones que proporcionan a nuestro posgrado para que podamos tener un laboratorio en dónde trabajar y salones donde podamos tomar las materias que se nos imparten.

Agradezco a los Doctores que se encuentran en el Posgrado en Ciencias en Ingeniería en Sistemas por ser parte de mi crecimiento académico y por darme sus enseñanzas sobre lo que un investigador y un docente puede llegar a ser. En particular agradezco a los miembros de mi comité de tesis, por haber leído mi tesis y haberme apoyado en las correcciones de la misma. De manera especial agradezco a mi asesor Roger Z. Ríos Mercado y a mi coasesora Yasmín A. Ríos Solís que durante toda mi investigación estuvieron apoyándome y dirigiendo la misma.

Agradezco al Consejo Nacional de Humanidades Ciencias y Tecnología (CONAHCyT) por la beca de manutención que me fue otorgada con la cual pude facilitarme mis estudios al utilizar ese ingreso en transporte, comida e incluso en la investigación que durante cuatro años estuve realizando.

Abstract

Beatriz Alejandra García Ramos.

Candidato para obtener el grado de Doctorado en Ciencias en Ingeniería de Sistemas.

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: Stochastic methodologies for locating and dispatching two types of ambulances with partial coverage.

Número de páginas: 51.

The thesis aims to study the Emergency Medical Service (EMS) systems problems and implement algorithms to improve them. The solution's methodology proposed is to determine ambulance location and dispatching based on scenarios. These scenarios show how the system is, i. e. if a demand point, which is a place where a patient could need attention, have to be served or not by an ambulance or more than one ambulance. In this investigation, we study a finite number of scenarios to determine where to locate ambulances and how to dispatch them to demand points according to the system.

The study method analyzes integer stochastic models to adapt some ideas for a practical solution. We are interested in improving a particular Mexico's EMS system, which is different from the first world's EMS systems. These differences lead us not to be able to use mathematical models as we find them in the literature; nevertheless, we can build an integer stochastic model based on combining ideas proposed before and new concepts from us.

One of the contributions is to introduce partial rate coverage to this type of model. Commonly partial coverage is used in deterministic models due to its simplicity. Another contribution is to propose and intelligent feedback to solve the ambulance location and used it as an input for the stochastic programming model proposed.

The objective is to improve Mexico's 911 system by locating and dispatching ambulance to maximize patient attention at the minimum response time possible.

Firma del director: _____
Dr. Roger Z. Ríos Mercado

Chapter 1

Introduction

Emergency Medical Services (EMS) systems provide medical care for people who suffer a medical incident. These systems control emergency calls' services received at the emergency number established for emergencies, commonly the 9-1-1. These systems have two phases. The first phase is the response to an emergency call: an operator responds to the call and identifies the emergency type, such as medical emergencies, security emergencies, and fire emergencies. The operator asks some questions to identify the type of emergency (dismissing prank calls). If the patient needs medical care, the operator contacts an ambulance (commonly the nearest) and asks for attention at the emergency scene. The second phase is the response of an ambulance: paramedics prepare to go to the emergency scene, the ambulance is ready with material resources needed to attend to the patient, ambulance leaves its base, arrives to scene, treat the patient, leaves the scene, and arrive at a hospital (commonly the nearest) if it is necessary, and finally return to its base to wait for another emergency call.

EMS systems have significantly impacted operational research and medical investigations in the last decades. Scientists are concerned about the impact of calls emergency' average response time for attending a patient who suffers a medical incident. Moreover, the cost of buying material resources, medical vehicles, or building a new medical center, among other things, can limit the service to patients. Not having human and material resources can cause deficient patients' attention.

The most studied problem is reducing the average response time when an emergency call arrives at a call center and someone needs medical attention. The objective is to provide, as soon as possible, the initial treatment for a patient that has a medical problem caused by an accident, a trauma, or a natural disaster to reduce patients' mortality. For short response time, more likely people survive. Another objective that EMS system problems consider is to maximize coverage to attend

all emergency calls that enter the system. Also, some problems exist that consider improving the patients' survival or reducing the patients' mortality.

There exist different types of problems. Various focus on statics systems, where the decisions are fixed after they are taken. Others are dynamic [13], where decisions change throughout time passing or after a change on the system, for example, when a call is received. Many investigations focus their problems on solving optimal locations of their ambulances to improve the system. In contrast, others want to obtain optimal policies to decide which ambulance or ambulances are the best to attend the emergency calls received.

EMS systems have different strategic, tactical, and operational types of planning [15]. Strategic planning focuses on long-term decisions, such as fixed potential locations or the acquisition of resources [1]. Tactical plannings made decisions for a middle time, such as locating ambulances at potential sites or planning which is the best option to dispatch an ambulance for all emergency calls types. Finally, operational planning takes decisions in a short time. These decisions are made frequently when a call enters the system and are divided into online and offline decisions. Online decisions are taken over the full service of a call, and offline decisions are taken when a call is received following the plannings made at tactical plannings, for example, to decide which ambulance will send to attend to the patient [17].

Our interest is in the EMS systems of Mexico. In Mexico exists the 9-1-1 number controled by the C-5 organization (Centro de Coordinación Integral, de Control, Comando, Comunicaciones y Cómputo del Estado), which receives emergency calls. Some calls are for medical emergencies, others for police emergencies, and others for fire emergencies. When a call enters the system, and an operator decides that it is a medical emergency, the operator has to determine if it is necessary to send an ambulance or not. Also, a doctor can continue the call to guide the person on the phone if the patient needs immediate attention while the ambulance arrives. Then the paramedics can attend to the patient and transfer the patient to a hospital.

We propose a two-stage stochastic programming model with recourses for ambulance location and dispatching, considering two service providers to obtain a coordinated EMS system to solve those problems. The following sections present the background investigations about EMS systems (Chapter 2) and the usually used models. We describe the problem and factors that affect the EMS system in Mexico (Chapter 3). Then, we solve the problem and define the model (Chapter **??**) used to do the experiments. Finally, we show conclusions (Chapter 6) that we obtain from experiments that we describe in the previous section.

1.1 Motivation

Our interest is to improve the Emergency Medical Service System in Mexico, particularly in Nuevo León. World Health Organization (WHO) establishes that it has to be four ambulances per one thousand habitants, which is not available in the states throughout our country.

Due to the lack of available ambulances, emergency calls are answered late. However, buying more ambulances so that there are a greater number of them available to distribute is not an option. Improving the distribution of ambulances, and locating and dispatching them in a better way, could improve the EMS systems.

1.2 Problem Description

We address a problem where we have to locate a limited number of two heterogeneous types of ambulances in different city points and dispatch them to the sites where accidents occur. Our problem considers uncertainty of the accident (demand) points. Our goal to maximize the total and partial coverage and the response time in which the patients receive medical first aids. We propose a two-stage quadratic stochastic program for this problem. In the first stage, the location of the limited number of two types of ambulances is decided. In the second stage, the dispatching of the ambulances to accidents is determined. This stochastic model allows partial coverage of the accidents by the ambulances based on a decay function. Given the model is intractable even for medium-sized instances, we propose a location-allocation methodology that relies on the solution of an auxiliary surrogate model, which is faster to solve. This location-allocation heuristic consists of two phases. In the location phase, the location of the ambulances is obtained by solving the surrogate model. Then, this information is the input for the allocation phase, where the original model is solved. Experimental results show the effectiveness and efficiency of this proposed approach, obtaining high-quality solutions in reasonable times.

1.3 Hypothesis

This investigation hypothesizes that we can model the Emergency Vehicle Covering and Planning problem as a stochastic programming model with resources based on different scenarios. These scenarios consider accident types in each demand point; many of them can help know what to do when a situation occurs in the 9-1-1 system.

The ambulance location and dispatching in the system are optimized.

1.4 Objectives

This investigation aims to improve an Emergency Medical Services System considering partial coverage. The main idea is to obtain an optimal ambulance location and optimal policies for ambulance dispatching. The system that we consider for the problem to solve includes different factors that affect the system. Those factors are:

- Various types of accidents and variations on maximal response times depending on accident types;
- Different ambulances types, which are ambulances for basic life support (BLS) and ambulances for advanced life support (ALS);
- And variation in demand points depending on the day of the week and the hour of the day, which can be considered making different scenarios.

The objective for solving the problem is to create a scenario-based stochastical programming model with resources considering more than one service provider involved in the system to attend incoming emergency calls.

Chapter 2

Background

Emergency Medical Services (EMS) systems provide basic but urgent in-situ medical care for people who suffer a medical incident and then transport patients to hospitals [4, 7, 28]. When scientists talk about EMS systems, many terms explain the problem. Two of these terms are demand points and potential sites. Demand points are sites where an emergency call is usually done. Commonly, there is a different demand for each point depending on the number of calls made within a period. Potential sites are places where a vehicle (ambulance) could be located if necessary to cover some demand points either statically or dynamically.

The first phase of an EMS is the response to an emergency call by an operator that identifies the emergency type: accident, medical, security, fire, etc. The second phase is dispatching one or several ambulances to the emergency scene to provide urgent medical care. Some emergency situations, such as a multiple-car accident, may involve several people; thus, more than one ambulance could be needed. Moreover, different types of ambulances may be required in an emergency: Basic Life Support (BLS), usually with two Emergency Medical Technicians (EMTs), and Advanced Life Support (ALS) units with an EMT, an advanced EMT, and one or two Paramedics. The third phase involves the paramedics' treatment of the patients and transporting them to a hospital [4].

EMS systems in developing countries, as is the case in Mexico, lack around 30-60%[1] of the number of ambulances suggested by the World Health Organization (WHO), which is at least four ambulances per 100,000 people [12]. For the Red Cross, an EMS operating with this small number of ambulances is considered similar to a war situation[1]. Thus, one of the main contributions of this work is to deal with the problem of deciding if an emergency will be totally or partially covered. Sadly, some emergencies may remain uncovered by an emergency unit.

[1] Anonymous interviews done by the authors.

Commonly, emergency vehicle planning problems' main objective is to reduce the average response time of a patient's initial treatment given by a paramedic in an emergency [2, 6, 31, 32]. Indeed, the quickness and the number of ambulances dispatched to the accidents are crucial. Each ambulance has a response time for travel from the potential site where it is located to the demand point where the patient will be cared for. Every minute of treatment delay in a cardiac patient reduces the survival probability by 24% [27].

There are many models to solve the problems of EMS systems divided into deterministic, probabilistic, and stochastic problems, which use different solution methods to solve them. The first problems that we studied are the statics.

2.1 Static models

These models are used to solve a system that only considers a particular point in time. When these models are used to solve EMS systems, it refers to allocating ambulances that will not be moved from the base.

There are two early models for statics problems: Location Set Covering Model (LSCM) and Maximal Covering Location Problem (MCLP), which are problems focused on covering the maximal demand points in the entire zone. However, over time these problems evolved according to the needs of the Emergency Medical Services, as will be defined below.

2.1.1 Deterministic models

Deterministic models were proposed to solve static problems because sometimes the emergency calls need to be attended for different vehicle types. Most of them are covered once, like the Backup Coverage Problem (BACOP) or the Double Standard Model (DSM), which use two different radii of coverage [19]. Alternative deterministic models are the tandem equipment allocation model (TEAM) or the facility-location equipment-emplacement technique (FLEET), which consider two types of vehicles (one for basic life support and another for advanced life support), or the fact that sometimes more than one ambulance has to be located on a potential site to maximize that a demand point is covered twice.

In the thousands was introduced by Berman et al. [9] a decay function to classify coverage as full, none, and partial coverage in a generalized MCLP model.

They added a weighted demand for each node covered, considering the distance between facilities and demand points. The objective aims to maximize the total demand weight covered by all facilities when a determined number of facilities are located.

A year later, [18] introduced partial coverage to the MCLP problem. This problem aims to maximize coverage level for all demand points deciding where to locate a certain number of facilities within the available potential sites. The model was based on a p-median formulation and classified coverage into three levels: totally covered, partially covered, and not covered. They defined a decay function monotone decreasing according to the distance between the facility and demand point for partial coverage. The distance between a facility and a demand point has to be less or equal to the maximum full coverage distance established to consider total coverage. Demand points are considered not covered for a facility if the distance between it and the demand point is greater or equal to a maximum partial coverage distance. To solve large-size problems, they used a Lagrangian relaxation.

A decade after, Wang et al. [34] used an extension of the MCLP Problem to maximize coverage for fire emergencies establishing a travel cost between potential sites and demand points. This extension considers a partial distance and quantity coverage for multi-type vehicles to locate and dispatch them. Partial distance is calculated with a decay function, which decreases according to the vehicle response time increase. Quantity coverage determines if an emergency is completely served or not, comparing the number of vehicles dispatched with the necessary quantity. For this problem, they have to consider demand priority to know where vehicles must be located and the patient's classification to decide how to dispatch them.

As an extension of DSM, Dibene et al. [14] created the Robust Double Standard Model. They added demand scenarios to the original DSM problem. These scenarios divide weeks into workdays and weekends, divided into four periods: night, morning, afternoon, and evening. They added eight scenarios applied to optimize the Red Cross Tijuana, Mexico system, increasing the coverage of demand points to more than 95% locating ambulances on different points of the city that are not the original bases.

For us, it is imperative to gather all this information for our project as it provides insights into the different accident coverage types and the improvement for ambulances location. A thorough understanding of these models and their effectiveness will enable us to optimize mainly our ambulance location strategies.

2.1.2 Probabilistic models

In the eighties, some researchers thought about problems involved with probabilities. One of these probabilities involved in EMS systems is the probability of an ambulance being busy responding to an emergency call. This probability is called the *busy fraction*. The maximum expected covering location problem (MEXCLP) uses this probability. An extension of this model is the TIMEXCLP which considers travel speed variations during the day. Another extension is the adjusted MEXCLP model (AMEXCLP), which considers different busy probabilities for each potential site to locate ambulances. All these models can use the hypercube queueing model to calculate the busy fraction [15].

Other models were proposed to maximize the coverage of the demand points with a probability α used to calculate the busy fraction; one of them is the maximal availability location problem I (MALPI), which considers the busy fraction is the same for all potential sites. Another model is the MALPII which uses the hypercube model to assume different busy fractions for each potential site.

There exist more probabilistic models created in the nineties. The first is an extended version of the LSCM called Rel-P; this version considers that more than one ambulance can be located at the same potential site, but each potential site has a probability to has ambulances that are available to respond to a call and consider the probability of the busy fraction too.

The second model is the two-tiered model (TTM), which consider two types of vehicles to allocate at potential sites (BLS and ALS) considering two different coverage radii and having an associated probability for the combination of how many ALS vehicles can be located at the radius A, how many ALS can be located at the radius B and how many BLS vehicles can be located at the same radius B for each demand point.

Laura Albert and Maria Mayorga researched the EMS systems of Hanover, Virginia. All these investigations about Hanover, Virginia, were applied to this county to obtain practical solutions, but all models can be used to any other EMS system changing data inputs.

