Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado

Locating and Dispatching Two Types of Ambulances Considering Partial Coverage: Stochastic Integer Programming Models and Heuristics

por

Beatriz Alejandra García Ramos

como requisito parcial para obtener el grado de

doctora en ciencias en ingeniería de sistemas

30 de Septiembre de 2025

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado

Locating and Dispatching Two Types of Ambulances Considering Partial Coverage: Stochastic Integer Programming Models and Heuristics

por

Beatriz Alejandra García Ramos

como requisito parcial para obtener el grado de

doctora en ciencias en ingeniería de sistemas

30 de Septiembre de 2025

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
Facultad de Ingeniería Mecánica y Eléctrica
Posgrado

Los miembros del Comité de Evaluación de Tesis recomendamos que la tesis "Locating and dispatching two types of ambulances considering partial coverage: stochastic integer programming models and heuristics", realizada por la estudiante Beatriz Alejandra García Ramos, con número de matrícula 1550385, sea aceptada para su defensa como requisito parcial para obtener el grado de Doctorado en Ingeniería de Sistemas.

El Comité de Evaluación de Tesis

Dr. Roger Zirahuén Ríos Mercado
Director

Dra. Yasmín Águeda Ríos Solís
Co-directora

Dra. Iris Abril Martínez Salazar
Revisora

Dr. Vincent André Lionel Boyer
Revisor

Dr. César Emilio Villarreal Rodríguez
Revisor

Dra. Yajaira Cardona Valdés
Revisora

Dra. Irma Delia García Calvillo
Revisora

Vº. Bº.

Dr. Simón Martínez Martínez
Subdirector de Estudios de Posgrado

Institución	190001
Programa	557620
Acta núm.	372

Ciudad Universitaria, a 2 de febrero de 2026.

Contents

List of Figures

List of Tables

Acknowledgments

Agradezco infinitamente a Dios por el don de la vida, porque siempre pone en mi camino a las personas indicadas para crecer espiritual, profesional y personalmente día con día. Dios siempre ha sido mi refugio, mi fuerza, y siempre me muestra que no estoy sola porque me acompaña en todo momento.

Agradezco a mi madre Beatriz Ramos Larralde por estar al pendiente de mí en todo momento, porque se ha preocupado al ver mis desveladas, mis comidas a destiempo, mi cansancio, y mi ajetreo en mis días, y me ha apoyado de manera en que yo tenga menos preocupaciones, escuchándome, alentándome, y con sus deliciosas comidas que me ahorran tiempo para poder utilizarlo más en escribir esta tesis. Agradezco de igual manera a mi padre Jesús García Gámez, que se preocupa por mi bienestar, porque está al pendiente en todo momento de que no me falte nada y no duda en ofrecerme su ayuda a cada instante. Gracias a ellos dos es que he podido seguir adelante con menos preocupaciones, sabiendo que cuento con ellos siempre que los necesite en cualquier aspecto de mi vida.

Agradezco a mi hermana Karina Guadalupe García Ramos, que siempre me ayuda a aliviar mis tensiones con sus pláticas. Con ella he crecido como persona y como profesional, y con una que otra cosa de emprendimiento. Ella siempre ha estado con sus consejos, y aunque en ocasiones discutimos como cualquier par de hermanas, estamos la una para la otra.

Agradezco también a toda mi familia, tanto de sangre como aquella que he creado a través de vínculos. Gracias a mi segundo hogar con mi familia Beltrán: Martha Beltrán y Raúl Candelaria, que me ven como si fuera su propia hija y se preocupan por mí tanto como por mi hermana Valeria Beltrán, a quien amo también como a mi hermana de sangre. Mis primas García, que me hacen reír todo el tiempo y me ayudan a despejar mi mente y descansar un rato de las preocupaciones, aparte de mi familia Ramos, que ha estado al pendiente de mí y preguntándome cómo va mi progreso. Pero sobre todo les agradezco a todos por aguantar mis momentos de trabajo incluso estando en noches de fiestas y cenas familiares, y porque han comprendido que a veces no he podido acompañarlos porque esto es un trabajo 24/7.

Agradezco a todos mis grupos de amigos. Arturo, Paola y Obed, que han estado conmigo tantos años y han vivido junto conmigo muchos momentos que no se olvidan y que son anécdotas para contar durante mucho tiempo. A mis amigos hechos desde la maestría: Alberto, Delavy, Gabriela, Mayra, Citlali, Pablo, Astrid y Alan; ellos, más que nadie, entienden lo que es hacer investigación y siempre están dispuestos a ayudarme en mis dudas. Mis amigos del Tecnológico de Monterrey: Salvador, Saúl y Emmanuel, que me dan sus buenas vibras siempre y me llenan de risas. Amigos, gracias porque sé que a pesar de la distancia la amistad sigue. Mis amigos Kefas: Aliber, Adrián, Priscila, Óscar, Deba; que llegaron a mi vida en un momento en donde yo necesitaba volver a Dios y me han ayudado a crecer espiritualmente. A mis dos personas favoritas de San Jorge Mártir: Josué Loredo y Pedro Ramírez, quienes me han escuchado siempre, me han guiado, me han alegrado los días, me han dado paz y mucho cariño, siendo un gran ejemplo de Jesús a través de su sacerdocio. A Victor Domínguez, que ha estado a diario al pendiente de todo lo que pasa en mi vida y he contado con su gran amistad durante tanto tiempo, siempre siendo mi apoyo incondicional. A todos ustedes los quiero mucho, los llevo en mi corazón todos los días y espero que estén en mi vida siempre.

Agradezco a mis colegas y compañeros, y a todas las personas que han estado organizando los congresos nacionales e internacionales a los que he asistido, me ayudan a crecer profesionalmente.

Agradezco a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por la beca de manutención que me fue otorgada, con la cual pude facilitarme mis estudios al utilizar ese ingreso en transporte, alimentos e incluso en la investigación que durante cuatro años estuve realizando. Además, parte de mi investigación fue financiada por el Proyecto CF2023-I-880 del Programa de Ciencia de Frontera de la SECIHTI.

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 donde trabajar y salones donde podamos cursar las asignaturas que se nos imparten.

Agradezco a los Doctores que se encuentran en el Programa de 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, la Dra. Iris Abril Martínez Salazar, el Dr. César Emilio Villarreal Rodríguez, el Dr. Vincent André Lionel Boyer, la Dra. Yajaira Cardona Valdés y la Dra. Irma Delia García Calvillo, por haber leído mi tesis y haberme apoyado en las correcciones de la misma. De manera especial, agradezco a mi director Roger Z. Ríos Mercado y a mi co-directora Yasmín A. Ríos Solís que durante toda mi investigación estuvieron apoyándome y dirigiendo la misma.

Abstract

Beatriz Alejandra García Ramos.

Candidata para obtener el grado de Doctora en Ciencias en Ingeniería de Sistemas.

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: Locating and Dispatching Two Types of Ambulances Considering Partial Coverage: Stochastic Integer Programming Models and Heuristics.

Número de páginas: 101.

The thesis aims to study a decision-making problem on the Emergency Medical Service (EMS) systems and design and implement algorithms to solve it. The objective is to improve Mexico's 9-1-1 call emergency system by locating and dispatching ambulances to maximize patient attention at the minimum possible response time. The system is subject to uncertainty, as we do not know when, where, or how many emergencies there will be in the system. That is why we consider this uncertainty as part of the modeling assumptions. The problem we studied is referred to as the Emergency Vehicle Covering and Planning Problem (EVCPP), which is mathematically modeled as a two-stage integer stochastic program. In the first stage, the location of the ambulances must be determined prior to the occurrence of the accidents. Then, after the accidents become known, the ambulance dispatching decisions are taken in the second stage.

In order to address this problem, a mathematical model aiming at maximizing expected coverage is initially proposed, which is tested on relatively small instances with nearly 100% success. As the size of the instances increases, the gap between the best solution found and the best bound found increases. Consequently, to address larger instances, a methodology referred to as Surrogate-based Feedback is introduced. This methodology makes ambulance dispatch decisions without taking into account the type of coverage considered in the first model. Efficiently and optimally determines ambulance locations and dispatches; however, it cannot be guaranteed that these represent the optimal solutions to the EVCP problem, but it yields improvements when applied to larger instances. Lastly, a heuristic algorithm, named Scenario-based Matheuristic, is proposed and evaluated across instances of varying sizes. The results indicate that, for larger instances, although improvements are observed, they are comparatively modest compared with those achieved by the second methodology.