The first research is focused on considering a new approach to calculate the response time threshold (RTT), a class of EMS performance measures [21]. The approach uses the patient survival rate considering that patients have a cardiac arrest and random response times that depend on the distance between demand points and potential sites instead of patient outcomes, which is most used. Then, they use these measures on a hypercube model to evaluate different RTTs needed

to input a model that considers fire stations and rescue stations to be potential sites where ambulances could be located distributed on Hanover's rural and urban areas. This model optimizes the location of ambulances on potential sites to maximize patient survival.

Later, Albert and Mayorga et al. used the performance measures as an input of the performance measure dispatching problem. According to survival patient rate, they used a Markov decision process that identifies the best and most robust RTT to maximize the covering level, prioritizing patient location. The research concludes that the optimal survival rate is obtained when the system has an eight minutes RTT [22]. However, this time for RTT does not apply to Hanover because of the number of the ambulance that they have, so they started a pilot program called quick-response vehicle to have more vehicles for patient attention obtaining a nine minutes RTT, these new vehicles are as ALS vehicles without transporting patients to the hospital, only attending patients at the scene and BLS ambulances transport patients if is necessary [23]. The idea of including these quick-response vehicles is to minimize the need to use ALS ambulances.

When talking about optimizing EMS systems, one can also speak about dispatching. Bandara et al. [6] considers demand priorities for the different emergency calls arriving at the system. The objective is to maximize the patient's survival probability when an ambulance is dispatched to demand points, calculating a reward for each dispatch. They used a Markov Decision Process model formulation to determine the optimal dispatching strategies for an EMS system.

[31] involves location and dispatching decisions for EMS vehicles in the same mathematical model with two focuses, minimizing the mean response time that takes since an emergency call is received and maximizing the expected coverage demand, using a continuous-time Markov process to balancing flow equations needed to control the busy fraction for each ambulance. Balancing these equations takes exponential time, and authors consider a genetic algorithm to obtain some solutions and combine them to create new solutions to reduce the computational time. This genetic algorithm was applied to Hanover, Virginia, and when they have mid-size problems, the nearest dispatch rule is the best solution. It can vary depending on the zone where it is applied.

[2] involve a simulation inputting an initial solution to decide if ambulances have to stay at the potential sites establishes when the mathematical model is solved or if some of them have to be moved to another potential site. To decide how to proceed, they used different day's period times when traffic in the city is changing on each week's days, which they called *scenarios*, to maximize the patient's survival.

Transitioning from probabilistic models to scenario-based models in the management of EMS systems is imperative, as scenarios provide a more robust framework for addressing uncertainty. Probabilistic models frequently require assumptions regarding the likelihood of various events, such as the busy fraction of ambulances, which can be challenging to estimate accurately and may not adequately capture real-world complexities. In contrast, scenario-based models allow for the incorporation of various demand and traffic conditions, enabling more realistic and flexible planning. By utilizing scenarios, researchers can ensure a more reliable and adaptive emergency response system, improving also dispatching decisions, as we can see in the next section.

2.1.3 Stochastic models

Recently, ambulance location, allocation, and dispatching problems involved uncertainty at demand points to have a more realistic model. This uncertainty is caused because it is impossible to know when the system will receive an emergency call.

In 2017, Boujemaa et al. [11] proposed a two-stage stochastic model with recourse. The model's first stage determines where to open ambulance stations with a fixed cost for opening them. For the second stage, allocation is determined considering the expected traveling cost from ambulance stations to demand points. A demand point is considered covered if an ambulance station is within a threshold value. And some important factors that they included are two different demand types: life-threatening calls and non-life-threatening calls; two ambulance types: ALS and BLS; and scenarios structured by two data for each demand point: number of life-threatening calls and number of non-life-threatening calls, respectively. This problem minimizes the ambulance location-allocation cost and is solved by a Sample Average Approximation (SAA) algorithm that allows computing lower and upper bounds for problem solutions and providing the corresponding optimality gaps.

Later, Bertsimas and Ng [10] implemented stochastic and robust formulation for ambulance deployment and dispatch for a problem constructed as a graph. These formulations were compared with MEXCLP and MALP problems and aimed to minimize the fraction of late-arrivals without requiring ambulances to be repositioned, sending to demand points the closest available ambulance, and maintaining a call at a queue if there are no ambulances available at the system. The demand has the problem's uncertainty, which was constructed by four demand types: single for each demand point, local for the demand point and the nearest points, regional for a region of the entire zone, and global for the whole area. They determined a deterministic equivalent model to solve the stochastic formulation, and for the robust

formulation, they did a column and constraint algorithm.

Recently, Yoon et al. [36] studied a two-stage stochastic problem for locating and dispatching two types of emergency vehicles: ALS and BLS. The first stage locates the ambulances at potential sites, while the second stage dispatch ambulances from places where they were located to demand points when a call arrives. The objective is to maximize the expected coverage considering a penalty when a call is not serviced. One difference from other problems is that the system manages multiple emergency call responses, divided into high priority and low priority calls. Any vehicle type can serve low priority calls. However, high-priority calls have two options for the service: the first option is that these calls can be responded to an ALS ambulance. The second option is that a nearby BLS ambulance can service the call first, followed by an ALS ambulance that is not necessarily closed. An SAA deterministic equivalent formulation solved this problem for small data, while a Branch-and-Benders-Cut Solution solved a large-scale problem. And they did another problem version considering non-transport vehicles which can attend patients without translating them to hospitals.

Some works propose stochastic programming models based on call-arrival scenarios as a bundle of calls, the total number of emergency calls in each demand node during a given period. As we do in this work, a two-stage stochastic program deploys the ambulances in the first stage and dispatches them to respond to demand in the second stage. Beraldi and Bruni [8] and Noyan [26] induce a reliability approach by using probabilistic constraints. Nickel et al. [25] minimize the total cost of locating the ambulances while assuring a minimum coverage level. By considering a bundle of calls, they address the volume of calls during a short period, such as the Friday night hours. Bertsimas and Ng [10] implemented stochastic and robust formulations for ambulance deployment and dispatch to minimize the fraction of late arrivals without requiring ambulances to be relocated, sending to demand points the closest available ambulance, and maintaining a call at a queue if there are no ambulances available at the system.

All the information gathered will be utilized collectively to formulate a novel problem that can integrate and utilize the previously mentioned knowledge and strategies. This approach aims to provide a comprehensive and advanced method for optimizing EMS systems.

2.2 Contribution

The novelty in this work resides in additionally maximizing the coverage of emergency situations and considering different types of ambulances. When a BLS ambulance is dispatched to an emergency requiring an ALS, it may reduce the patient's survival. Thus, this work considers that ALS ambulances can be used as BLS units, but the contrary is not allowed [5]. There are a few works that deal with different types of ambulances as we do in this work. McLay [20] determines how to optimally locate and use ambulances to improve patient survivability and coordinate multiple medical units with a hypercube queuing model. Grannan et al. [16] determine how to dispatch multiple types of air assets to prioritized service calls to maintain a high likelihood of survival of the most urgent casualties in a military medical evacuation by a binary linear programming model. In Yoon et al. [36], two types of vehicles are considered, but one of them is a rapid one that cannot offer the first care services of an ambulance. Moreover, neither of these works considers partial covering of the calls.

We denote our problem as the *Emergency Vehicle Covering and Planning* (EVCP) problem which consists of locating the limited number of two heterogeneous types of ambulances in different city points and dispatch them to the accident points, considering the uncertainty of the accident points, so as to maximize the coverage (even if partially) with short medical first aid response time. Usually, the location and dispatching decisions are made separately [7, 14, 35]. In the EVCP problem, these two interrelated decisions are simultaneously determined as done by Amorim et al. [2], Ansari et al. [3], Toro-Díaz et al. [32].

We propose a novel two-stage stochastic program for the EVCP problem. The stochastic program locates the limited number of heterogeneous types of ambulances in the first stage, and in the second stage, the dispatching of ambulances to accidents is determined. The EVCP stochastic model allows partial coverage of the accidents by the ambulances based on a decay function [34]. Similarly to Yoon et al. [36], we generate the call-arrival scenarios by sampling from emergency call logs to use them in the second stage of our stochastic model. In this manner, we address the volume of calls during a short period, such as Friday night hours. Thus, time is not explicitly measured, and it is assumed that a vehicle can be assigned only once during this high ambulance demand period [38]. Boujemaa et al. [11] use a bundle of calls but do not consider a heterogeneous ambulance fleet.

Another contribution of this work is the methodology to solve the EVCP stochastic model. Indeed, the proposed model can only be solved for relatively small instances with a restrictive number of scenarios. Thus, instead of decomposing the

model with Bender's methods as it is usually done [30, 37], we propose a location-allocation methodology [29, 33] that relies on the solution in an auxiliary surrogate model, which is faster to solve. We name this method *an intelligent feedback approach* because the location of the ambulances obtained by this surrogate model is used as input to the original model. Thus, we obtain high-quality solutions in a reasonable time with an off-the-shelf solver without complex decomposition techniques.

Some works use metaheuristic methods to solve their stochastic models. Toro-Díaz et al. [31] integrate location and dispatching decisions for EMS vehicles to minimize the mean response time of an emergency call and maximize the expected coverage demand, using a continuous-time Markov process to balance flow equations that control the busy fraction of each ambulance. A genetic algorithm can solve mid-size instances. Some others, such as Amorim et al. [2], use simulation to decide if ambulances stay at the potential sites established by a mathematical model or must be moved to another potential site to maximize the patient's survival. They work on a complete day period while we focus on high-demand periods of some hours. Moreover, we do not need a metaheuristic due to the high-quality solutions that we obtained with the Intelligent feedback approach. However, we would like to propose a matheuristic to improve the solution obtained from this approach.

Chapter 3

Problem desciption

The Emergency Vehicle Covering and Planning problem (EVCP) locates a limited number of two heterogeneous types of ambulances in different city points and dispatches them to the emergency scenes, considering the uncertainty of the emergency locations, to maximize the emergency total and partial coverage and the response time in which the patients receive medical first aids.

3.1 Description and assumptions

Let us formally describe the Emergency Vehicle Covering and Planning problem. Let set I include the possible demand points where patients may need medical attention in a city or region. This set can be very large, so we consider all the demand points observed in the historical data. In our case study, $|I|$ can be as large as 1500 demand points. Set L provides the potential sites or ambulance stations where ambulances could be located, such as hospitals, firehouses, malls, or similar places where the ambulance and the paramedics can wait for emergency calls. We consider instances with up to 30 potential sites for the experimental results. Set K contains the two types of ambulances available in the system: the BLS (labeled with index $k = 1$) and the ALS ambulances (labeled with index $k = 2$), which are limited by a known parameter η_k for each type $k \in K$. These ambulances must be allocated to a potential site $l \in L$ and dispatched toward a demand point $i \in I$ if there is an emergency situation.

The traveling time of any ambulance type from a potential site $l \in L$ to a demand point $i \in I$ is given by r_{li}. Ideally, ambulances should arrive in less than τ minutes in a life-threatening emergency. Usually, τ is a fixed value in the $[8, 15]$ range. This work also considers that the emergency is not covered if an ambulance

takes more than a maximum time $\tau_{\max}$ to arrive. In this case, sadly, the accident has probably been dealt with by other means.

Since the aim of the EVCP problem is to reduce the response time of the patient's first medical aid, even if it is in a partial or late way, we define a benefit decay function that only depends on the response time of a location $l \in L$ to any demand point $i \in I$:

$$c_{li} = \begin{cases} 1 & \text{if} \quad r_{li} \leq \tau, \\ 1 - \frac{r_{li}-\tau}{\tau_{\max}-\tau} & \text{if} \quad \tau < r_{li} < \tau_{\max}, \\ 0 & \text{if} \quad r_{li} \geq \tau_{\max}. \end{cases}$$

3.2 Information related to the scenarios

The operational level is represented by a set of scenarios S with a bundle list of arriving calls. Each scenario $s \in S$ represents a realization of accidents in the demand points. Thus, a scenario is represented by the number and type of ambulances needed at each demand points. Recall that an ALS ambulance can be sent instead of a BLS ambulance, but not the other way around. Thus, each scenario $s \in S$ indicates if there is an accident on a demand point $i \in I$ and provides the value a^s_{ki} related to the number of required ambulances of type $k \in K$.

For each scenario $s \in S$, let $I^s \subseteq I$ contain only the demand points $i \in I$ where ambulances are needed, that is, where $a^s_{ki} \neq 0$ for any $k \in K$. We define five different types of ambulance coverage related to the response times cases for each demand point $i \in I^s$:

- Total: the a^s_{ki} required ambulances of each type k are dispatched to i, and all arrive in less than τ time.
- Total-late: the a^s_{ki} required ambulances of each type k are dispatched, but at least one arrives between $(\tau, \tau_{\max})$ time.
- Partial: at least one of the a^s_{ki} required ambulances is not dispatched, for $k \in K$, but all the dispatched ones arrive in less than τ time.
- Partial-late: at least one of the a^s_{ki} required ambulances is not dispatched, for $k \in K$, but at least one of the dispatched arrives between $(\tau, \tau_{\max})$ time.
- Null: none of the a^s_{ki} required ambulances arrives in less than $\tau_{\max}$ time, for $k \in K$.

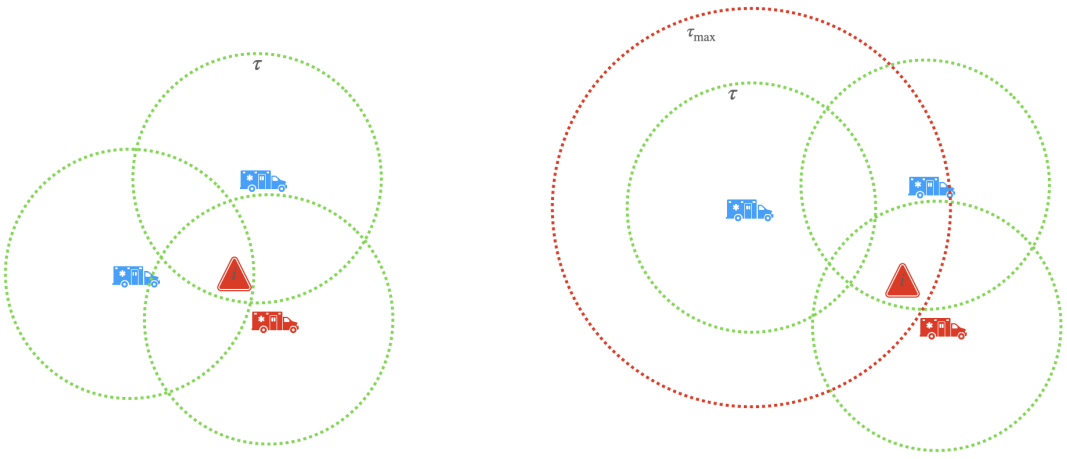


Figure 3.1: Two different coverage cases for a scenario $s \in S$ where $i \in I^s$ requires $a^s_{1i} = 2$ basic ambulances (blue) and $a^s_{2i} = 1$ advanced ones (red). Total coverage in the left: all ambulances arrive in less than the ideal time τ. Total-late coverage in the right: at least one of the ambulances arrives between $(\tau, \tau_{\max})$.

Figure 3.1 illustrates two different coverage cases for a scenario $s \in S$ where $i \in I^s$ requires $a^s_{1i} = 2$ basic ambulances (indicated in blue) and $a^s_{2i} = 1$ advanced one (indicated in red). All ambulances arrive in less than the ideal time τ for the Total coverage (left-hand-side figure). In the Total-late coverage line, at least one of the ambulances is late since it arrives between $(\tau, \tau_{\max})$ (right-hand-side figure). In the Partial coverage, the number of required ambulances is not met, but at least they arrive in less than the ideal time τ. In the Partial-late coverage, not only are there not enough ambulances to cover the demand point, but they arrive late, that is, between $(\tau, \tau_{\max})$. In the Null coverage, ambulances may be dispatched to the demand point, but since the arrival times are larger than $\tau_{\max}$, the demand point is considered uncovered.

Table 3.1 summarizes the sets and parameters used to describe the EVCP problem.

3.3 Maximum Expected Coverage stochastic formulation for the EVCP problem

The Maximum Expected Coverage (MEC) formulation is a stochastic integer quadratic programming model in which the first stage variables x_{lk} correspond to the number

I	set of possible demand points (possible accident places)
L	set of possible ambulance location sites
K	set of ambulance types
η_k	total number of ambulances in the system of type $k \in K$
r_{li}	response time from potential site $l \in L$ to demand point $i \in I$
τ	ideal response time to give the patients the first medical aid in an emergency
τ_{max}	maximum response time to cover an accident
c_{li}	benefit from traveling from potential site $l \in L$ to demand point $i \in I$
S	set of scenarios
a^s_{ki}	number of needed ambulances of type $k \in K$ at demand point $i \in I, s \in S$
I^s	set of demand points for $s \in S$ with at least a value $a^s_{ki} \neq 0$ for $i \in I, k \in K$

Table 3.1: Sets and parameters to describe the EVCP problem.

of ambulances of type $k \in K$ located at $l \in L$, and the second-stage variables correspond to the ambulance dispatching decisions at each demand point for each scenario $s \in S$:

$$y^s_{lki} = \begin{cases} 1 & \text{if an ambulance of type } k \in K \text{ in location } l \in L \\ & \text{is dispatched to demand point } i \in I^s, \text{ for scenario } s \in S, \\ 0 & \text{otherwise.} \end{cases}$$

We defined the following binary variables related to the *total* and *total-late* coverages related to the response times of the ambulances to the demand point $i \in I^s, s \in S$:

$$f^s_i = \begin{cases} 1 & \text{if demand point } i \in I^s \text{ has a } \textit{total} \text{ coverage,} \\ 0 & \text{otherwise,} \end{cases}$$

$$g^s_i = \begin{cases} 1 & \text{if demand point } i \in I^s \text{ has a } \textit{total-late} \text{ coverage,} \\ 0 & \text{otherwise.} \end{cases}$$

The following sets of binary variables are for the *partial* and *partial-late* coverages of the ambulances to the emergencies:

$$h^s_i = \begin{cases} 1 & \text{if demand point } i \in I^s \text{ has a } \textit{partial} \text{ coverage,} \\ 0 & \text{otherwise,} \end{cases}$$

$$w^s_i = \begin{cases} 1 & \text{if demand point } i \in I^s \text{ has a } \textit{partial-late} \text{ coverage,} \\ 0 & \text{otherwise.} \end{cases}$$

Finally, to indicate a null coverage of a demand point, we define

$$z_i^s = \begin{cases} 1 & \text{if active demand point } i \in I^s \text{ has a } \textit{null} \text{ coverage,} \\ 0 & \text{otherwise.} \end{cases}$$

The MEC formulation is as follows.

$$\max_x \mathbb{E}_{s\in S}[\mathcal{Q}^s(x)] \tag{3.1}$$

where

$$\mathcal{Q}^s(x) = \sum_{i\in I^s}(\alpha_1 f_i^s + \alpha_2 g_i^s + \alpha_3 h_i^s + \alpha_4 w_i^s - \phi z_i^s)$$

$$\text{s.t.} \sum_{l\in L} x_{lk} \le \eta_k \qquad k \in K \tag{3.2}$$

$$\sum_{i\in I^s} y_{lki}^s \le x_{lk} \qquad l \in L, k \in K, s \in S \tag{3.3}$$

$$f_i^s \sum_{k\in K} a_{ki}^s \le \sum_{l\in L}\sum_{k\in K} c_{li} y_{lki}^s, \quad a_{2i}^s f_i^s \le \sum_{l\in L} c_{li} y_{l2i}^s \qquad i \in I^s, s \in S \tag{3.4}$$

$$g_i^s \sum_{k\in K} a_{ki}^s \le \sum_{l\in L}\sum_{k\in K} y_{lki}^s, \quad a_{2i}^s g_i^s \le \sum_{l\in L} y_{l2i}^s \qquad i \in I^s, s \in S \tag{3.5}$$

$$g_i^s \le M\left(\sum_{l\in L}\sum_{k\in K} y_{lki}^s - \sum_{l\in L}\sum_{k\in K} c_{li} y_{lki}^s\right) \qquad i \in I^s, s \in S \tag{3.6}$$

$$h_i^s \le \sum_{k\in K} a_{ki}^s - \sum_{l\in L}\sum_{k\in K} y_{lki}^s, \quad h_i^s \le a_{2i}^s - \sum_{l\in L} y_{l2i}^s \qquad i \in I^s, s \in S \tag{3.7}$$

$$\sum_{l\in L}\sum_{k\in K} y_{lki}^s h_i^s \le \sum_{l\in L}\sum_{k\in K} c_{li} y_{lki}^s \qquad i \in I^s, s \in S \tag{3.8}$$

$$w_i^s \le \sum_{k\in K} a_{ki}^s - \sum_{l\in L}\sum_{k\in K} y_{lki}^s, \quad w_i^s \le a_{2i}^s - \sum_{l\in L} y_{l2i}^s \quad i \in I^s, s \in S \tag{3.9}$$

$$w_i^s \le M\left(\sum_{l\in L}\sum_{k\in K} y_{lki}^s - \sum_{l\in L}\sum_{k\in K} c_{li} y_{lki}^s\right) \qquad i \in I^s, s \in S \tag{3.10}$$

$$\sum_{l\in L}\sum_{k\in K} y_{lki}^s + z_i^s \ge 1 \qquad i \in I^s, s \in S \tag{3.11}$$

$$f_i^s + g_i^s + h_i^s + w_i^s + z_i^s = 1 \qquad i \in I^s, s \in S \tag{3.12}$$

$$x_{lk} \in \mathbb{Z}^+, y_{lki}^s \in \{0,1\} \qquad l \in L, k \in K, i \in I^s, s \in S \tag{3.13}$$

$$f_i^s, g_i^s, h_i^s, w_i^s, z_i^s \in \{0,1\} \qquad i \in I^s, s \in S. \tag{3.14}$$

The objective function (3.1) maximizes the expected value of the weighted coverage of the emergencies. The parameters $\alpha_1, \alpha_2, \alpha_3$ and α_4 are normalized weights that ponder the coverage type, and ϕ is the penalty for the null coverage. We assume that every scenario is equally probable since each $s \in S$ represents a sample of the high-demand period we are interested in.

Constraints (3.2) establish the available number of ambulances per type. Constraints (3.3) establish the relationship between the first and second-stage variables, meaning no ambulances can be dispatched from a potential site if no ambulances are located there. The *total* coverage of an emergency is defined by constraints (3.4). Indeed, if the time response of the location of the ambulances to the emergency is less than τ, then all $c_{li} = 1$ and total coverage variables f_i^s can be equal to one, for $l \in L, i \in I^s, s \in S$. The *total-late* coverage is defined by constraints (3.5) and (3.6). Constraints (3.5) allow the total-late coverage variables g_i^s to be one when dispatching variables are active. Meanwhile, constraints (3.6) track the demand points where the response time is between $(\tau, \tau_{\max})$ when the difference in the right-hand side of the equation is positive, that is, when there is a value $c_{lj} < 1$ associated to a dispatched ambulance, for $l \in L, i \in I^s, s \in S$. Note that this difference may be decimal, so we include a big M value. The *partial* coverage is defined by constraints (3.7) and (3.8). Recall that, in this case, not all the needed ambulances are dispatched to the emergencies, but the ones that are dispatched have an ideal response time. Thus, constraints (3.7) activate variables h_i^s if the number of dispatched ambulances is less than the required ones. Quadratic constraints (3.8) guarantee that the dispatched ambulances arrive within the ideal response time, that is, their corresponding value $c_{li} = 1$, for $l \in L, i \in I^s, s \in S$. Constraints (3.9) and (3.10) define the *partial-late* coverage. Constraints (3.9) activate the w_i^s variables when the number of required ambulances exceeds the number of dispatched ones. Similarly to the total-late coverage, constraints (3.10) track the ambulances with a response time larger than the ideal one and must be multiplied by a big M. The *null* coverage is activated by constraints (3.11). All the coverage constraints are related to constraint (3.12) that ensures only one type of coverage for each emergency. Finally, (3.13) and (3.14) establish the nature of the decision variables.

The novelty of the MEC model is the stochastic total/partial coverage per emergency by two types of ambulances. Nevertheless, the related number of variables and constraints is usually large. Moreover, constraints (3.8) are quadratic. An integer linear stochastic model with a classical linearization method could be easily formulated. Still, previous experiments showed similar times between the linearized and the quadratically constrained models when solved with integer programming solvers, so we keep the quadratic one for the Intelligent Feedback methodology presented in the next section.

3.4 Surrogate-based feedback method for the EVCP problem

The EVCP problem is $\mathcal{NP}$-hard since the classical $\mathcal{NP}$-hard facility location problem [24] could be polynomially reduced to it. The MEC model is experimentally challenging to solve, even for medium-sized instances, as shown in Section 4. Thus, we propose a surrogate-based feedback method (SBFM) to obtain approximate solutions to the EVCP problem based on an auxiliary disaggregated model, named *Surrogate Ambulance-Based Coverage* (SABC) model, which is faster to solve.

The SABC model's essential characteristic is that its objective function does not rely on emergency coverage, as in the MEC model; it only counts the number of ambulances sent on time, late, or null to emergency demand points. Moreover, its resolution time is extremely fast since it requires fewer variables and constraints than the MEC model. However, disaggregating an emergency situation into the number of ambulances needed does not capture emergency coverage, which is crucial for an EMS system.

In addition to the location variables x_{li}, the SABC model requires the following ambulance dispatching binary variables for $k \in K, l \in L, i \in I^s, s \in S$:

$$u^s_{lki} = \begin{cases} 1 & \text{if ambulance of type } k \text{ is dispatched from site } l \text{ to point } i \\ & \text{with response time less than } \tau, \\ 0 & \text{otherwise,} \end{cases}$$

$$v^s_{lki} = \begin{cases} 1 & \text{if ambulance of type } k \text{ is dispatched from site } l \text{ to } i \\ & \text{with response time in } (\tau, \tau_{\max}), \\ 0 & \text{otherwise.} \end{cases}$$

Variables u^s_{lki} indicate the ambulances with an ideal response time dispatched from the location sites corresponding to a decay function value $c_{li} = 1$. While variables v^s_{lki} indicate the ones with a larger than τ response time which have a value $c_{li} < 1$. The number of required ambulances k in an emergency demand point i that are not dispatched are counted by integer variable ζ^s_{ki}, for $k \in K, i \in I^s, s \in S$. The SABC is as follows.

$$\max_x \ \mathbb{E}_s[\mathcal{G}^s(x)], \tag{3.15}$$

$$\text{where } \mathcal{G}^s(x) = \left[\sum_{l \in L} \sum_{k \in K} \sum_{i \in I^s} (\beta_1 u^s_{lki} + \beta_2 v^s_{lki}) - \sum_{k \in K} \sum_{i \in I^s} \phi \zeta^s_{ki} \right] \tag{3.16}$$

$$\text{s.t.} \sum_{l\in L} x_{lk} \leq \eta_k \qquad k \in K \tag{3.17}$$

$$\sum_{i\in I^s}(u^s_{lki} + v^s_{lki}) \leq x_{lk} \qquad l \in L, k \in K, s \in S \tag{3.18}$$

$$u^s_{lki} \leq c_{li} \qquad l \in L, i \in I^s, k \in K, s \in S \tag{3.19}$$

$$u^s_{lki} + v^s_{lki} \leq 1 \qquad l \in L, i \in I^s, k \in K, s \in S \tag{3.20}$$

$$a^s_{1i} + a^s_{2i} = \sum_{l\in L}\sum_{k\in K}(u^s_{lki} + v^s_{lki} + \zeta^s_{ki}) \qquad i \in I^s, s \in S \tag{3.21}$$

$$a^s_{2i} \leq \sum_{l\in L}(u^s_{l2i} + v^s_{l2i} + \zeta^s_{2i}) \qquad i \in I^s, s \in S \tag{3.22}$$

$$x_{lk}, \zeta^s_{ki} \in \mathbb{Z}^+, u^s_{lki}, v^s_{lki} \in \{0,1\} \qquad l \in L, k \in K, i \in I^s, s \in S$$

The objective function (3.15) maximizes the expected value of the on-time and late dispatched ambulances minus a penalty ϕ for the required ambulances that could not be dispatched in less than $\tau_{\max}$ time response. The weights $\beta_1 > \beta_2$ are normalized parameters that prioritize the ambulances dispatched with a response time less than τ. As in the previous model, no more than the available ambulances can be located on the sites, corresponding to constraints (3.17). The number of ambulances dispatched on time or late is less than the number of ambulances located, as indicated by constraints (3.18). Constraints (3.19) define the ambulances dispatched with an ideal response time of less than τ. Thus, if $c_{li} = 1$, then the ambulance will have an ideal response time, while constraints (3.20) activate the late variables for which their response time is between $(\tau, \tau_{\max})$. With constraints (3.21) and (3.22), the non-covered emergencies, ζ^s_{ki} variables are defined for $i \in I^s, s \in S$. Recall that advanced ambulances can be dispatched instead of basic ones. Finally, the nature of the variables is stated.

The surrogate-based feedback method: Under the SBFM, the SABC stochastic model is solved first. From its optimal solution, we obtain the location of the ambulances of the first stage corresponding to the value of x_{lk} variables, for $l \in L, k \in K$. Let the solution vector of these values be called x^{SABC}. Then, in the allocation stage, we solve MEC taking x^{SABC} as input. We call this model MEC(x^{SABC}), or simply MEC(SABC), implying that it is the solution of the MEC model with the location variables fixed with the solution of the surrogate model SABC. Since the first stage variables are fixed, the MEC(SABC) model becomes easier to solve and yields high-quality solutions. We could implement a local search neighborhood based on the location variables x_{lj} to diversify the solution yield by variables x^{SABC}. However, experimental results show that the quality of the SBFM solutions is exceptionally high with a single feedback.