One of the main contributions is the introduction of partial coverage to this type of model. When discussing accident coverages in general, especially in problems that handle stochastic programming due to the uncertainty present in the EMS system, it usually refers to whether a demand point is covered or not. The uncertainty already makes this problem challenging, and managing more decisions related to different types of coverage further complicates it. However, we consider that a demand point can be uncovered or covered in various manners, which we categorize as total, total-late, partial, and partial-late, depending on the type of ambulances dispatched and the time it takes for them to reach each demand point.

The novelty of involving different types of coverage in our problem is that, in addition to considering various scenarios having variability to address the existing uncertainty, we provide the system with the flexibility to have greater ambulances' availability to attend to more patients, even if it is as a first response. After stabilization, patients can receive more appropriate care, either by an ambulance with better equipment that arrives later or by being transported to a hospital.

Another contribution is the proposal of a surrogate model, which enables us to develop a surrogate-based solution technique. This methodology makes decisions about the location and dispatching of ambulances without considering the type of coverage present at each demand point. By not considering partial coverage in this methodology, an initial solution for the first model can be obtained in a shorter computational time. Once this solution is achieved, the location of the ambulances is taken and their quality is evaluated considering the dispatching that would occur if the type of coverage from the first model is considered. The advantage of this methodology is that there are improvements in the first model solution when larger instances are considered. In addition, we also propose a heuristic algorithm, in which we consider enhancing the solutions of this methodology.

The solution method was fully assessed in a wide collection of problem instances.

Firma del director: ______________________________

Dr. Roger Z. Ríos Mercado

Firma del co-director: ______________________________

Dra. Yasmín Á. Ríos Solis

Chapter 1

Introduction

Emergency Medical Services (EMS) systems provide medical care for people who suffer a medical incident. These systems control emergency call services received at the emergency number established for emergencies, commonly 9-1-1. Typically, 9-1-1 emergency systems consist of two phases. As we can see in Figure 1.1, the first phase is the response to an emergency call: an operator responds to the call and identifies the type of emergency, such as medical, security, or fire. The operator asks some questions to identify the type of emergency (dismissing prank calls) [46]. If the patient needs medical attention, the operator contacts an ambulance (usually the nearest from a public institution) 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 equipped with the material resources needed to attend to the patient, the ambulance leaves its base, arrives at the scene, treats the patient, leaves the scene and arrives at a hospital (commonly the nearest) if necessary, and finally returns to its base to wait for another emergency call [24].

The complex decision-making process involved in EMS systems have motivated a plethora of research work on operational research models and algorithms in the last decades [1, 8, 58]. Scientists are concerned about the impact of calls emergency'

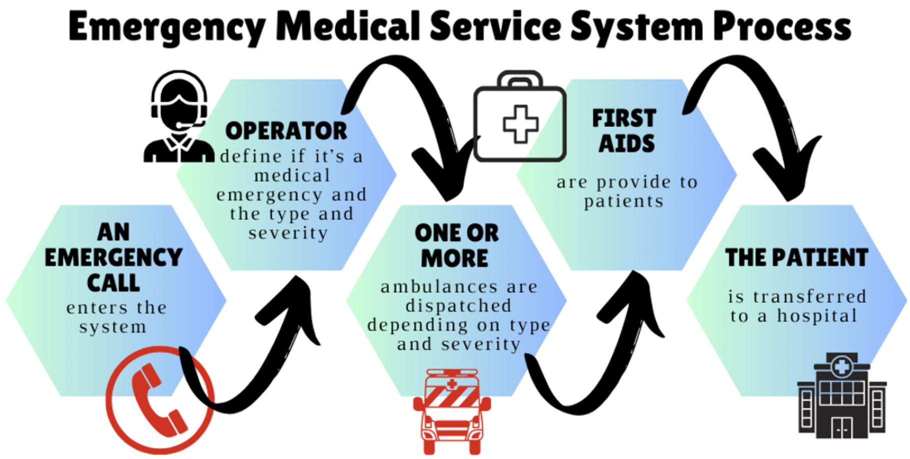


Figure 1.1: An Emergency Medical Service System Process.

average response time for attending a patient who suffers a medical incident. Moreover, the cost of purchasing material resources, medical vehicles, or building a new medical facilities, among other things, can limit patient service [39]. Lack of human and material resources can cause insufficient attention in patients [2, 32, 42].

One of the most studied performance measures is reducing the average response time when an emergency call arrives at a call center and someone needs medical attention [17]. The objective is to provide, as soon as possible, the initial treatment for a patient who has a medical problem caused by an accident, trauma, or a natural disaster to reduce the mortality of the patients. Clearly, there are some situations where shorter response times lead to better survival rates. Another objective that EMS system problems consider is to maximize coverage to handle all emergency calls that enter the system [22]. In addition, there are some problems that consider improving patient survival or reducing the patients' mortality [73].

Our interest is in the EMS systems within the Mexican Healthcare System. In Mexico, there exists the 9-1-1 number controlled by the C-5 organization (*Integrated Coordination, Control, Command, Communications, and Computing Center*), 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. In addition, a doctor can continue the call to guide the person on the phone if the patient needs immediate attention while the ambulance arrives. Paramedics take care of the patient on site and then transfer the patient to a hospital if needed [29].

The particular problem under study has a number of features that are further discussed in detail in Chapter 3. Some of these involve handling two different types of ambulances, consideration of partial coverage, scenarios for accidents' uncertainty, response time from potential sites to demand points, among others. For this particular problem, we propose a two-stage stochastic integer programming model with recourse for ambulance location and dispatching, considering two types of vehicles 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 the factors that affect the EMS system in Mexico (Chapter 3). In addition, we describe (Chapter 4.3) the instance generation and experimental assessments. Then, we solve the problem and define the model (Chapter 4.5) used to do the experiments. Finally, we show conclusions (Chapter 5) that we obtain from experiments described in the previous section. In this section, we also propose future work.

1.1 Motivation

Our interest is to improve the Emergency Medical Service System in Mexico, particularly in Nuevo León. World Health Organization establishes that there must be four ambulances per 1,000 people, which is not available in all states of our country.

Due to the lack of available ambulances, emergency calls are answered late. However, buying more ambulances so that there are more available to distribute is not sufficient [56]. 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 must locate a limited number of ambulances of two heterogeneous types at different city points and dispatch them to the sites where accidents occur. Our problem considers the uncertainty of the accident (demand points). Our goal is to optimize total and partial coverage, in which patients receive medical first aid. We propose a two-stage integer stochastic program for this problem. In the first stage, the location of a limited number of ambulances of two types is decided. In the second stage, the dispatching of the ambulances to accidents is determined. This stochastic model allows for partial coverage of accidents by ambulances based on a decay function.

1.3 Hypothesis

The hypotheses of this thesis focus on the implementation of partial coverage within an EMS system under uncertainty and the use of a surrogate model for an initial approach to an optimal response. These hypotheses are detailed below:

- The introduction of different types of coverage in a stochastic integer programming problem allows for a greater reach in patient coverage by expanding the ability to cover more demand points when considering partial coverage. Using this model, which focuses on the type of coverage for each demand point, the number of uncovered points can be minimized, thus potentially saving more lives.

- A surrogate model for this problem allows to significantly reduce the computational time required to obtain an initial response and can serve as an initial solution for further improvement.

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 problem under study includes different factors that affect the system. Those factors are:

- The gravity of the accident, which can be categorized into basic care, which refers to those that do not pose a threat to the patient's life, and advanced care, which requires immediate first aid to stabilize the patient and prevent life-threatening conditions;
- Two different ambulance types, which are ambulances for basic life support (BLS) and ambulances for advanced life support (ALS);
- There is variation in demand points depending on the day of the week and the hour of the day, which can be considered different scenarios.

The objectives of solving the problem under study are the following:

- To provide a scenario-based stochastic program integer model with resources considering partial coverage that includes all factors we described above. The objective of this model is to maximize the expected coverage for the accidents.
- To develop a solution methodology for large scale instances, which achieves solutions in reduced computational time. This methodology obtains a feedback approach for the model proposed.
- To implement a heuristic algorithm that improves the solution found in the previous methodology, which features different neighborhoods that should be explored to achieve more significant variability in the decisions regarding the location of ambulances.

1.5 Contributions

Summary of research contributions:

- We present a maximum expected coverage model considering partial coverage. To this end, we use a decay function to handle partial coverage. Although decay functions have been used in facility location problems, to the best of our knowledge, this is the first time they have been used in a stochastic ambulance location problem.
- We develop a surrogate model that is faster to solve, facilitating the solution of the problem.
- We propose a location-allocation method based on intelligent exploitation of the surrogate model solution.
- We introduce a heuristic approach to improve the surrogate model solutions.

- We present empirical evidence that shows: (i) the benefit of considering gradual coverage, (ii) the value of the stochastic solution, and (iii) the effectiveness of the proposed solution method.

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 [6, 11, 60]. When scientists talk about EMS systems, many terms are used to explain the problem. Two of these terms are demand points and potential sites for locations. Demand points are sites where an accident occurs. Typically, 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. In our particular case, we consider two types: 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 treatment of the patients by paramedics and their transport to a hospital [6].

EMS systems in developing countries, such as Mexico, lack around 30-60% [26, 56] of the number of ambulances suggested by the World Health Organization (WHO), which is at least four ambulances per 100,000 people [19]. For the Cruz Roja, an EMS operating with this small number of ambulances is considered similar to a war situation. Thus, one of the main contributions of this work is to deal with the problem of deciding whether an emergency will be totally or partially covered. Sadly, some emergencies may remain uncovered by an emergency unit.

In general, the main objective of emergency vehicle planning problems is to reduce the average response time of a patient's initial treatment administered by a paramedic in an emergency [4, 9, 67, 68]. In fact, the speed and number of ambulances dispatched to 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 taken care of. Every minute of delay in treatment in a cardiac, for instance, reduces the probability of survival by 24% [57].

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 [20], and can be static and dynamic problems. The first problem that we studied is one of the static and stochastic ones.

There are many models that have been developed to address EMS systems. These can be divided into deterministic and stochastic models. Within models that handle uncertainty, there are models based on queuing theory concepts and others based on mathematical programming approaches. Models are also classified as static or dynamic. All of these different approaches involve different solution methodologies [20]. In particular, in this thesis, we consider a stochastic static model modeled as a stochastic integer programming model.

2.1 EMS models

In the literature, two important types of models are used to address issues in EMS systems: static models and dynamic models. Static models represent a specific moment in time and do not consider changes that occur over time. In contrast, dynamic models include time as a variable and can simulate how a system evolves over time.

Static models are employed within EMS systems to allocate ambulances that remain stationed at their bases. These models typically utilize a response time to evaluate whether a demand location is adequately serviced. By focusing on a specific point in time, static models improve the decision-making process for ambulance deployment [13]. However, they often overlook variations in demand and the need for ambulance relocation in dynamic circumstances. These limitations can be effectively addressed through mathematical programming, queuing theory, and simulation models, some of which are stochastic while others operate based on discrete events [1, 45].

In this section, we will explore static models in detail, highlighting their essential role in improving ambulance localization. Additionally, we will survey the most significant models relevant to our investigation. Accurate localization is critical for emergency response, as it directly impacts response times and patient outcomes.

2.1.1 Deterministic models

There are two early models for static problems: Location Set Covering Model (LSCM) [62] and Maximal Covering Location Problem (MCLP) [23], 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.

Deterministic models were proposed to solve static problems because sometimes emergency calls need to be attended for different vehicle types. Most of the demand points are covered once, like Backup Coverage Problems [30, 38] or the Double Standard Models [27, 31, 33], which uses two different radii of coverage [47]. Alternative deterministic models are the tandem equipment allocation model or the Facility-Location Equipment-Emplacement Technique model [63], which consider two types of vehicles (one for BLS and another for ALS), 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 their research, Berman et al. [14] introduced a decay function to classify coverage into three categories: full, none, and partial coverage, within a generalized the 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.

Karasakal and Karasakal [43] introduced partial coverage for the MCLP problem. This problem aims to maximize the 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, and they define their problem as the MCLP-P problem. They propose a monotonic decay function that decreases according to the distance between the facility and the demand point for partial coverage. The distance between a facility and a demand point must be less than 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 the facility and the demand point is greater than or equal to a maximum partial coverage distance. To solve the MCLP-P problem, they used a Lagrangian relaxation based solution procedure.

Jian et al. [41] 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. The partial distance is calculated with a decay function that decreases according to the increase in vehicle response time. Quantity coverage determines whether an emergency is fully served or not, comparing the number of vehicles dispatched with the necessary quantity. For this problem, the authors consider demand priority to know where vehicles must be located and the patient's classification to decide how to dispatch them.

Comprehending the initial static location models is essential for our research because they underscore the primary challenges in optimizing emergency services. These models, including their advancements to deterministic and partial coverage approaches, reveal problems that require more advanced solutions. The evolution from basic coverage models to more complex ones, which factor in different vehicle types, demand prioritization, and decay functions, emphasizes the importance of addressing challenges in emergency services logistics effectively.

2.1.2 Probabilistic models

There are two classes of models that attempt to handle uncertainty. We called probabilistic models to those who introduce the modeling of some probabilities such as probability that an ambulance is busy or available at a given time.

The pioneer models of this sort can be found since the 1980s. The Maximum Expected Covering Location problem was introduced by Daskin [25]. In this model, the authors use probabilities to determine if potential sites can operate and cover demand points, considering the same probability for all potential sites. By incorporating these probabilities, the model can more effectively evaluate various situations and optimize ambulance placement. An extension of this problem is the Adjusted Maximum Expected Covering Location model [10], which considers different busy probabilities for each potential site to dispatch the closest available unit. All these models can use the hypercube queueing model to calculate the busy fraction [36].

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 (MALP) I, which considers that the busy fraction is the same for all potential sites. Additionally, there is the MALP II, which utilizes the hypercube model and assumes different busy fractions for each potential site [61]. Moreover, there is the Queuing MALP, which employs a queuing theory model to determine server availability [49].

There exist more probabilistic models introduced in the 1990s. The first is an extended version of the LSCM called Rel-P [8]; this version considers that more than one ambulance can be located at the same potential site, but each potential site has a probability to have ambulances that are available to respond to a call and considers the probability of the busy fraction, too.

The second model is the two-tier model, which considers two types of vehicles to allocate at potential sites (BLS and ALS) [48], 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.

McLay and Mayorga [52] 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 focuses on considering a new approach to calculate the Response Time Threshold (RTT), a class of EMS performance measure. The approach uses the patient survival rate, considering that patients have 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 and distributed in Hanover's rural and urban areas. This model optimizes the location of ambulances on potential sites to maximize patient survival.

When talking about optimizing EMS systems, one can also speak about dispatching. Bandara et al. [9] 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.

Toro-Díaz et al. [67] 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 Decision 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 computational time. This genetic algorithm was applied to Hanover, Virginia, and when they have midsize problems, the nearest dispatch rule is the best solution. It can vary depending on the zone where it is applied.

Enayati et al. [28] continue the investigation mentioned before incorporating different levels of call emergencies in the decision-making process. They employed a multicriteria optimization approach to analyze solutions derived from various fairness objective functions while simultaneously addressing efficiency within a realistic service system characterized by multiple demand types. Authors allow that more than one available ambulance per location site can cover each demand zone, considering a preference dispatching list. A genetic algorithm is used with a queuing submodel that accounts for busy probabilities of the ambulances.

Amorim et al. [4] involve a simulation considering an initial solution to decide if ambulances have to stay at the potential sites established 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 management provides a more practical and realistic approach for addressing uncertainty [3, 59, 72]. Probabilistic models often require assumptions about the likelihood of various events, such as the busy fraction of ambulances, which can be challenging to estimate accurately and may not adequately capture the complexities of the real world. In contrast, scenario-based models enable the incorporation of various demand and traffic conditions, allowing for more realistic and flexible planning. Scenario-based models enable the design of emergency response systems that are more operationally stable and adaptable to varying conditions, leading to superior dispatching performance, as illustrated in the next section [44, 65].

2.1.3 Stochastic programming models

A second class of models handling uncertainty are those based on stochastic programming. In this type of models, the different scenarios have an associated probability of occurrence that is explicitly modeled. The random parameters are assumed to be drawn from a probability distribution that may be continuous or discrete. Typical models are stochastic programming models with recourse, robust optimization problems, or chance/constrained models. If the model employs integer variables, it is called a stochastic integer program. In this section, we survey the most relevant works based on stochastic programming approaches to EMSs.

In Boujemaa et al. [18], a two-stage stochastic model with recourse is proposed for solving an ambulance location-allocation problem that decides if an ambulance base station is open and the number of ambulances to locate at each station. The first stage of the model determines where to open ambulance stations with a fixed cost to open 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 algorithm that allows computing lower and upper bounds for problem solutions and providing the corresponding optimality gaps.