As mentioned, the SABC auxiliary model is a surrogate for the MEC formulation. Thus, the solutions obtained by the MEC and SABC are not equivalent. Nevertheless, the solutions of the SABC model can be mapped into solutions for the EVCP problem, as shown in Algorithm 1. In this manner, we can compare both models regarding emergency coverage, even if the SABC model is short-sighted regarding this objective. Step 3 activates the total coverage when all the required ambulances arrive in less than τ response time. Step 4 verifies if a dispatched ambulance has a response time in $(\tau, \tau_{\max})$, corresponding to the total-late coverage. Step 6 checks that not all the required ambulances are dispatched but they arrive between the ideal time, while Step 8 verifies that the dispatched ambulances are not all the required ones and at least one of them has a response time in $(\tau, \tau_{\max})$. Finally, Step 10 activates the null variable.

Algorithm 1 Transformation of a SABC solution into a MEC solution

1: **require** solution of the SABC model $(\bar{x}, \bar{u}, \bar{v})$
2: **for** $i \in I^s, s \in S$ **do**
3: **if** $\sum_{l \in L, k \in K} \bar{u}^s_{lki} = a^s_{ki}$ **then** $f^s_i = 1$ ▷ total coverage
4: **if** $\sum_{l \in L, k \in K} \bar{u}^s_{lki} < a^s_{ki}$ and $\sum_{l \in L, k \in K} \bar{u}^s_{lki} + \bar{v}^s_{lki} = a^s_{ki}$
5: **then** $g^s_i = 1$ ▷ total-late coverage
6: **if** $\sum_{l \in L, k \in K} \bar{u}^s_{lki} < a^s_{ki}$ and $\sum_{l \in L, k \in K} \bar{v}^s_{lki} = 0$
7: **then** $h^s_i = 1$ ▷ partial coverage
8: **if** $\sum_{l \in L, k \in K} \bar{u}^s_{lki} + \bar{v}^s_{lki} < a^s_{ki}$ and $\sum_{l \in L, k \in K} \bar{v}^s_{lki} > 0$
9: **then** $w^s_i = 1$ ▷ partial-late coverage
10: **otherwise** $z^s_i = 1$ ▷ null coverage
11: **return** MEC solution $(\bar{x}, f, g, h, w, z)$

3.5 MATHEURISTIC TO IMPROVE THE MEC MODEL

The Surrogate-based feedback method for the MEC(SABC) model has good results. However, a disadvantage of this method is that we obtain only one solution for the ambulance location from the surrogate model SABC. This is a problem because we disown the optimal solution for the MEC model and we are not sure if the SABC solution is close to that optimality. Trying to improve the solution x^{SABC} we proposed a local search procedure, which is a matheuristic considering four neighborhoods, named as SABC Matheuristic. These four different neighborhoods are as follows:

- Neighborhood 1, (N_1): exchange one active potential site with another active potential site.

- Neighborhood 2, (N_2): pick half of an active potential site and add it to a non-active potential site.
- Neighborhood 3, (N_3): pick half of an active potential site and add it to another active potential site.
- Neighborhood 4, (N_4): exchange one active potential site with one non-active potential site.

N_1 and N_4 change BLS ambulances with BLS ambulances and then ALS ambulances with ALS ambulances. N_2 and N_3 change only BLS ambulances with BLS ambulances due to the small quantity of ALS ambulances in the EMS system.

The algorithm to solve the SABC Matheuristic has the x^{SABC} as an initial solution. First, we obtain the value of the objective function for this initial solution, defined as t^{SABC}, which is the best solution at this point. Then, we construct the first neighborhood from the x^{SABC}. The t^{SABC} objective value is compared with each neighbor's objective value, obtained as the MEC(SABC) methodology. If a neighbor's solution is better than t^{SABC}, we consider this solution as the best one for the Matheuristic and we save the objective value and variables results. Otherwise, we have the initial solution as the best one when the algorithm is finished. Regardless new best solution is the initial solution or not, we construct the second neighborhood from the x^{SABC}, and each neighbor is compared with the best solution at the moment. We repeat this procedure for the other two neighborhoods and, when the comparisons are finished, we obtain the best solution, as we can see at the Algorithm 2.

This procedure calculates each neighbor's objective value as in the MEC(SABC) methodology in the four different neighborhoods, which makes the SABC Matheuristic take so much time to check each neighborhood. This Matheuristic aids in improving x^{SABC} solutions but not for all the instances, as we can see in Chapter 5.

The next section compares the MEC, MEC(SABC), and even the SABC solutions.

Algorithm 2 SABC Matheuristic to improve the MEC model

require solution of the SABC model x^{SABC}
$x^* = x^{SABC}$
$t^* = MEC(x^{SABC}) = t^{SABC}$
while not all neighborhoods N_i, where $i \in \{1, 2, 3, 4\}$, have been visited **do**
    for $x' \in N_i$ **do**
        Evaluate $t' = MEC(x')$
        if $t' > t^*$ **then** $x^* = x', t^* = t'$
return solution (x^*, f, g, h, w, z) and objective value t^*

Chapter 4

Experimental assessment

This chapter presents an empirical assessment of models and the solution methodology previously described to solve the EVCP problem. We used Gurobi Optimizer 10.0.2 with Python 3.10 to solve the integer programming models MEC, SABC, and MEC(SABC). The experiments were carried out on an Intel Core i7 at 3.1 GHz with 16 GB of RAM under the macOS Catalina 10.15.7 operating system. Each execution of the integer linear programming solvers had a CPU time limit of 10800 seconds. For SABC Matheuristic the solver had a CPU limit of 300 seconds for neighbor's evaluation and 10800 (**CHECAR**) second in total.

4.1 Instance generation

The value ranges of our instance generator are based on real-world data taken from Monterrey, Mexico. In the literature, there are no suitable benchmarks for our problem. The databases for the Monterrey case study showed a larger number of possible demand points, $|I| \in \{168, 270, 500, 900, 1500\}$ compared to the one from the literature with $|I| \leq 270$ [36]. The number of possible locations for ambulances in Monterrey is $|L| \in \{16, 50, 100\}$, which is also larger than the one from the literature ($\leq$30) since not only hospitals and fire stations can be considered. We consider the whole city of Monterrey, so the number of ambulances $(\eta_1, \eta_2) = (35, 20)$ is also greater than the ones from the literature cases (6 ambulances per type [36]). The number of scenarios is set to be as large as that in the literature $|S| \in \{10, 50, 100, 150, 200\}$. Thus, our benchmark has 15 instances for which five different scenario settings were built.

For each instance, we simulated a two-hour high-demand period. Each scenario $s \in S$ consists of a set of demand values per ambulance type and per demand point

$\{a^s_{ki}\}_{k\in K, i\in I, s\in S}$. Fewer demand points imply a larger city grid and a larger proportion of emergencies per demand point. Therefore, when $|I| = 168$, around 30% of the demand points may have a value different from 0. In contrast, when $|I| = 1500$, only 1% of the demand points will require ambulances. This setting reflects the number of emergencies per hour observed in the case study. Instances are built such that most emergencies require a single ambulance, but as observed in real cases, some of them may require up to three ambulances.

The ideal ambulance response time is $\tau = 10$ minutes, while the maximum response time is $\tau_{\max} = 30$ minutes. For the MEC formulation, we use the following weights in the objective function (3.1): $\alpha_1 = 0.65, \alpha_2 = 0.2, \alpha_3 = 0.1$, and $\alpha_4 = 0.05$. In this manner, the total coverage is the most sought-after, while the partial-late cover has less benefit. Surprisingly, the value of the big M of the model is not the main cause of the execution time of the MEC model. Thus, a simple value $M = 1000$ is set.

For the SABC objective function (3.16) we use $\beta_1 = 0.7$ and $\beta_2 = 0.3$. These values reflect the aim to send primordially the required ambulances with an ideal response time. The penalty for null coverage in the MEC model or when a required ambulance cannot be dispatched to the emergency in less than $\tau_{\max}$ time in the SABC model is set to $\phi = 1/|S| + 0.0005$.

All instances with their related scenarios and detailed solutions are available at `https://doi.org/10.6084/m9.figshare.25928401`.

Chapter 5

Experimental work

In this chapter, we analyze the parameters of the EVCP problem that impact the performance of the objective values of our stochastic methodologies. Several questions arise. We wish to investigate how sensitive the model is to the number of scenarios in terms of solution quality and solution time. We also want to determine the size of tractable instances.

5.1 Objective values for the MEC, MEC(SABC) and SABC Matheuristic

In this first experiment, we solve the instances using the original MEC model. Figure 5.1 consists of six plots. The three plots in the left column vary the number of demand points (x-axis), comparing each one to the value of the objective function when different scenarios are tested. The three plots in the right-hand side column vary the tested number of scenarios and show the variation in the solution value for each number of demand points. The upper plots consider a number of possible locations for the ambulances of $|L| = 16$, the middle plots of $|L| = 50$, and the lower plots of $|L| = 100$. Straight lines are the best objective values, while dotted ones are the best bounds found.

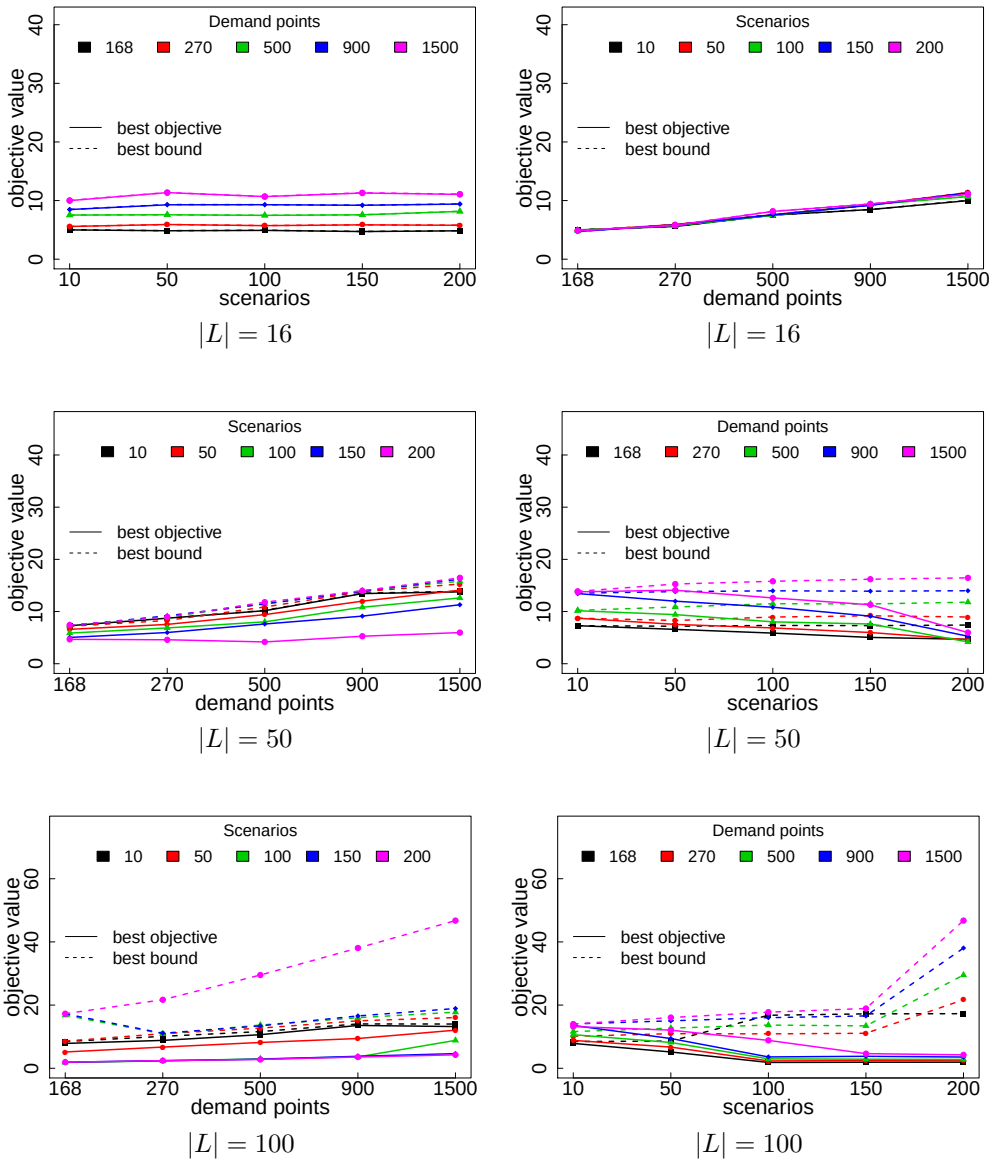


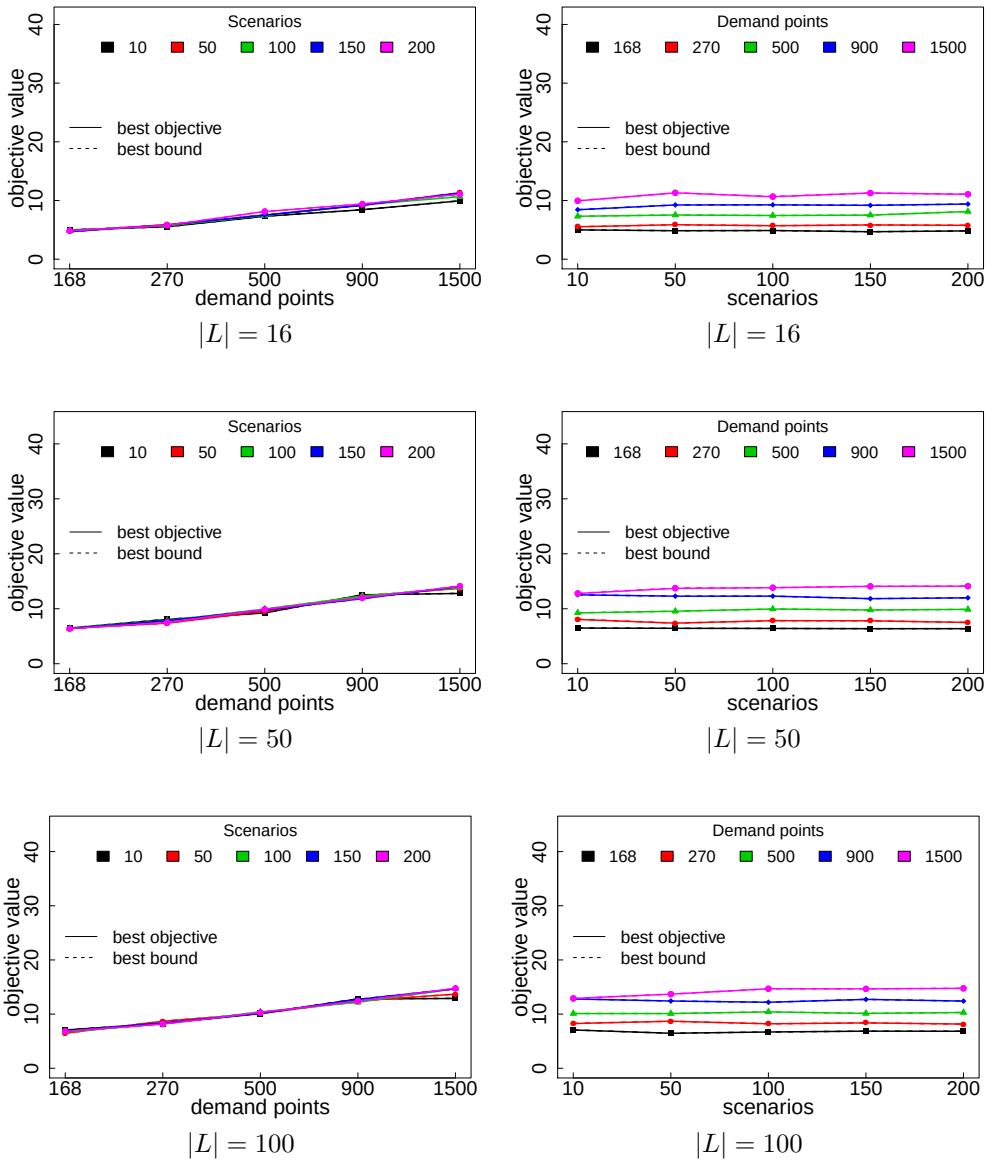
Figure 5.1: Objective MEC.