In 2019, Bertsimas and Ng [15] implemented a stochastic and robust formulation for ambulance deployment and dispatch for a problem constructed as a graph. These formulations were compared with Maximum Expected Covering Location problem and Maximal Availability Location 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 developed a column and constraint algorithm.

As an extension of the Double Standard Model mentioned in subsection 2.1.1, Dibene et al. [27] created the Robust Double Standard Model. They added demand scenarios to the original Double Standard Model 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 at different points of the city that are not the original bases.

In 2021, Yoon et al. [71] 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 dispatches ambulances from the places where they were located to demand points when a call arrives. The objective is to maximize 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 by 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. A Sample Average Approximation deterministic equivalent formulation solved this problem for small data, while a Branch-and-Benders-Cut algorithm solved a large-scale problem. And they did another problem version considering non-transport vehicles that 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 in this thesis, 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 [12] and Noyan [55] induce a reliability approach using probabilistic constraints. Nickel et al. [54] minimize the total cost of locating ambulances while ensuring a minimum level of coverage. 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 [15] 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 in the system.

2.2 Heuristics, metaheuristics and matheuristics

Recent advancements in the field of heuristics have significantly influenced the approach to stochastic combinatorial optimization problems, which are defined by their inherent uncertainty and discrete nature. In the past decades traditional optimization methods often fall short in effectively addressing these challenges; hence, heuristics have emerged as a powerful alternative. Recent trends highlight the development of hybrid metaheuristics that combine multiple approaches or integrate exact methods, adaptive metaheuristics that dynamically adjust parameters based on intermediate results, and multi-objective optimization techniques to manage conflicting objectives [35].

In terms of EMSs, we have seen some successful heuristic approaches. Mayorga et al. [50] proposed a constructive heuristic to implement districting/dispatching strategies that improve the patient survival probability depending on the response times. The ambulance location remains fixed, and the dispatching decisions consider a patient's priority.

Toro-Díaz et al. [68] aim to minimize the mean response time in an EMS system by addressing a queuing model through a Tabu Search technique. Their approach utilizes a non-linear stochastic mixed integer programming model that considers the busy fraction of each open station where multiple ambulances are located. However, the queuing sub-model, which represents a finite-state continuous-time stochastic process, proves to be particularly challenging to solve for large-scale systems. Consequently, it is crucial to apply the previously proposed heuristic to determine the value of the objective function.

Chanta et al. [21] present a novel approach to optimize the location and dispatch of emergency response units. The proposed methodology combines Tabu Search with an embedded queuing model to address both the spatial placement of units and the dynamic dispatching decisions necessary to respond to emergencies effectively. By incorporating queuing theory, the model accounts for the stochastic nature of emergency incidents and response times, resulting in a responsive and efficient emergency service system. This hybrid approach demonstrates significant improvements in minimizing response times and maximizing coverage, which is crucial for effective emergency management.

While there is limited information regarding heuristics specifically for solving two-stage integer stochastic programming models, it's evident that other techniques have garnered more extensive study. However, as previously mentioned, researchers utilize heuristic processes to optimize specific aspects of the proposed model. By leveraging these heuristics strategically, we can enhance specific functionalities within the model, leading to improved overall performance in addressing the challenges posed by stochastic programming in EMS systems. All the information gathered will be used collectively to formulate a novel problem that can integrate and utilize the knowledge and strategies mentioned above. This approach aims to provide an advanced and comprehensive method of optimizing EMS systems.

2.3 Contribution

From the modeling perspective, the novelty of this work is to introduce a stochastic programming model that considers both the coverage maximizing that could be total or partial, and the location and dispatching of different types of ambulances. To the best of my knowledge, some of these have been addressed but not all in the same modeling framework. When a BLS ambulance is dispatched to an emergency that requires ALS, it can reduce the patient's survival. Thus, this work considers that ALS ambulances can be used as BLS units, but the contrary is not allowed [7]. There are a few works dealing with different types of ambulances as we do in this work. McLay [51] determines how to optimally locate and use ambulances to improve patient survivability and coordinate multiple medical units with a hypercube queueing model. Grannan et al. [34] determine how to dispatch multiple types of air assets to prioritized service calls to maintain a high probability of survival of the most urgent casualties in a military medical evacuation by a binary linear programming model. In Yoon et al. [71], two types of vehicles are considered, but one of them is a rapid vehicle that cannot offer the first care services of an ambulance. Moreover, neither of these works considers partial coverage of the calls.

We denote our problem as the *Emergency Vehicle Covering and Planning Problem* (EVCPP) which consists of locating a limited number of two heterogeneous types of ambulances in different city points and dispatch them to the accidents (demand points), considering the uncertainty of accident points, so as to maximize coverage (even if partially) with short medical first aid response time. In early works, the location and dispatching decisions are made separately [11, 27, 70]. However, there are also studies where these aspects are handled simultaneously as in Ansari et al. [5], Toro-Díaz et al. [68], and Amorim et al. [4].

The proposed stochastic integer 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 EVCPP stochastic model allows partial coverage of accidents by ambulances based on a decay function [41].

Similarly to Yoon et al. [71], we generate the call-arrival scenarios by sampling 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 on Friday night. Thus, time is not explicitly measured, and it is assumed that a vehicle can be assigned only once during this high ambulance demand period [75]. Boujemaa et al. [18] uses a bundle of calls but does not consider a heterogeneous ambulance fleet.

The methodological contribution consists of two solution procedures. The first proposed algorithm is based on the solution of a surrogate model and is aimed at small- to medium-scale instances. The second is a metaheuristic aimed at larger instances. In fact, the proposed model can only be solved for relatively small instances with a restricted number of scenarios. Thus, instead of decomposing the model with Bender's methods as is usually done [66, 74], we propose a location-allocation methodology [64, 69] that relies on the solution of an auxiliary surrogate model, which is faster to solve. We name this method *a surrogate-based 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. [67] integrate location and dispatch 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 midsize instances. Some others, such as Amorim et al. [4], use simulation to determine whether ambulances should remain at potential sites established by a mathematical model or be relocated to another potential site to maximize patient survival. They work on a complete day period, while we focus on high-demand periods of some hours.

In an attempt to tackle larger instances, we also develop a heuristic algorithm. This heuristic is a local search procedure that enhances the solution to the ambulance location obtained from the surrogate-based feedback approach, considering four neighborhoods to explore different disturbances for the solution. The solution is taken as an initial solution of the heuristic procedure; it is disturbed based on a neighborhood proposed; each neighbor is evaluated at the proposed model, and a local search procedure with a *first-found* criteria is done. Once the first local search is completed for the first neighborhood, the next one is processed in the same manner, and the same applies to the other two neighborhoods. The primary objective of this heuristic is to explore various solutions that can enhance the objective function values. By fixing the location variables, the model can evaluate the second phase more efficiently.

Chapter 3

Problem description

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

3.1 Description and assumptions

Let us formally describe the EVCPP. Let I be the set of possible demand points for patients who 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. Let L be the set of potential sites or ambulance stations where ambulances could be located, such as hospitals, firehouses, malls, or similar places where ambulances and paramedics can wait for emergency calls. We consider instances with up to 30 potential sites for the experimental results. Let K be the set of 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 travel time of any type of ambulance from a potential site $l \in L$ to a demand point $i \in I$ is given by r_{li}. While it is true that these travel times may be affected by many factors such as traffic conditions, within a specific time period where the model is applied, vehicle speed is more or less constant. Thus, we assume that these parameters are known following common practice in the literature. Ideally, ambulances should arrive in less than τ minutes in a life-threatening emergency. Usually, τ is a fixed value in the $[8, 15]$ minute 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, unfortunately, the accident has likely been handled by other means.

One special feature incorporated in our model is the consideration of allowing partial covering. This is achieved by considering a gradual coverage decay function that have not been used in stochastic ambulance location problems, to the best of our knowledge. In our particular real-world setting, the number of available ambulances is very scarce, which implies that when ambulances are needed, accidents may not be fully covered. Therefore, it makes sense to consider the benefits of partial coverage to help customers get urgent care. In the ambulance location literature, a total coverage function is typically used.

Since the EVCPP aims 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, based on the defined by Berman et al. [14], that only depends on the response time of a location $l \in L$ to any demand point $i \in I$:

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

Thus, c_{li}=1 is the normal coverage definition if the ambulance can arrive in less than τ minutes. However, if the ambulance i in location l can arrive at the emergency in more than τ minutes but in less than τ_{max}, then it takes a decreasing value with respect to the number of minutes. That is, the farther the ambulance, the smaller the value of c_{li}. If an ambulance takes more than $\tau_{\max}$ minutes, then it is too far and is not considered to able to arrive to emergency i from location l.