As can be seen from the plots, the difference between the best objective and the best bound (and thus, the relative optimality gaps[1]) are negligible for small instances with 16 potential locations sites. Still, the gaps become larger for the instances with 50 and 100 potential sites. The number of demand points where emergencies may occur and the scenarios considered make the instances harder to solve optimally within the time limit. Thus, the deterministic equivalent integer program of MEC can only handle small instances with a few scenarios, demand points (emergency points), and potential ambulance sites. Note that the larger the number of scenarios

[1](best objective - best bound)/best objective.

in the plots on the left-hand side, the better the objective function. This implies that a better sampling of the emergency demand points benefits the quality of the solution related to the ambulance response time. The plots on the right side show that the larger the size of the demand point set, the harder it is to solve the instance.

For the MEC(SABC) methodology, we have the solution represented in Figure 5.2, which have a similar structure to the previous one. For this methodology, the number of scenarios not affects the results for the objective value due to the MEC(SABC) only considers the ambulances serving the accidents. To this methodology, the optimal is found in a easier and faster way than the MEC, but it is different to the optimal at the MEC so we need to compare the objective values of both of them.



$|L| = 16$ $|L| = 16$

$|L| = 50$ $|L| = 50$

$|L| = 100$ $|L| = 100$

Figure 5.2: Objective M2M1.

Now, we compare the solution values of the equivalent integer program of MEC with the ones obtained by MEC(SABC) in Figure 5.3. As can be seen, while the number of scenarios, demand points, and potential sites slightly affects the performance of MEC(SABC), it obtains better objective function values than those obtained by the MEC model for the larger instances that reported positive gaps. Indeed, the optimality gaps of the MEC(SABC) model always equal 0 within the time limit that we established. In addition, the MEC(SABC) model tends to be less dependent on the number of scenarios. Thus, although we cannot guarantee optimality with the MEC(SABC) model, it obtains faster and higher-quality solutions

than those obtained by the MEC equivalent model.

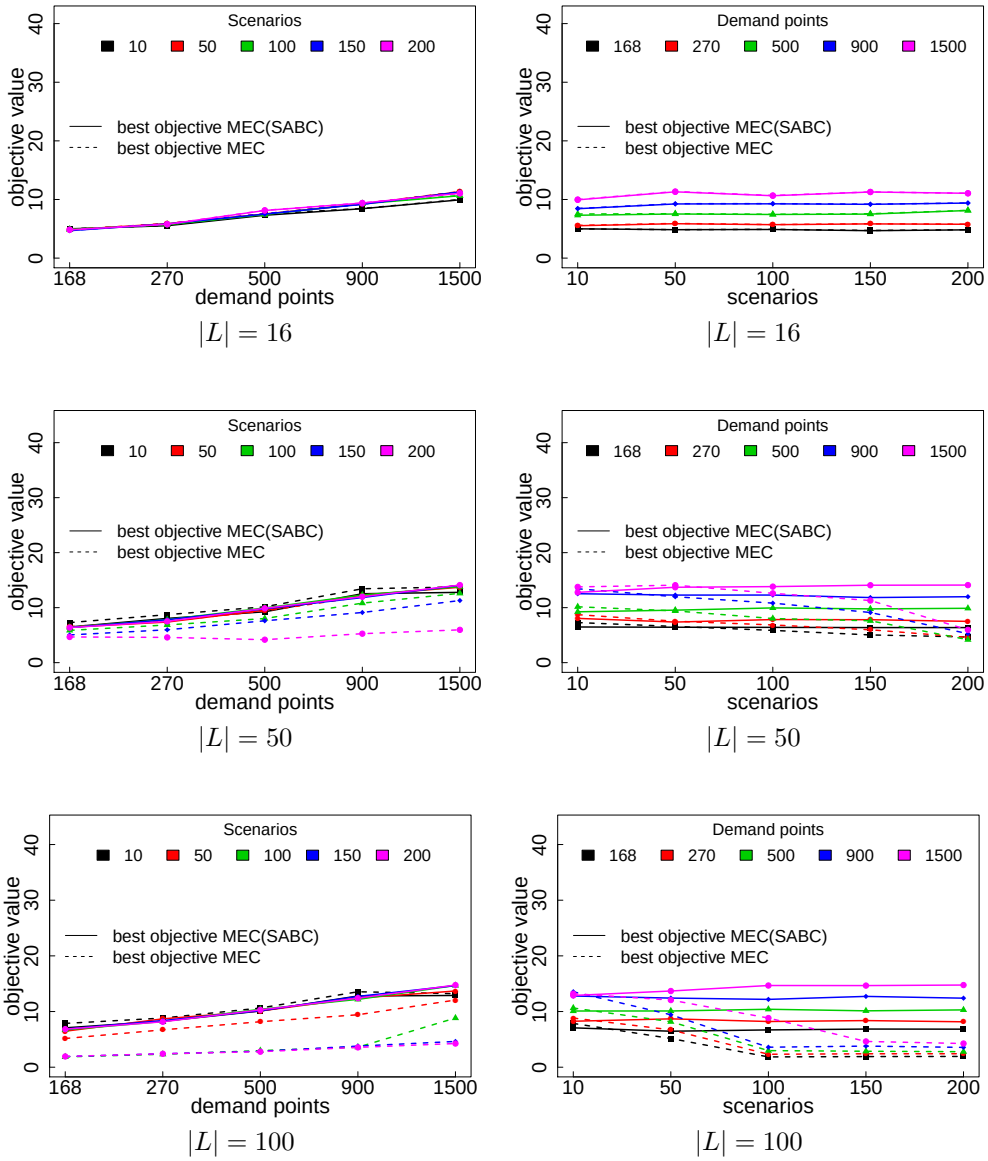


Figure 5.3: Objective M2M1-M1.

Another experimental results that we have in Figure 5.4 are for the SABC Matheuristic.

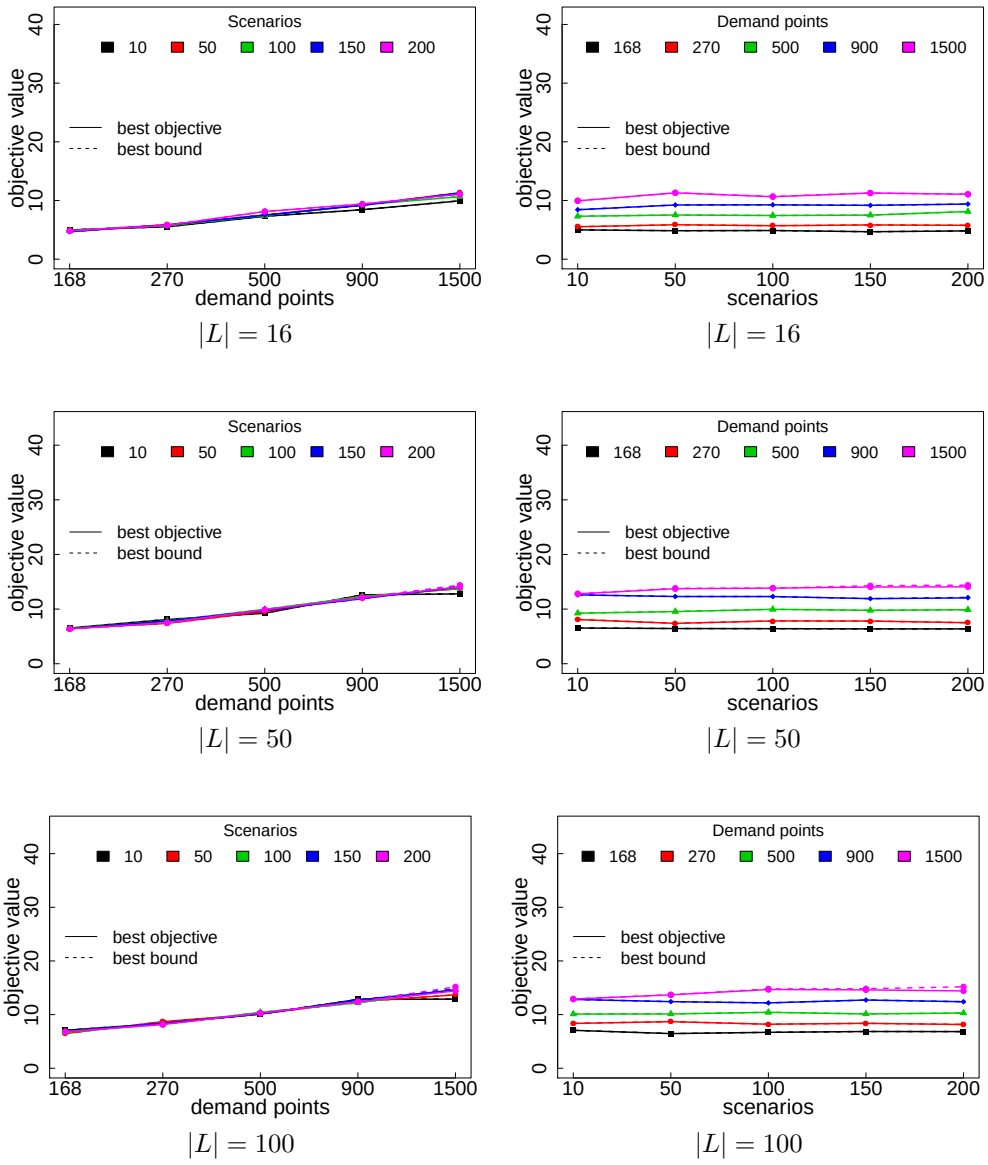


Figure 5.4: Objective Math.

To see if the Matheuristic shows better results than MEC(SABC) methodology we compare the results between the objective values obtained from both of them. As we can see, there is not a significant difference in the objective value result. This small improvement may be due to the computational limit time the solver had to solve each instance. Although, we can see the improvement percentage between the MEC and the MEC(SABC) in column five of the Table 5.1, and the improvement percentage between the MEC(SABC) and the SABC Matheuristic in column six. Something interesting is that SABC Matheuristic mostly improves MEC solution for those instances where MEC(SABC) not improves the objective value.

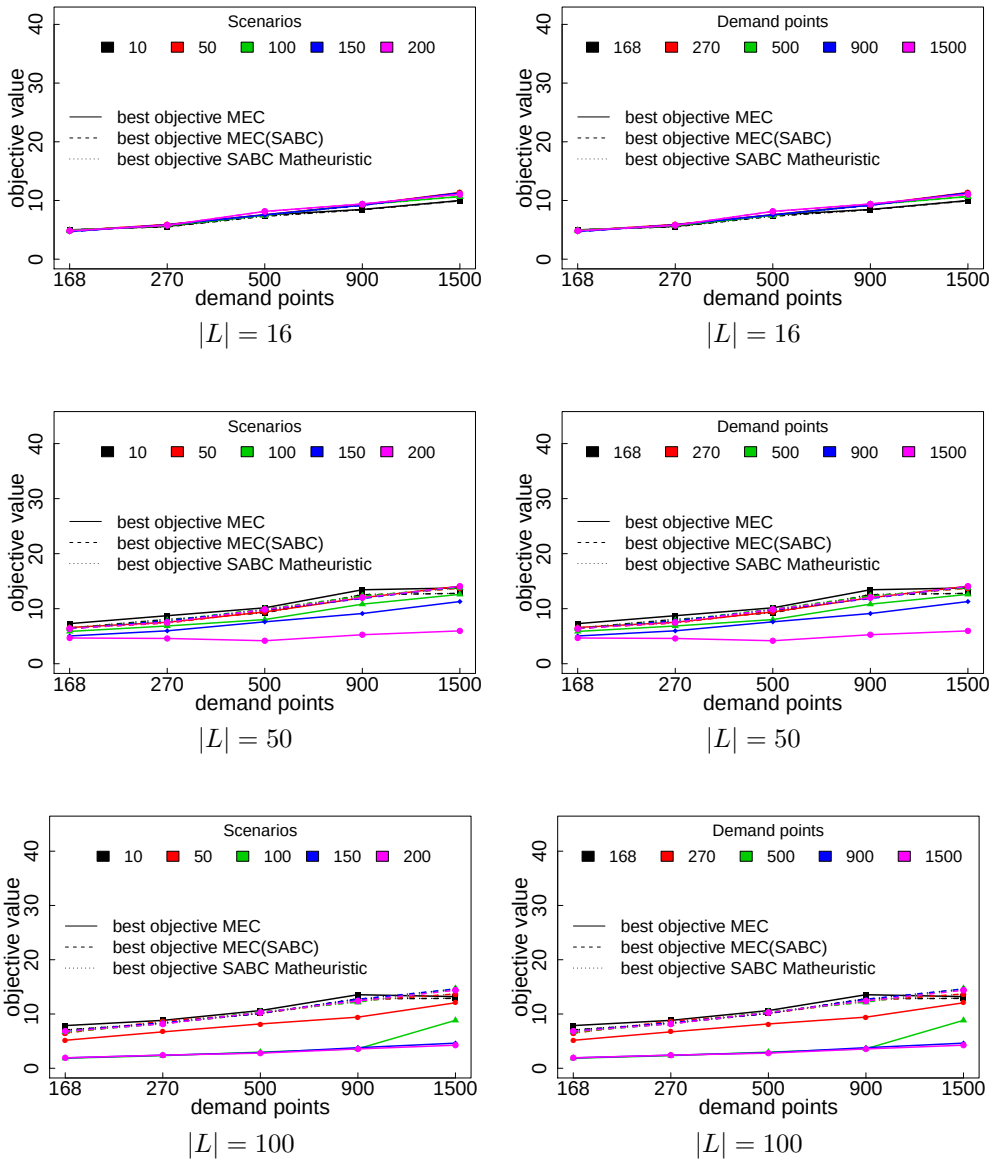


Figure 5.5: Objective Comp.