3.2 Information related to the scenarios

The operational level is represented by a set of scenarios Ω with a bundle list of arriving calls. Each scenario $\omega \in \Omega$ represents a set of emergencies in the demand points. Thus, a scenario is defined by the number and type of ambulances needed at each demand point. Recall that an ALS ambulance can be sent instead of a BLS ambulance, but not vice versa. Thus, each scenario $\omega \in \Omega$ indicates if there is an accident at a demand point $i \in I$ and provides the value a_{ki} related to the number of ambulances required of type $k \in K$.

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

- Total: the $a_{ki}(\omega)$ required ambulances of each type k are dispatched to i, and all arrive in less than τ time.

- Total-late: the $a_{ki}(\omega)$ required ambulances of each type k are dispatched, but at least one arrives between $(\tau, \tau_{\max})$ time.

- Partial: at least one of the $a_{ki}(\omega)$ 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 $a_{ki}(\omega)$ 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 $a_{ki}(\omega)$ required ambulances arrives in less than $\tau_{\max}$ time, for $k \in K$.

Figure 3.1 illustrates a solution to the EVCPP considered in this work. Five emergencies (red triangles) occur in the city during a rush hour period. There are three ALS ambulances (blue) and four BLS ambulances (dark blue) located in the city, which are dispatched to emergency situations. Emergency 1 has *total coverage* as it needs one ALS and one BLS that arrive within the ideal response time (green circle). Emergency 2 needs two BLS and one ALS. It has a *total-late coverage* since one ALS and one BLS arrive after the ideal response time (orange circles), while a BLS arrives within the ideal time. Emergency 3 needs one ALS and one BLS. It has a *partial coverage* since only the BLS can respond to the emergency within the ideal response time. Emergency 4 requires two BLS. It has a *partial-late coverage* because only one ALS (replacing a BLS) arrives, but with a longer response time than ideal. Unfortunately, Emergency 5 has a *null coverage* since the BLS ambulance that is required does not arrive within the maximum tolerated response time. Considering the number of ambulances available, their type and requirements, the coverage obtained by solving the EVCPP is the best. Note that

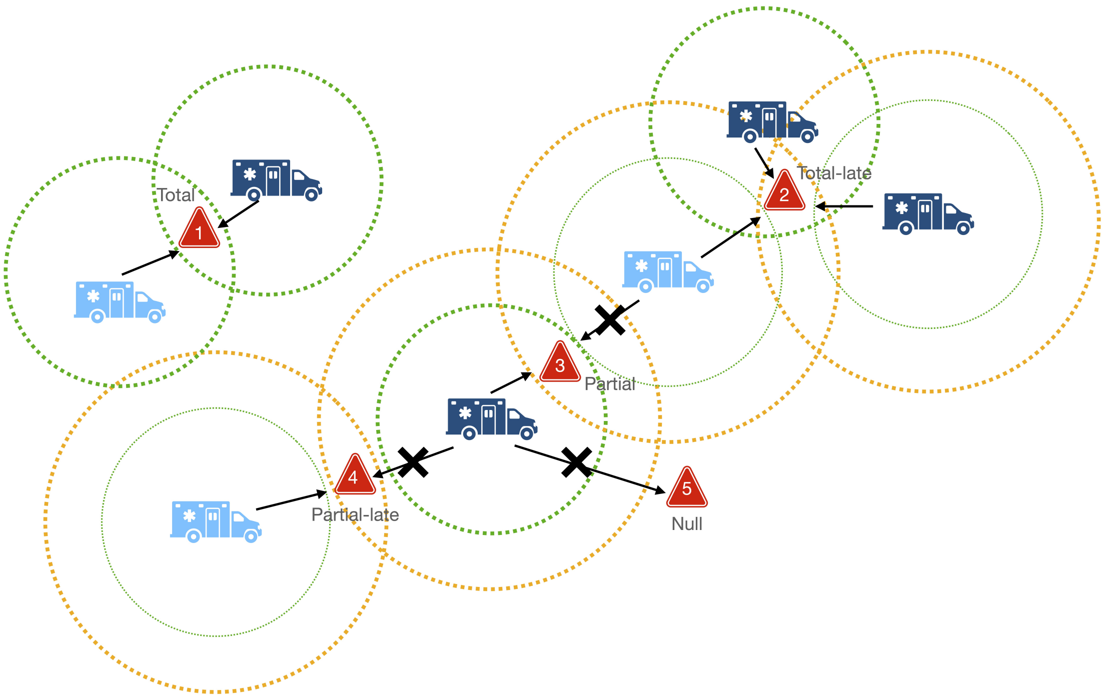


Figure 3.1: Five emergencies (red triangles), three ALS (blue) ambulances, and four BLS ones (dark blue). Emergency 1 has a *total coverage*; Emergency 2 has a *total-late coverage*; Emergency 3 has a *partial coverage*; Emergency 4 has a *partial-late coverage*; Emergency 5 has a *null coverage*.

every emergency is treated as a whole event, and the solution tries to cover most of them, if not fully, at least partially, which in reality translates into saving lives.

Table 3.1 summarizes the sets and parameters used to describe the EVCPP.

3.3 Maximum expected coverage stochastic formulation for the EVCPP

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 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

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 $i \in I$
Ω	set of scenarios
$a_{ki}(\omega)$	number of needed ambulances of type $k \in K$ at demand point $i \in I, \omega \in \Omega$
$I(\omega)$	set of demand points for $\omega \in \Omega$ with at least a value $a_{ki}(\omega) \neq 0$ for $i \in I, k \in K$

Table 3.1: Sets and parameters to describe the EVCPP.

scenario $\omega \in \Omega$:

$$y_{lki}(\omega) = \begin{cases} 1 & \text{if an ambulance of type } k \in K \text{ in location } l \in L \\ & \quad \text{is dispatched to demand point } i \in I(\omega), \text{ for scenario } \omega \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

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 $i \in I(\omega), \omega \in \Omega$:

$$f_i(\omega) = \begin{cases} 1 & \text{if demand point } i \in I(\omega) \text{ has a } \textit{total} \text{ coverage,} \\ 0 & \text{otherwise,} \end{cases}$$

$$g_i(\omega) = \begin{cases} 1 & \text{if demand point } i \in I(\omega) \text{ has a } \textit{total-late} \text{ coverage,} \\ 0 & \text{otherwise.} \end{cases}$$

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

$$h_i(\omega) = \begin{cases} 1 & \text{if demand point } i \in I(\omega) \text{ has a \textit{partial} coverage,} \\ 0 & \text{otherwise,} \end{cases}$$

$$w_i(\omega) = \begin{cases} 1 & \text{if demand point } i \in I(\omega) \text{ has a \textit{partial-late} coverage,} \\ 0 & \text{otherwise.} \end{cases}$$

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

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

Let $\mathcal{Q}(x, a)$ denote the maximum coverage given decision x and random parameter array a. Given the notation introduced above, $a(\omega) = (a_{ki}(\omega))$, represents a vector of realizations of parameter array $a_{ki}(\omega)$ under scenario $\omega \in \Omega$. Thus, we aim to find x that maximizes the expected coverage. For simplicity, let $Q(x, \omega)$ denote the maximum coverage under the specific realization of scenario ω and let $\pi(\omega)$ the probability of occurrence of scenario ω. In our work, we assume scenarios are equally likely, so $\pi(\omega) = 1/|\Omega|$ for all $\omega \in \Omega$.