Table 5.1: Objective Values Comparison

Objective values					
Instance $\|I\|$ $\|L\|$ $\|S\|$	MEC	MEC(SABC)	Matheuristic (Math)	% Improve MEC vs MEC(SABC)	% Improve MEC(SABC) vs Math
1500 16 200	11.0536	11.05635	11.05635	0.024878773	-
168 50 10	7.285	6.485	6.53	-	0.693909021

Continued on next page

Table 5.1: Objective Values Comparison (Continued)

168 50 50	6.5825	6.441	6.441	-	-
168 50 100	5.86825	6.4167	6.4167	9.346057172	-
168 50 150	5.04052222	6.375988893	6.375988893	26.4946093	-
168 50 200	4.6735	6.3773	6.3773	36.4566171	-
270 50 10	8.725	8.05	8.085	-	0.434782609
270 50 50	7.535	7.381	7.389	-	0.108386398
270 50 100	6.85685	7.83305	7.83305	14.23685803	-
270 50 150	5.9789	7.80876667	7.80876667	30.60540685	-
270 50 200	4.59045	7.504750002	7.504750002	63.48615064	-
500 50 10	10.1585	9.2465	9.2465	-	-
500 50 50	9.4105	9.547000009	9.547000009	1.450507504	-
500 50 100	8.02305	9.9521	9.9521	24.04384866	-
500 50 150	7.61198889	9.7781	9.7781	28.45657216	-
500 50 200	4.17615	9.883800001	9.883800001	136.6725333	-
900 50 10	13.4365	12.535	12.603	-	0.542481053
900 50 50	11.987	12.286	12.301	2.4943689	0.122090184
900 50 100	10.8234	12.29315	12.29565	13.57937432	0.020336529
900 50 150	9.11984444	11.83602222	11.91226667	29.78315908	0.644172873
900 50 200	5.26805	11.98405	12.04855	127.4855022	0.53821538
1500 50 10	13.7735	12.78400001	12.78400001	-	-
1500 50 50	14.084	13.709	13.7235	-	0.10576994
1500 50 100	12.6053	13.82855	13.82855	9.704251386	-
1500 50 150	11.2970444	14.03975556	14.03975556	24.27812983	-
1500 50 200	5.9664	14.0559	14.0723	135.5842719	0.116676984
168 100 10	7.865	7.070000095	7.070000095	-	-
168 100 50	5.156	6.463000009	6.463000009	25.34910801	-
168 100 100	1.8485	6.6976	6.6976	262.3262105	-
168 100 150	1.9081	6.85182223	6.85182223	259.0913594	-
168 100 200	1.93335	6.8329	6.8329	253.4228153	-
270 100 10	8.82500001	8.255000064	8.355	-	1.211386255
270 100 50	6.71	8.6855	8.6855	29.44113264	-

Continued on next page

Table 5.1: Objective Values Comparison (Continued)

270 100 100	2.32665	8.208650004	8.208650004	252.8098341	-
270 100 150	2.41097778	8.375244449	8.375244449	247.3795788	-
270 100 200	2.4159	8.15665	8.15665	237.6236599	-
500 100 10	10.6315	10.0945	10.0945	-	-
500 100 50	8.169	10.101	10.124	23.6503856	0.22770023
500 100 100	2.97905	10.40835	10.4282	249.384871	0.190712265
500 100 150	2.86453333	10.12661111	10.12661111	253.5169584	-
500 100 200	2.78645	10.29585	10.29585	269.4970304	-
900 100 10	13.5535	12.79	12.8315	-	0.324472244
900 100 50	9.42	12.4165	12.4165	31.80997877	-
900 100 100	3.59	12.18375	12.18375	239.3802228	-
900 100 150	3.78992222	12.71902222	12.71902222	235.6011411	-
900 100 200	3.5589	12.4005	12.4005	248.4363146	-
1500 100 10	13.206	12.887	12.889	-	0.015519516
1500 100 50	12.063	13.6875	13.6875	13.4667993	-
1500 100 100	8.8318	14.67245	14.67245	66.13204556	-
1500 100 150	4.62758889	14.60712222	14.60712222	215.6529798	-
1500 100 200	4.2533	14.3678	14.41435	237.8035878	0.323988387

AGREGAR LA TABLA QUE COMPARA OBJETIVOS CON TIEMPOS DE EJECUCIÓN MÁS LARGOS.

5.2 Response time for the MEC, MEC(SABC) and SABC Matheuristic

The following experiment compares the running times of the equivalent MEC model with the MEC(SABC) method. Recall that MEC(SABC) attempts to exploit that the surrogate model SABC is very tractable and solved relatively quickly. To this end, Figure 5.6 shows plots of the running time in seconds of the instances with $|L| = 16$, $|L| = 50$ and $|L| = 100$ potential location sites for the equivalent MEC, and Figure 5.7 shows response times for the SBFM with the same potential sites. The x-axis of the plots corresponds to the number of scenarios, and we vary the number of emergency demand points. Recall that the MEC model with $|L| = \{50, 100\}$

reaches the time limit even for ten scenarios and few demand points.

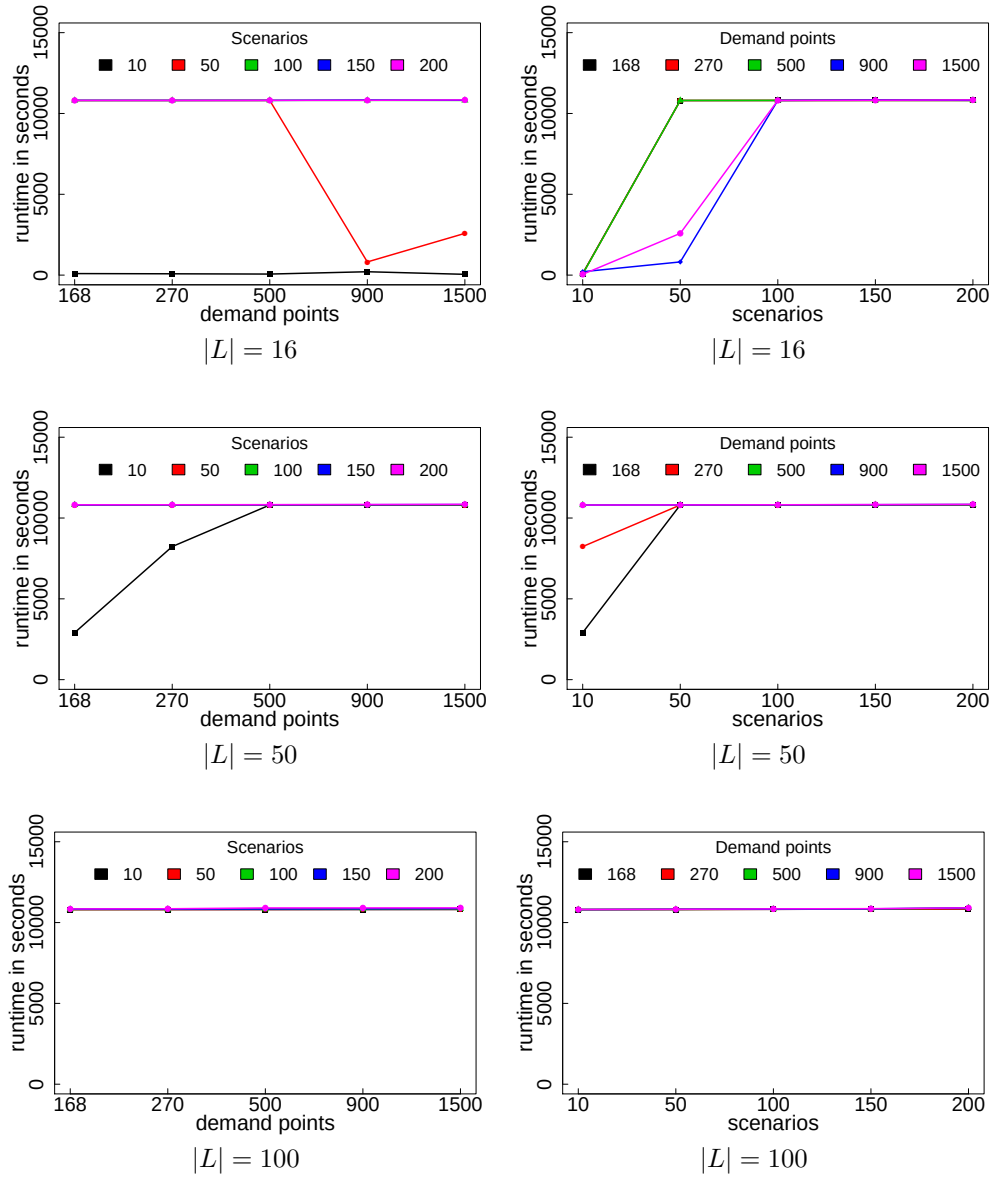


Figure 5.6: Time MEC.

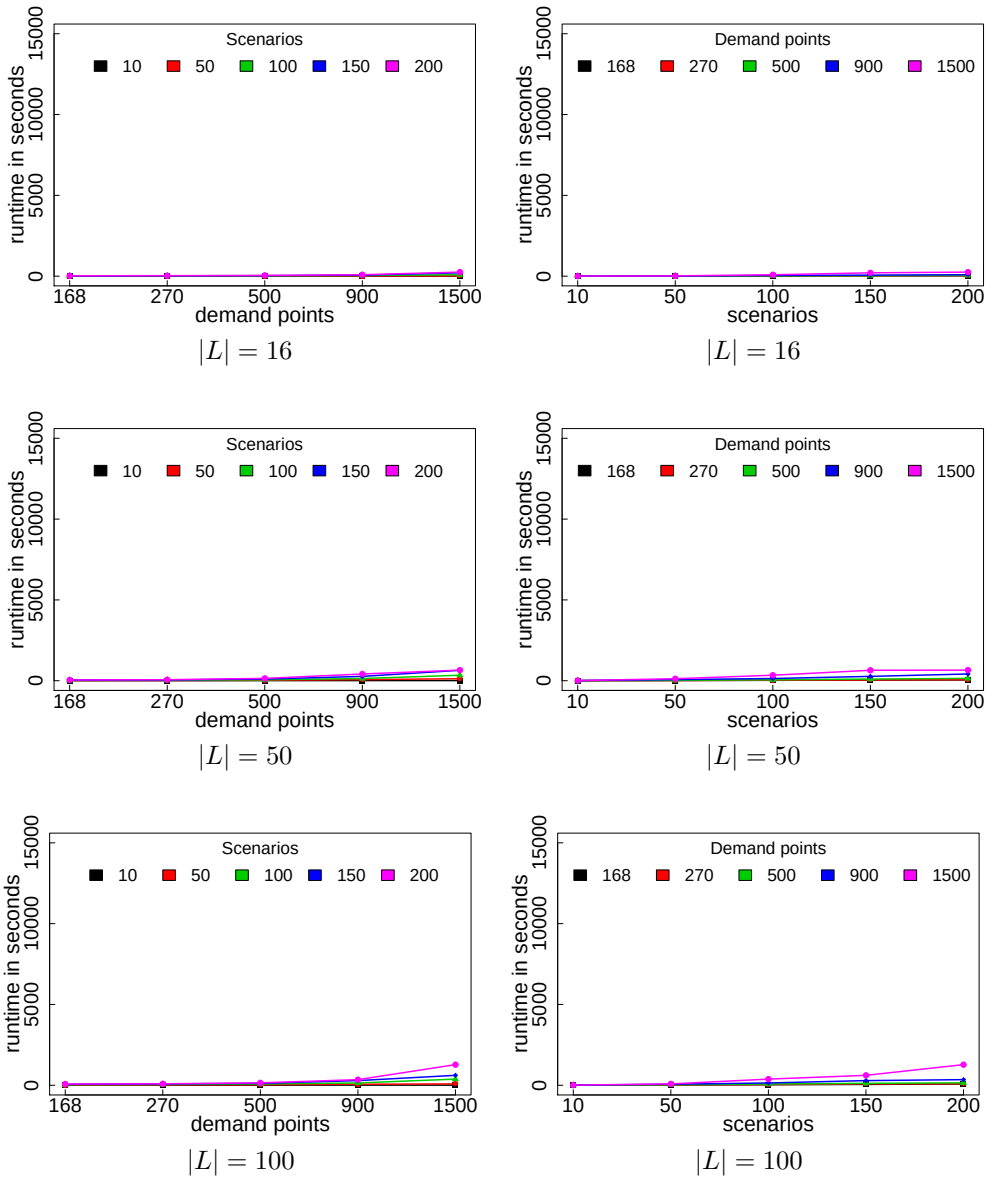


Figure 5.7: Time M2M1.

Figure 5.6 shows that the main disadvantage of the MEC model is its computational time, which increases significantly with the number of demand points, potential sites, and scenarios, even for small instances with 16 potential location sites for ambulances. The SABC model is extremely fast, even for large instances, and yields an initial solution to the assignment of ambulance location in a short time to allow the MEC(SABC) model, Figure 5.7, to be solved faster than the MEC model and obtain high-quality solutions. The MEC(SABC) location-allocation strategy inherits not only its fast computational time from the SABC, but also yields coverage per emergency situation, which is the main objective for the EVCP problem. The

MEC(SABC) model is an approximated mathod, but it gives solutions that are as good as the MEC and even better when the MEC instances do not reach optimality and its gaps are large. The SBFM solves most instances in less than a minute.

An interesting advantage of the SBFM is that only one iteration is needed. In fact, once the location of the ambulances has been retrieved from the SABC model and fed back to the MEC model, we could perturb the ambulances either randomly or with a local search, the allocation of the ambulances, and iterate again. Nevertheless, we could not systematically generate a neighborhood around a location solution that yields better solutions with the MEC(SABC) approach. This implies that local maximums are often reached with this first feedback and that complex or more diverse neighborhoods should be built to allow escaping from these solutions. It would probably be interesting to enable local search movements that do not yield immediate benefits.

Comparing SABC Matheuristic runtime with MEC and MEC(SABC) we show Figure 5.8

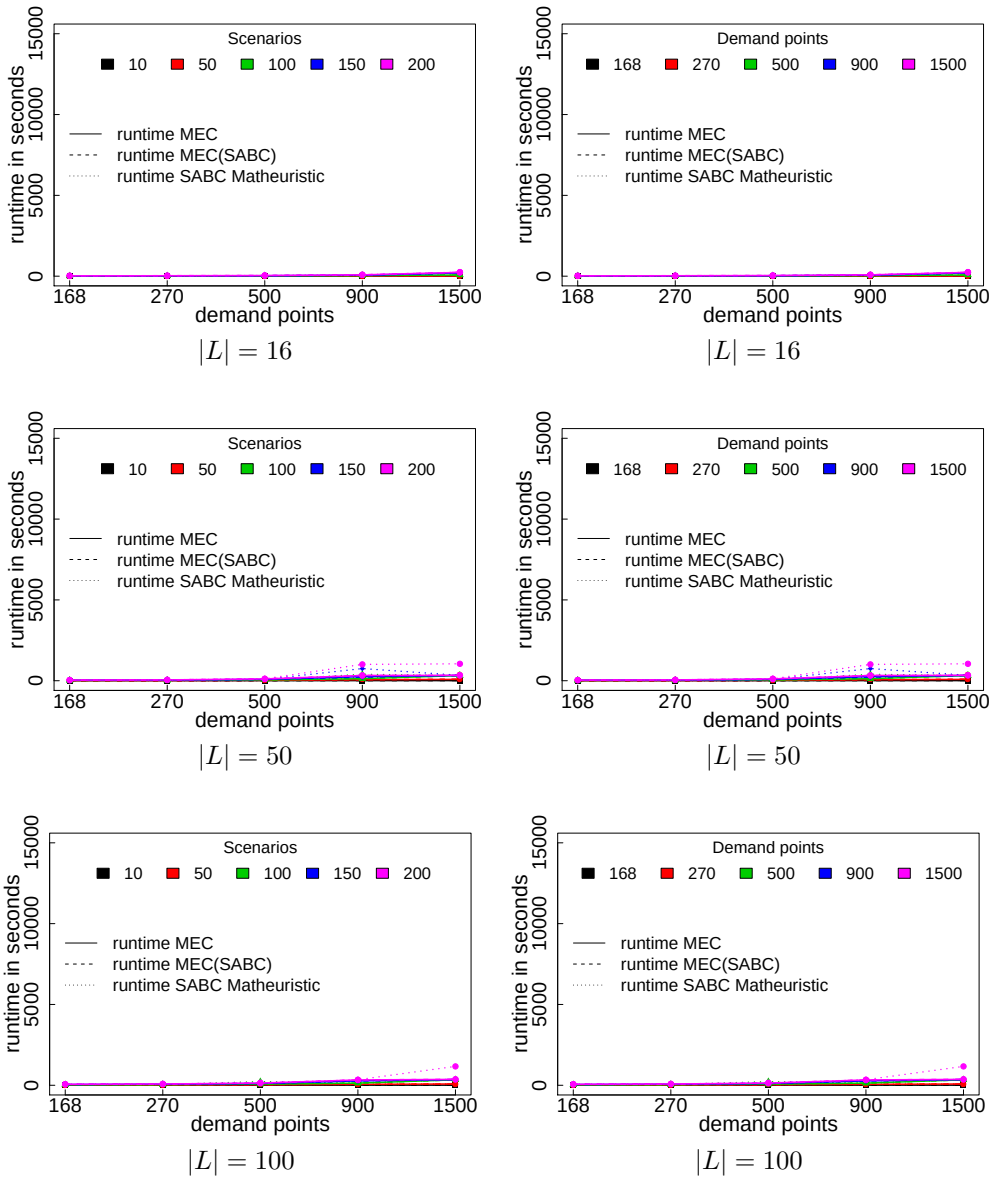


Figure 5.8: Time Comp.

5.3 Coverage for the MEC, MEC(SABC) and SABC Matheuristic

The objective values and running times are crucial for evaluating the models' performance. However, the most critical objective of the EVCP problem is to cover the largest number of demand points within a fixed response time. Thus, a central question arises: Is the emergency coverage quality of the MEC(SABC) as good as the one

yielded by the MEC model?

The percentage of emergency coverage for all instances is presented with the equivalent MEC model and the SBFM in Figure **??**. Two columns with three plots each, varying the number of scenarios and the location sites. Each plot shows the type of ambulance percentage coverage obtained by the a) MEC and b) SBFM: T is for Total coverage (all required ambulances on time), TL is for Total-late coverage (all required ambulances, but at least one arrives late), P is for Partial coverage (at least one required ambulance is not dispatched, but the dispatched ones all arrive in time), PL is for Partial-late coverage (at least one required ambulance is not dispatched, at least one of the dispatched arrives late), and N for Null (no ambulances assigned to the demand point). The upper plots are for $|L| = \{16\}$ potential sites, the middle ones for $|L| = 50$, and the lower ones for $|L| = 100$.

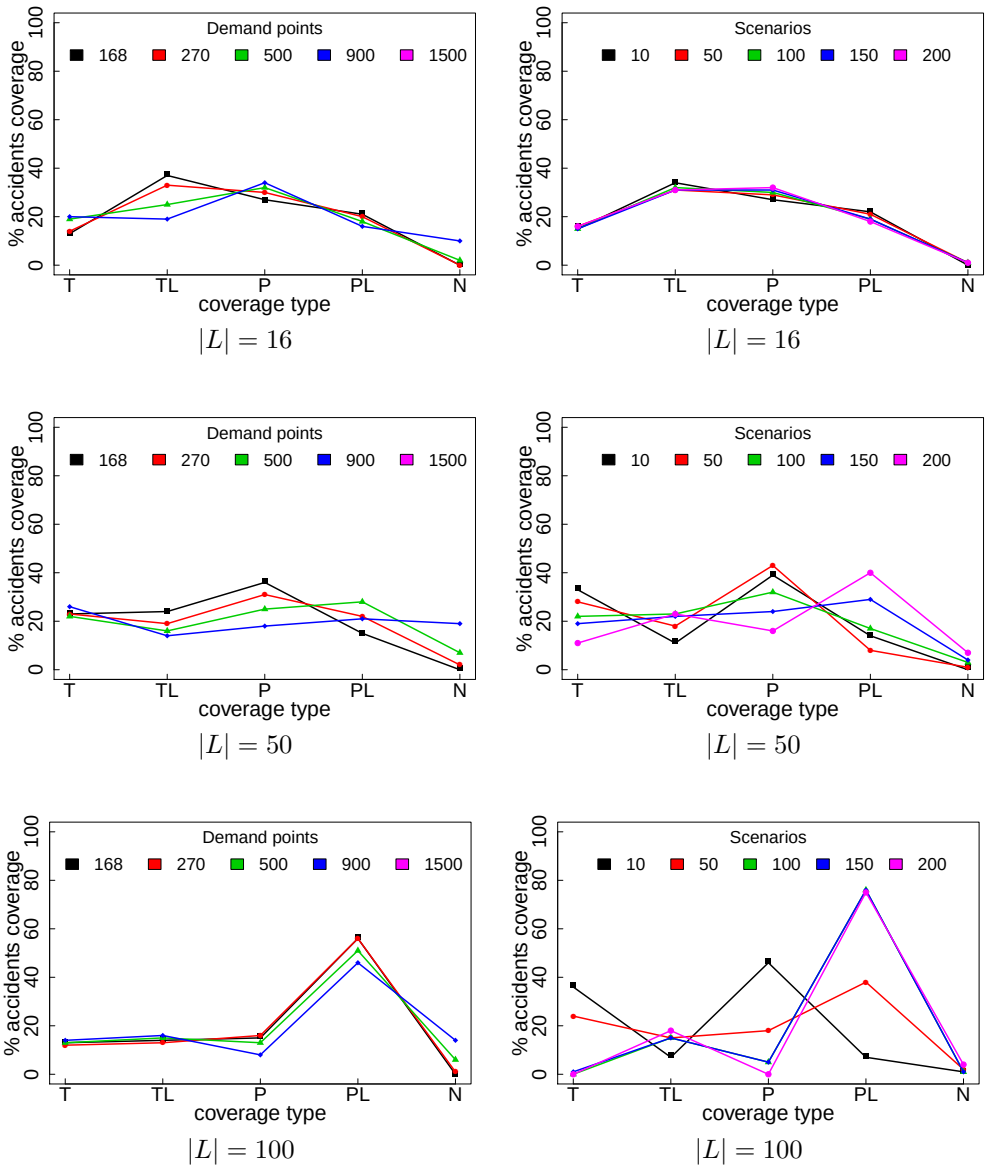


Figure 5.9: Coverage MEC.

Figure **??**a, shows that the MEC model tends to leave very few demand points with null coverage, which is the primary concern of the emergency services in our case study. As the number of potential sites $|L|$ increases, the coverage tends to be partial-late for the MEC model. This behavior is probably related to the large gaps obtained by the MEC model for large instances, but the number of null coverage is still remarkably low. Column b) shows that the SBFM is robust in terms of the number of scenarios. That is, the demand point coverage is independent of the scenario number. In this way, 100 scenarios are sufficient to handle a high-quality coverage solution. Moreover, the MEC(SABC) model inherits the characteristic of

having very few null demand point coverage from the MEC model. Interestingly, partial coverage tends to be larger than partial late coverage, which is mainly desired in real life because it can be translated into first-aid medical care on time, increasing the probability of saving lives.

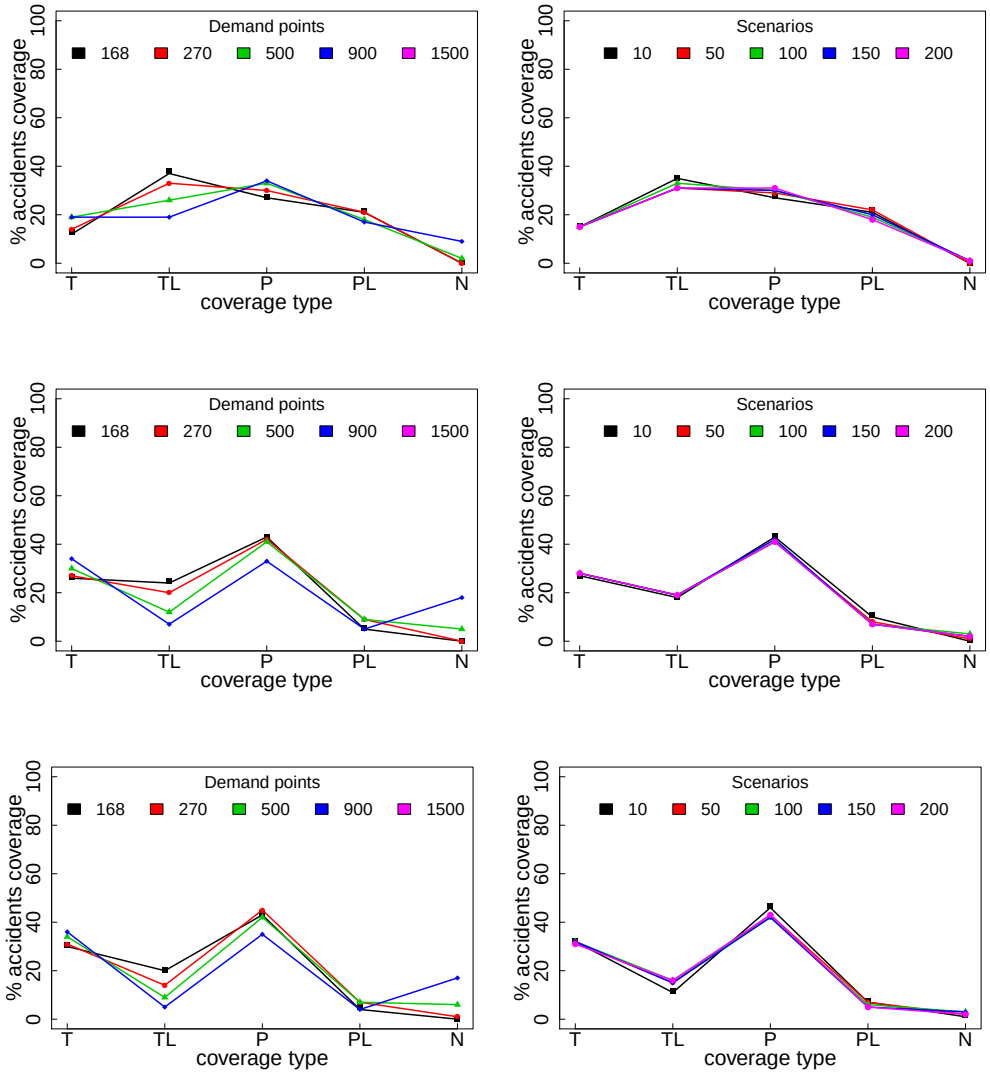


Figure 5.10: Coverage M2M1.

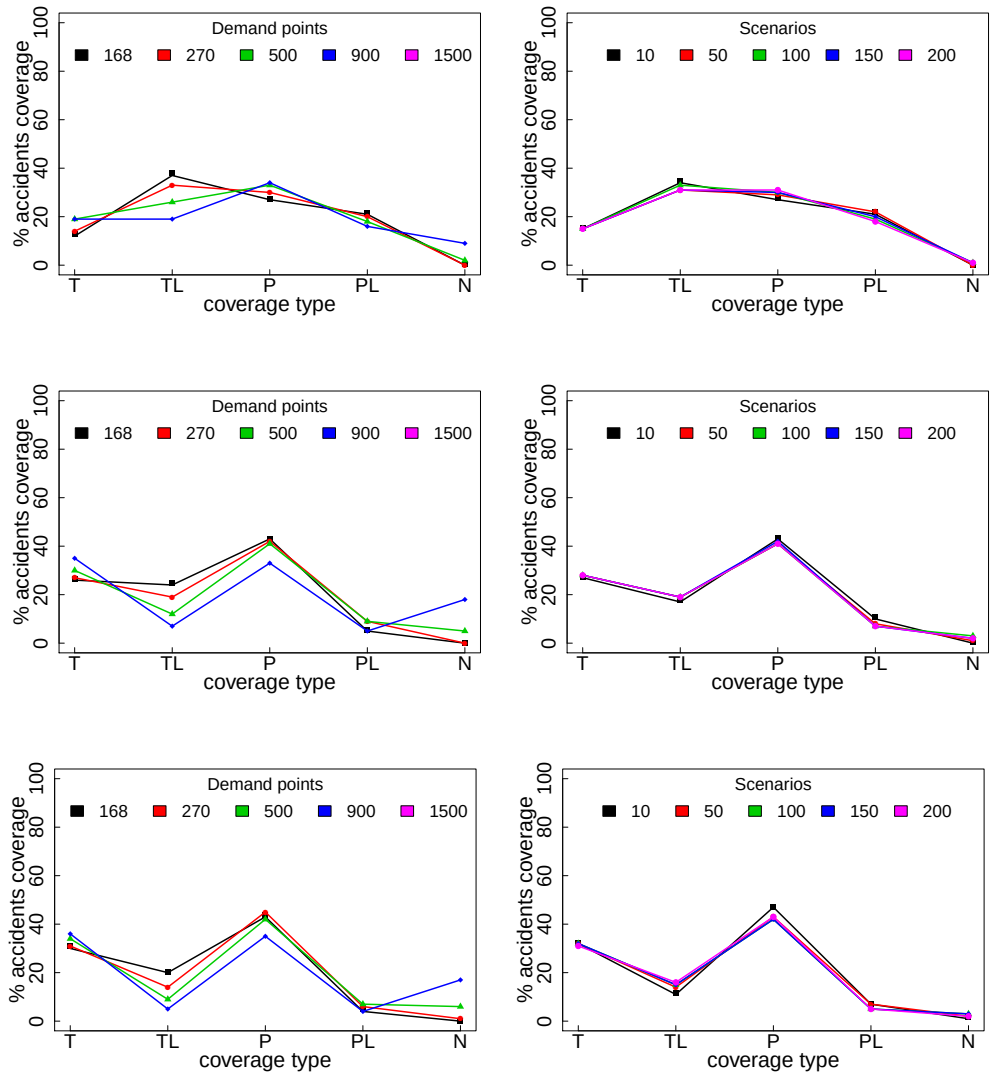


Figure 5.11: Coverage Math.