The MEC formulation is as follows.

$$\max_x \mathbb{E}[\mathcal{Q}(x, a)] \tag{3.1}$$

$$\text{where} \quad \mathbb{E}[\mathcal{Q}(x, a)] = \sum_{\omega \in \Omega} \pi(\omega) \mathcal{Q}(x, \omega) \quad \text{and}$$

$$\mathcal{Q}(x, \omega) = \max_{(y,f,g,h,w,z)} \sum_{i \in I(\omega)} (\alpha_1 f_i(\omega) + \alpha_2 g_i(\omega) + \alpha_3 h_i(\omega) + \alpha_4 w_i(\omega) - \phi z_i(\omega))$$

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

$$\sum_{i \in I(\omega)} y_{lki}(\omega) \leq x_{lk} \qquad l \in L, k \in K, \omega \in \Omega \tag{3.3}$$

$$f_i(\omega) \sum_{k \in K} a_{ki}(\omega) \leq \sum_{l \in L} \sum_{k \in K} c_{li} y_{lki}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.4}$$

$$a_{2i}(\omega) f_i(\omega) \leq \sum_{l \in L} c_{li} y_{l2i}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.5}$$

$$g_i(\omega) \sum_{k \in K} a_{ki}(\omega) \leq \sum_{l \in L} \sum_{k \in K} y_{lki}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.6}$$

$$a_{2i}(\omega) g_i(\omega) \leq \sum_{l \in L} y_{l2i}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.7}$$

$$g_i(\omega) \leq M \left(\sum_{l \in L} (1 - c_{li}) \sum_{k \in K} y_{lki}(\omega) \right) \qquad i \in I(\omega), \omega \in \Omega \tag{3.8}$$

$$h_i(\omega) \leq \sum_{k \in K} a_{ki}(\omega) - \sum_{l \in L} \sum_{k \in K} y_{lki}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.9}$$

$$h_i(\omega) \leq a_{2i}(\omega) - \sum_{l \in L} y_{l2i}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.10}$$

$$\sum_{l \in L} (h_i(\omega) - c_{li}) \sum_{k \in K} y_{lki}(\omega) \leq 0 \qquad i \in I(\omega), \omega \in \Omega \tag{3.11}$$

$$w_i(\omega) \leq \sum_{k \in K} a_{ki}(\omega) - \sum_{l \in L} \sum_{k \in K} y_{lki}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.12}$$

$$w_i(\omega) \leq a_{2i}(\omega) - \sum_{l \in L} y_{l2i}(\omega) \qquad i \in I(\omega), \omega \in \Omega \tag{3.13}$$

$$w_i(\omega) \leq M \left(\sum_{l \in L} (1 - c_{li}) \sum_{k \in K} y_{lki}(\omega) \right) \qquad i \in I(\omega), \omega \in \Omega \tag{3.14}$$

$$\sum_{l \in L} \sum_{k \in K} y_{lki}(\omega) + z_i(\omega) \geq 1 \qquad i \in I(\omega), \omega \in \Omega \tag{3.15}$$

$$f_i(\omega) + g_i(\omega) + h_i(\omega) + w_i(\omega) + z_i(\omega) = 1 \qquad i \in I(\omega), \omega \in \Omega \tag{3.16}$$

$$x_{lk} \in \mathbb{Z}^+, y_{lki}(\omega) \in \{0, 1\} \qquad l \in L, k \in K, i \in I(\omega), \omega \in \Omega \tag{3.17}$$

$$f_i(\omega), g_i(\omega), h_i(\omega), w_i(\omega), z_i(\omega) \in \{0,1\} \qquad i \in I(\omega), \omega \in \Omega \tag{3.18}$$

The objective function (3.1) maximizes the expected value of the weighted coverage of emergencies. This sum is weighted to give greater importance to the types of coverage we want to occur more frequently than others. For practical reasons, the coverage types were arranged in descending order, taking into account the importance that we, along with experts in the field, considered in the case of studies from the state of Nuevo León, where we based our experimentation. The parameters $\alpha_1, \alpha_2, \alpha_3$ and α_4 are weights that ponder the coverage type, and ϕ is a penalty for null coverage. This parameters represent the value of each coverage type. The quantity for each one of them is fixed according to experts suggestions, which analyze the 9-1-1 system. After experts finish the analysis, they decide the coverage's importance, and the parameters' values decrease as the importance decreases. We assume that every scenario is equally probable since each $\omega \in \Omega$ represents a sample of the high-demand period in which we are interested. Constraints (3.2) establish the number of ambulances available per type. Constraints (3.3) establish the relationship between the first and second-stage variables, which means that 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)–(3.5). We have in the first constraint that, if $f_i(\omega) = 1$ and is multiplied by the total number of ambulances of the same type needed at a demand point, then c_{li} are equal to 1. However, if at least one ambulance is late, i.e. if $c_{li} < 1$ for one or more ambulances, then $f_i(\omega) = 0$. In the second constraint, we guarantee that ALS ambulances can only cover demand points that need ALS ambulances. If $f_i(\omega) = 1$, then the number of ambulances dispatched of type 2 is not late, that is, $c_{li} = 1$. Otherwise, if the number of dispatched ambulances of this type with $c_{li} = 1$ is not sufficient to cover at least the ALS needed, then $f_i(\omega) = 0$.

The *total-late* coverage is defined by constraints (3.6)–(3.8). Constraints (3.6)–(3.7) allow the total-late coverage variables $g_i(\omega)$ to be one when the dispatch variables are active. In (3.6), $g_i(\omega) = 1$ multiplied by the total number of ambulances of both types needed, implying that all must be dispatched, but not necessarily in a time less than τ. Note that c_{li} could be less than 1, allowing for late coverage. If not all needed ambulances are dispatched, then $g_i(\omega) = 0$. Constraints (3.8) indicate that at least all needed ALS ambulances must be dispatched.

Meanwhile, constraints (3.8) track the demand points where the response time is between $(\tau, \tau_{\max})$ when the difference on the right-hand side of the equation is positive, that is, when there is a value $c_{li} < 1$ associated to a dispatched ambulance, for $l \in L, i \in I(\omega), \omega \in \Omega$. Note that, when $c_{li} = 1$ for all ambulances dispatched, then $g_i(\omega) = 0$ because that case implies a total coverage. Otherwise, the term multiplying M is a positive fractional number. The value of M introduced has to be large enough to make the constraint redundant in this case. In our testing, we observed that a value of $M = 1000$ suffices.

The *partial* coverage is defined by constraints (3.9)–(3.11). Recall that in this case not all needed ambulances are dispatched to the emergency, but the ones dispatched have an ideal response time. Thus, constraints (3.9)–(3.10) activate variables $h_i(\omega)$ if the number of ambulances dispatched is less than the required. In the first equation, $h_i(\omega) = 1$ implies that the difference between the needed and the dispatched ambulances is one or more. If the difference is zero, then $h_i(\omega)$ must be zero. For the second equation, we guarantee that the ambulances type 2 needed are covered only for ALS ambulances although not all must be dispatched. Quadratic constraints (3.11) ensure that ambulances dispatched arrive within the ideal response time, that is, their corresponding value $c_{li} = 1$, for $l \in L, i \in I(\omega), \omega \in \Omega$. If

$h_i(\omega) = 1$ and the ambulances arrive at an ideal time, the difference between $h_i(\omega)$ and c_{li} is zero and guarantees that all ambulances must be dispatched at a time less than τ. If at least one ambulance is late, the difference would be positive and the constraint is infeasible because that case is a *partial-late* coverage.

Constraints (3.12)–(3.14) define the *partial-late* coverage. Note that the case of the ALS ambulances ($k = 2$) replacing a BLS one is also considered. Constraints (3.12)–(3.13) activate the variables $w_i(\omega)$ when the number of ambulances required exceeds the number of ambulances dispatched. Similarly to the total-late coverage, the constraints (3.14) track ambulances with a response time larger than the ideal and must be multiplied by a M since there could be cases with a sum that is less than 1. Here we also deal with the ALS ambulances that may replace the BLS.

The *null* coverage is activated by constraints (3.15). All coverage restrictions are related to the constraint (3.16) which ensures only one type of coverage for each emergency. Finally, (3.17) and (3.18) 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. However, the related number of variables and constraints is usually large. In addition, constraints (3.11) are quadratic. Although integer linear stochastic models could be formulated with classical linearization methods, 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 scenario-based feedback methodology presented in the next section.

Chapter 4

Solution methodologies

4.1 Surrogate-based feedback method for the EVCPP

The EVCPP is $\mathcal{NP}$-hard since the classical $\mathcal{NP}$-hard facility location problem could be polynomially reduced to it, as we can see in Megiddo and Tamir [53]. The MEC model is experimentally challenging to solve, even for medium-sized instances, as shown in Section 4.5. The main drawback to obtaining a simple solution to the proposed problem is that integer variables are used. Typically, stochastic programming problems use continuous variables in most cases, making it challenging to apply solution approximation methods commonly used in the literature. 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 $k \in K, l \in L, i \in I(\omega), \omega \in \Omega$:

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

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

Variables $u_{lki}(\omega)$ indicate the ambulances with an ideal response time dispatched from the location sites corresponding to a decay function value $c_{li} = 1$. While variables $v_{lki}(\omega)$ 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 variables $\zeta_{ki}(\omega)$, for $k \in K, i \in I(\omega), \omega \in \Omega$.

Let $\mathcal{G}(x, a)$ denote the maximum expected value of the on-time and late dispatched ambulances minus a penalty for the required ambulances that could not be dispatched on time, given given decision x and random parameter array a. For simplicity, let $G(x, \omega)$ denote the maximum coverage under the specific realization of scenario ω and let $\pi(\omega)$ the probability of occurrence of scenario ω. In our work, we assume scenarios are equally likely, so $\pi(\omega) = 1/|\Omega|$ for all $\omega \in \Omega$. Then SABC can be expressed as:

$$\max_{x} \mathbb{E}[\mathcal{G}(x,a)] \tag{4.1}$$

$$\text{where} \quad \mathbb{E}[\mathcal{G}(x,a)] = \sum_{\omega\in\Omega} \pi(\omega)\mathcal{G}(x,\omega) \quad \text{and}$$

$$\mathcal{G}(x,\omega) = \max_{u,v,\zeta} \left[\sum_{l\in L}\sum_{k\in K}\sum_{i\in I(\omega)} (\beta_1 u_{lki}(\omega) + \beta_2 v_{lki}(\omega)) - \sum_{k\in K}\sum_{i\in I(\omega)} \phi\zeta_{ki}(\omega) \right]$$

$$\text{s.t.} \quad \sum_{l\in L} x_{lk} \leq \eta_k \qquad k \in K \tag{4.2}$$

$$\sum_{i\in I(\omega)} (u_{lki}(\omega) + v_{lki}(\omega)) \leq x_{lk} \qquad l \in L, k \in K, \omega \in \Omega \tag{4.3}$$

$$u_{lki}(\omega) \leq c_{li} \qquad l \in L, i \in I(s), k \in K, \omega \in \Omega \tag{4.4}$$

$$u_{lki}(\omega) + v_{lki}(\omega) \leq 1 \qquad l \in L, i \in I(s), k \in K, \omega \in \Omega \tag{4.5}$$

$$a_{1i}(\omega) + a_{2i}(\omega) = \sum_{l\in L}\sum_{k\in K}(u_{lki}(\omega) + v_{lki}(\omega) + \zeta_{ki}(\omega)) \qquad i \in I(\omega), \omega \in \Omega \tag{4.6}$$

$$a_{2i}(\omega) \leq \sum_{l\in L}(u_{l2i}(\omega) + v_{l2i}(\omega) + \zeta_{2i}(\omega)) \qquad i \in I(\omega), \omega \in \Omega \tag{4.7}$$

$$x_{lk}, \zeta_{ki}(\omega) \in \mathbb{Z}^+, u_{lki}(\omega), v_{lki}(\omega) \in \{0,1\} \qquad l \in L, k \in K, i \in I(\omega), \omega \in \Omega \tag{4.8}$$

The objective function (4.1) 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 β_1 and β_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 (4.2). The number of ambulances dispatched on time or late is less than the number of ambulances located, as indicated by constraints (4.3). Constraints (4.4) 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 (4.5) activate the late variables for which their response time is between $(\tau, \tau_{\max})$. With constraints (4.6) and (4.7), the non-covered emergencies, $\zeta_{ki}(\omega)$ variables are defined for $i \in I(\omega), \omega \in \Omega$. Recall that advanced ambulances can be dispatched instead of basic ambulances. 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 the MEC(SABC) model, 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. However, 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 in terms of 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** $i \in I(\omega), \omega \in \Omega$ **do**
3:     **if** $\sum_{l\in L,k\in K} \bar{u}_{lki}(\omega) = a_{ki}(\omega)$ **then** $f_i(\omega) = 1$ ▷ total coverage
4:     **if** $\sum_{l\in L,k\in K} \bar{u}_{lki}(\omega) < a_{ki}(\omega)$ and $\sum_{l\in L,k\in K} \bar{u}_{lki}(\omega) + \bar{v}_{lki}(\omega) = a_{ki}(\omega)$
5:         **then** $g_i(\omega) = 1$ ▷ total-late coverage
6:     **if** $\sum_{l\in L,k\in K} \bar{u}_{lki}(\omega) < a_{ki}(\omega)$ and $\sum_{l\in L,k\in K} \bar{v}_{lki}(\omega) = 0$
7:         **then** $h_i(\omega) = 1$ ▷ partial coverage
8:     **if** $\sum_{l\in L,k\in K} \bar{u}_{lki}(\omega) + \bar{v}_{lki}(\omega) < a_{ki}(\omega)$ and $\sum_{l\in L,k\in K} \bar{v}_{lki}(\omega) > 0$
9:         **then** $w_i(\omega) = 1$ ▷ partial-late coverage
10:     **otherwise** $z_i(\omega) = 1$ ▷ null coverage
11: **end for**
12: **return** MEC solution $(\bar{x}, f, g, h, w, z)$

4.2 Heuristic to improve the SBFM

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 heuristic considering four neighborhoods, named as Neighborhoods Local Search Heuristic (NLSH) procedure. These four different neighborhoods are as follows:

- Neighborhood 1 (N_1): exchange one active potential site with a non-active potential site.

- Neighborhood 2 (N_2): transfer one ambulance from an active potential site to a non-active potential site.
- Neighborhood 3 (N_3): transfer one ambulance from an active potential site to another active potential site.
- Neighborhood 4 (N_4): exchange one active potential site with another active potential site.

N_1 and N_4 change both ambulance types at the same time, while N_2 and N_3 transfer only one BLS ambulance, evaluate the MEC model, and then transfer one ALS ambulance, and evaluate the MEC model if there were no changes with the BLS ambulance evaluation. We can see examples for each neighborhood in Figure 4.1.

In the Algorithms 2 (part 1) and 3 (part 2), we can see the NLSH procedure to improve the SBFM. The NLSH procedure has the x_{SABC} variables as an initial solution, which are the location variables of the SBFM's best solution s^*_{SABC}. We define a stop criterion of 3600 seconds. In lines 2 and 3, we define the location variables x_{SABC} as x^*, and the objective value of the best solution s^*_{SABC} as s^*, respectively. While the runtime is less than the stop criterion (line 4), we can explore the neighborhoods mentioned before. The heuristic procedure explores neighborhoods until a local optimum is found. To ensure this, we set a flag to TRUE (line 5). While the flag is equal to TRUE (line 6), the flag turns to FALSE (line 7), and a neighborhood N is created by modifying the best solution at the moment x^* as we described in neighborhood N_1 (line 8). The next step is to solve the MEC model, fixing each neighbor as an initial solution x' for the MEC model (lines 9-10); this solution is named s'. If s' is better than s^* (line 11), then s^* changes its value to s', x^* changes its value to x' (line 12), and the flag turns its value to TRUE (line 13). The for loop breaks (line 14), and the algorithm repeats the procedure until no further