All the previous experiments were executed with the number of ambulances equal to $(\eta_1, \eta_2) = (35, 20)$. A central feature of the EVCP problem is that an ALS ambulance can be sent instead of the BLS one, which gives a more flexible setting but may induce difficulty when solving the models. Thus, what is the effect of the number of available ambulances in the EVCP problem on the objective function value and the running time?

We execute all the instances with emergency demand points fixed to 900, 100 scenarios, and 50 ambulance location sites. For this experiment, we vary the number of ambulances. Figure 5.10 shows two columns of two plots each. The objective value (upper plots) and the running time (lower plots) are on the y-axis, while the x-axis

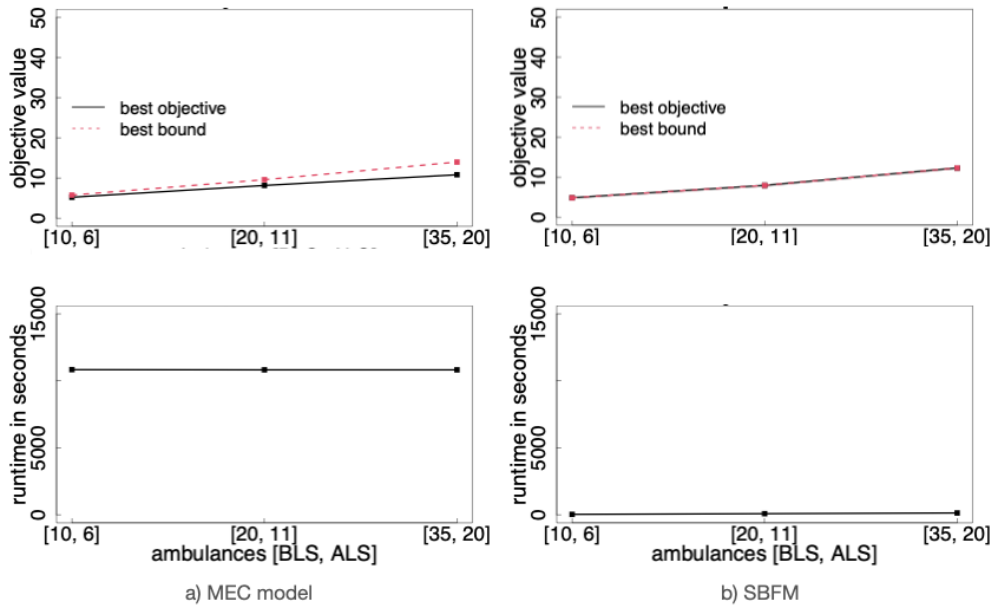


Figure 5.12: Objective value and running time versus the number of ambulances for a) MEC model and b) SBFM.

varies the number of ambulances: (η_1, η_2)=(10,6), (20,11), and (η_1, η_2)=m,(35,20). The left plots correspond to the MEC stochastic model, while the right ones are for the SBFM.

From Figure 5.10a, we observe that the difference between the best objective and the best bound for the MEC model (left-hand side plots) slightly increases with the number of ambulances. Thus, the larger the number of ambulances, the harder the instances for the MEC model. Furthermore, the time limit is reached for every tested instance of the MEC model. For the SBFM, the relative optimality gaps are equal to 0 for all instances. Moreover, the objective values are comparable to the MEC model for all different ambulance settings, which is a prominent characteristic. Furthermore, under the SBFM, all instances are solved in less than one minute, and this time is unaffected by the number of ambulances.

5.4 Coverage

5.5 Experimental analysis of the MEC, SABC, and MEC(SABC) stochastic formulations

Chapter 6

Conclusions

EMS systems in developing countries such as Mexico suffer from a shortage of ambulances. Thus, one of the main goals addressed in this work was to investigate and develop tools that allow us to decide whether an emergency can be uncovered, or totally or partially covered.

The *Emergency Vehicle Covering and Planning* (EVCP) problem consists of locating a limited number of two heterogeneous types of ambulances in different city locations and dispatching them to the emergency points so as to maximize the coverage with short medical first aid response time. In the EVCP problem, these two interrelated decisions are simultaneously considered in a novel two-stage stochastic program. The EVCP stochastic model allows for partial coverage of the accidents by the ambulances based on a decay function.

We propose a two-stage stochastic program for the EVCP problem that can be solved by branch-and-bound for small instances with a restrictive number of scenarios. We also propose a surrogate-based feedback method, which is essentially a location-allocation procedure that relies on the solution of an auxiliary surrogate model. This method is faster to solve and allows us to obtain high-quality solutions significantly faster than the previous model. The SBFM was tested over a broad set of randomly generated instances based on real-world data from a local system. An important feature of the proposed approach is that it can be implemented by calling any off-the-shelf integer solver without employing complex decomposition techniques.

Naturally, there are several lines of work that can be further investigated. For instance, one interesting aspect we observed is that there are some private EMS services that also dispatch vehicles to the accident sites. Some of these are not regulated nor coordinated by the state. In some cases, this provokes a conflict as too many ambulances arrived at the site, leaving other points unattended. This

situation could of course benefit if coordinated through decision-making tools as the ones developed here.

6.1 Main contributions and conclusions

Our stochastic model can solve large-scale instances but as the sizes of the instances increase, the convergence time becomes very long. This is why one of the main contributions of this thesis is the SAFM.

6.2 Future work

Our future work involves more than one service provider in the system, considering the differences between them and preferences that public ambulances can have compared with private ambulances.

To solve the preliminar model and future models, we will include Benders cuts or another solution method that we are studying. The difficulty is that our problem considers integer and binary variables, which are not the same as models in the literature.

Also, we want to include queues at hospitals. During the beginning of the Covid-19 pandemic, some hospitals only attended to Covid-19 patients, which caused other hospitals to have ambulance queues due to overdemand. This time wasted waiting for attention affects ambulance availability and has to be counted in the EMS system.

Bibliography

[1] L. Aboueljinane, E. Sahin, and Z. Jemai. A review on simulation models applied to emergency medical service operations. *Computers & Industrial Engineering*, 66(4):734–750, 2013.

[2] M. Amorim, S. Ferreira, and A. Couto. How do traffic and demand daily changes define urban emergency medical service (uems) strategic decisions?: A robust survival model. *Journal of Transport & Health*, 12:60–74, 2019.

[3] S. Ansari, L. A. McLay, and M. E Mayorga. A maximum expected covering problem for district design. *Transportation Science*, 51(1):376–390, 2015.

[4] R. Aringhieri, M. E. Bruni, S. Khodaparasti, and J. T. van Essen. Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, 78:349–368, 2017.

[5] G. Bakalos, M. Mamali, C. Komninos, E Koukou, A. Tsantilas, S. Tzima, and T. Rosenberg. Advanced life support versus basic life support in the pre-hospital setting: A meta-analysis. *Resuscitation*, 82(9):1130–1137, 2011.

[6] D. Bandara, M. E. Mayorga, and L. A. McLay. Optimal dispatching strategies for emergency vehicles to increase patient survivability. *International Journal of Operational Research*, 15(2):195–214, 2012.

[7] V. Bélanger, A. Ruiz, and P. Soriano. Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. *European Journal of Operational Research*, 272(1):1–23, 2019.

[8] P. Beraldi and M. E. Bruni. A probabilistic model applied to emergency service vehicle location. *European Journal of Operational Research*, 196(1):323–331, 2009.

[9] O. Berman, D. Krass, and Z. Drezner. The gradual covering decay location problem on a network. *European Journal of Operational Research*, 151(3):474–480, 2003.

[10] D. Bertsimas and Y. Ng. Robust and stochastic formulations for ambulance deployment and dispatch. *European Journal of Operational Research*, 279(2): 557–571, 2019.

[11] R. Boujemaa, A. Jebali, S. Hammami, A. Ruiz, and H. Bouchriha. A stochastic approach for designing two-tiered emergency medical service systems. *Flexible Services and Manufacturing Journal*, 30(1):123–152, 2018.

[12] O. Braun, R. McCallion, and J. Fazackerley. Characteristics of midsized urban EMS systems. *Annals of Emergency Medicine*, 19(5):536–546, 1990.

[13] L. Brotcorne, G. Laporte, and F. Semet. Ambulance location and relocation models. *European journal of operational research*, 147(3):451–463, 2003.

[14] J. C. Dibene, Y. Maldonado, C. Vera, M. de Oliveira, L. Trujillo, and O. Schütze. Optimizing the location of ambulances in tijuana, mexico. *Computers in biology and medicine*, 80:107–115, 2017.

[15] R. D. Galvao and R. Morabito. Emergency service systems: The use of the hypercube queueing model in the solution of probabilistic location problems. *International Transactions in Operational Research*, 15(5):525–549, 2008.

[16] B. C. Grannan, N. D. Bastian, and L. A. McLay. A maximum expected covering problem for locating and dispatching two classes of military medical evacuation air assets. *Optimization Letters*, 9:1511–1531, 2015.

[17] P. J. H. Hulshof, N. Kortbeek, R. J. Boucherie, E. W. Hans, and P. J. M. Bakker. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems*, 1(2):129–175, 2012.

[18] O. Karasakal and E. K. Karasakal. A maximal covering location model in the presence of partial coverage. *Computers & Operations Research*, 31(9):1515–1526, 2004.

[19] X. Li, Z. Zhao, X. Zhu, and T. Wyatt. Covering models and optimization techniques for emergency response facility location and planning: a review. *Mathematical Methods of Operations Research*, 74(3):281–310, 2011.

[20] L. A. McLay. A maximum expected covering location model with two types of servers. *IIE Transactions*, 41(8):730–741, 2009.

[21] L. A. McLay and M. E. Mayorga. Evaluating emergency medical service performance measures. *Health care management science*, 13(2):124–136, 2010.

[22] L. A. McLay and M. E. Mayorga. Evaluating the impact of performance goals on dispatching decisions in emergency medical service. *IIE Transactions on Healthcare Systems Engineering*, 1(3):185–196, 2011.

[23] L. A. McLay and H. Moore. Hanover county improves its response to emergency medical 911 patients. *Interfaces*, 42(4):380–394, 2012.

[24] Nimrod Megiddo and Arie Tamir. On the complexity of locating linear facilities in the plane. *Operations research letters*, 1(5):194–197, 1982.

[25] S. Nickel, M. Reuter-Oppermann, and F. Saldanha-da Gama. Ambulance location under stochastic demand: A sampling approach. *Operations Research for Health Care*, 8:24–32, 2016.

[26] N. Noyan. Alternate risk measures for emergency medical service system design. *Annals of Operations Research*, 181:559–589, 2010.

[27] C. O'Keeffe, J. Nicholl, J. Turner, and S. Goodacre. Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest. *Emergency Medicine Journal*, 28(8):703–706, 2011.

[28] M. Reuter-Oppermann, P. L. van den Berg, and J. L. Vile. Logistics for emergency medical service systems. *Health Systems*, 6(3):187–208, 2017.

[29] L. Shaw, S. K. Das, and S. K. Roy. Location-allocation problem for resource distribution under uncertainty in disaster relief operations. *Socio-Economic Planning Sciences*, 82:101232, 2022.

[30] I. Sung and T. Lee. Scenario-based approach for the ambulance location problem with stochastic call arrivals under a dispatching policy. *Flexible Services and Manufacturing Journal*, 30:153–170, 2018.

[31] H. Toro-Díaz, M. E. Mayorga, S. Chanta, and L. A. Mclay. Joint location and dispatching decisions for emergency medical services. *Computers & Industrial Engineering*, 64(4):917–928, 2013.

[32] H. Toro-Díaz, M. E. Mayorga, L. A. McLay, H. K. Rajagopalan, and C. Saydam. Reducing disparities in large-scale emergency medical service systems. *Journal of the Operational Research Society*, 66(7):1169–1181, 2015.

[33] M. van Buuren, R. van der Mei, and S. Bhulai. Demand-point constrained ems vehicle allocation problems for regions with both urban and rural areas. *Operations Research for Health Care*, 18:65–83, 2018.

[34] J. Wang, H. Liu, S. An, and N. Cui. A new partial coverage locating model for cooperative fire services. *Information Sciences*, 373:527–538, 2016.

[35] Y. Wang, K. L. Luangkesorn, and L. Shuman. Modeling emergency medical response to a mass casualty incident using agent based simulation. *Socio-Economic Planning Sciences*, 46(4):281–290, 2012.

[36] S. Yoon, L. A. Albert, and V. M. White. A stochastic programming approach for locating and dispatching two types of ambulances. *Transportation Science*, 55(2):275–296, 2021.

[37] Y. Zhang, Z. Li, and Y. Zhao. Multi-mitigation strategies in medical supplies for epidemic outbreaks. *Socio-Economic Planning Sciences*, 87:101516, 2023.

[38] Z. Zhou, D. S. Matteson, D. B. Woodard, S. G. Henderson, and A. C. Micheas. A spatio-temporal point process model for ambulance demand. *Journal of the American Statistical Association*, 110(509):6–15, 2015.

Resumen autobiográfico

Beatriz Alejandra García Ramos

Candidato para obtener el grado de
Doctorado en Ciencias en Ingeniería de Sistemas

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica

Tesis:
Stochastic methodologies for locating and dispatching two types of ambulances with partial coverage

Nací el 20 de enero de 1995 en el municipio de San Nicolás de los Garza en el estado de Nuevo León. Mis padres Jesús García Gámez y Beatriz Ramos Larralde me han cuidado y educado desde mi nacimiento, al igual que a mi hermana Karina Guadalupe García Ramos. Concluí mis estudios como Licenciado en Matemáticas en junio del año 2017 en la Facultad de Ciencias Físico Matemáticas perteneciente a la Universidad Autónoma de Nuevo León. Obtuve mi grado como Maestro en Ingeniería de Sistemas en noviembre de 2019 en la Facultad de Ingeniería Mecánica y Eléctrica perteneciente a la Universidad Autónoma de Nuevo León, en donde también inicié mis estudios de Doctorado en Ingeniería en Sistemas en agosto del año 2020.