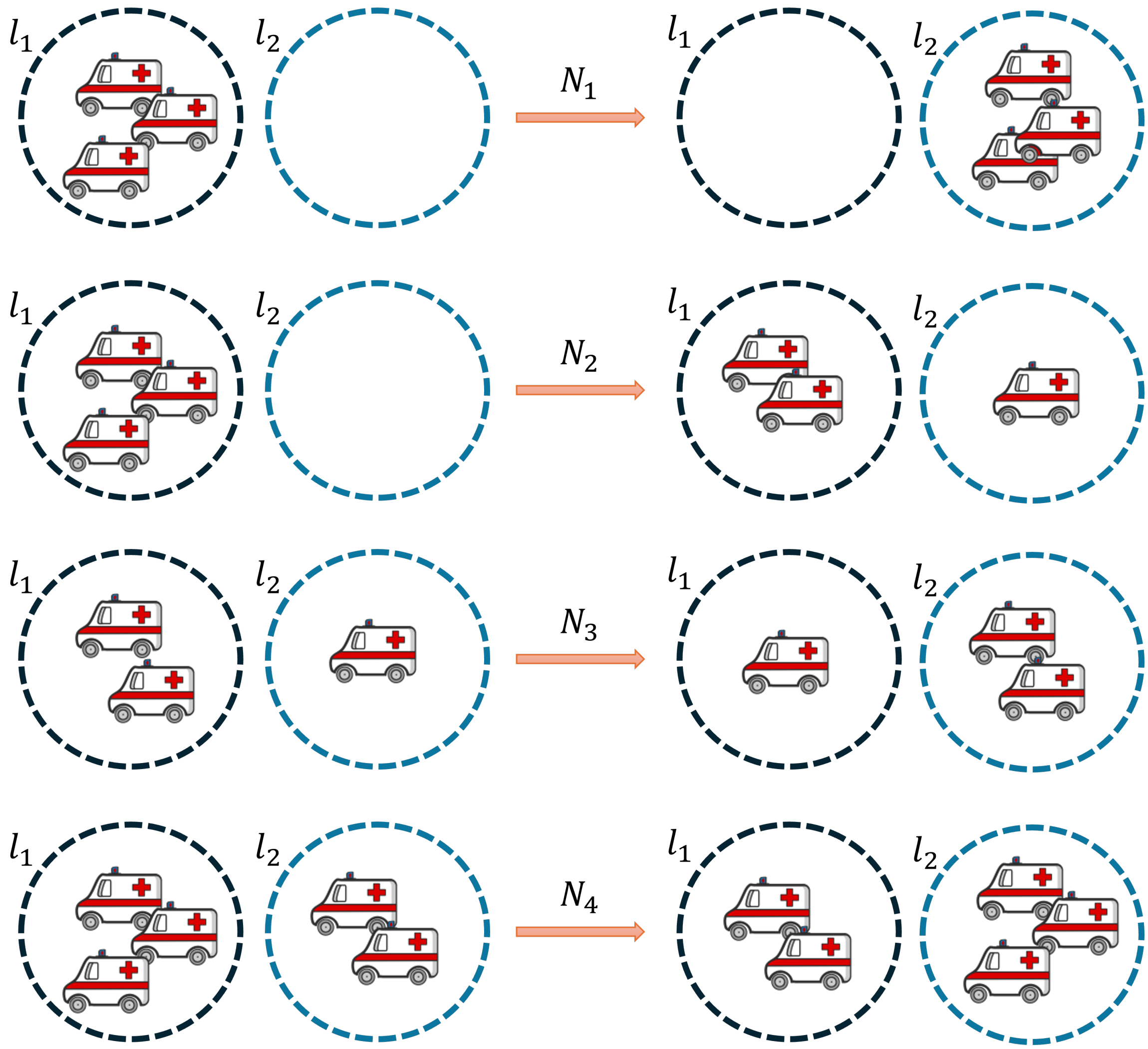


Figure 4.1: Four neighborhoods: N_1 exchange one active potential site with a non-active potential site, N_2 transfer one BLS or ALS ambulance from an active potential site to a non-active potential site, N_3 transfer one BLS or ALS ambulance from an active potential site to another active potential site, and N_4 exchange one active potential site with another active potential site.

improvements are made to the solution. The step from lines 5 to 17 is repeated for the four neighborhoods proposed. The improvement using neighborhood N_2 is defined from line 18 to line 30. We can see that the N_3 procedure is at lines 31 to 43, and lines 44 to 57 define the N_4 improvement procedure. At the end of the heuristic, the best solution is returned (line 58).

Algorithm 2 NLSH procedure to improve the SBFM (Part 1)

1: **require** solution of the SABC model s^*_{SABC}, $stop\_criteria = 3600$ seconds
2: $x^* = x_{SABC}$
3: $s^* = MEC(x^*_{SABC})$
4: **while** runtime $<$ $stop\_criteria$ **do**
5:     flag = TRUE
6:     **while** flag = TRUE **do**
7:         flag = FALSE
8:         Let $N = N_1(x^*)$
9:         **for** $x' \in N$ **do**
10:             $s' = MEC(x')$
11:             **if** $s' > s^*$ **then**
12:                 $s^* = s'$, $x^* = x'$
13:                 flag = TRUE
14:                 **break**
15:             **end if**
16:         **end for**
17:     **end while**
18:     flag = TRUE
19:     **while** flag = TRUE **do**
20:         flag = FALSE
21:         Let $N = N_2(x^*)$
22:         **for** $x' \in N$ **do**
23:             $s' = MEC(x')$
24:             **if** $s' > s^*$ **then**
25:                 $s^* = s'$, $x^* = x'$
26:                 flag = TRUE
27:                 **break**
28:             **end if**
29:         **end for**
30:     **end while**

Algorithm 3 NLSH procedure to improve the SBFM (Part 2)

```
    flag = TRUE
    while flag = TRUE do
        flag = FALSE
        Let N = N_3(x*)
        for x' ∈ N do
            s' = MEC(x')
            if s' > s* then
                s* = s', x* = x'
                flag = TRUE
                break
            end if
        end for
    end while
    flag = TRUE
    while flag = TRUE do
        flag = FALSE
        Let N = N_4(x*)
        for x' ∈ N do
            s' = MEC(x')
            if s' > s* then
                s* = s', x* = x'
                flag = TRUE
                break
            end if
        end for
    end while
end while
return solution (x, f, g, h, w, z)*
```

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

4.3 Experimental assessment

This section presents an empirical assessment of models and the solution methodology previously described to solve the EVCPP. 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 neighbors' evaluation and 10800 seconds in total.

4.3.1 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 $|I| \leq 270$ [71]. 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 (≤ 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 [71]). The number of scenarios is set to be as large as that in the literature $|\Omega| \in \{10, 50, 100, 150, 200\}$. 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 $\omega \in \Omega$ consists of a set of demand values per ambulance type and per demand point $\{a_{ki}(\omega)\}_{k \in K, i \in I, \omega \in \Omega}$. 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): ; however, as observed in real cases, some $\alpha_1 = 0.65, \alpha_2 = 0.2, \alpha_3 = 0.1$, and $\alpha_4 = 0.05$. In this manner, the total coverage is the most sought after, while the partial-late cover has less benefit. The values of these parameters can change, and they would affect decisions since they are part of the objective function, which is intended to be maximized. Depending on the values assigned to the parameters, different variables could be activated than those obtained in our results. The values assigned to the parameters are left to the user's discretion, depending on the importance given to each type of coverage based on the functioning of the EMS system. 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.

The parameter values established for this thesis were based on information provided by the managers of the Medical Emergency Regulatory Center (CRUM) in the state of Nuevo León, Mexico. CRUM is an institution responsible for managing the 9-1-1 system in various states of Mexico, and we had the opportunity to meet with the experts in charge of making decisions about the system. They recommended that we consider the established values for ambulance response times and the importance that should be given to the types of partial coverage considered. There is also a reference from PRONAMED Salud Integral, a private institution that offers ambulance dispatch services for medical emergencies [40].

For the SABC objective function (4.1) 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`. These instances are structured as follows:

Line 1: $|I|$, number of demand points.

Line 2: $|L|$, number of potential sites.

Line 3: $|\Omega|$, number of scenarios.

Next $|\Omega|$ lines: number of maximum accidents at each demand point.

Next $|\Omega|$ lines: number of ambulances needed at each demand point of each type (BLS or ALS ambulances).

Next $|L|$ lines: response time from each potential site to each demand point.

Next $|L|$ lines: benefit c_{li} from traveling from each potential site to each demand point.

4.4 Solution representation

The solutions of our models and our heuristic are provided by the ambulances located at the potential sites, along with the number of ambulances dispatched from each potential site to each demand point. For example, if we have four potential sites and four demand points, as we can see in Figure 4.2 a), and if we have four BLS ambulances and two ALS ambulances, we can obtain the solution $x_{11} = x_{22} = x_{41} = x_{42} = 0$, $x_{12} = x_{32} = 1$, $x_{21} = x_{31} = 2$, then the active potential sites are the ones in Figure 4.2 b).

Suppose that potential site 1 can cover demand points 1,2, and 3; potential site 2 can cover demand points 2 and 3; and potential site 3 can cover demand points 3 and 4. The locations of the ambulances are as follows: potential site 1 has 1 ALS ambulance, potential site 2 has 2 BLS ambulances, and potential site 3 has 2 BLS ambulances and 1 ALS ambulance. This ambulance distribution at potential sites can cover the demand points, as shown in Figure 4.3. Based on that location, dispatch decisions are then made, and the necessary or available ambulances are dispatched from each potential location (1, 2, or 3) to each demand point (1, 2, 3, or 4) it can cover.

a)

b)

Figure 4.2: Example of a solution.

Figure 4.3: Example of a solution.

4.5 Experimental work

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

4.5.1 Assessment of benefit of partial coverage

In this experiment, we aim to assess the benefit of the proposed partial coverage model. To this end, we solve our MEC model with partial coverage ($\alpha_1 = 0.65, \alpha_2 = 0.2, \alpha_3 = 0.1$, and $\alpha_4 = 0.05$), and then we solve the Total-MEC model, which corresponds to the MEC model but with $\alpha_2 = \alpha_3 = \alpha_4 = 0$ in the objective function. The Total-MEC is equivalent to eliminating the partial covering terms from the MEC model. Thus, the objective function does not consider partial coverages; only the full and null coverage terms are considered.

Table 4.1 displays the results. In the first column, the size of the instance is indicated in terms of the number of potential location sites, the number of demand points, and the number of scenarios. The second, third, and fourth columns show the objective function value, the running time (CPU seconds), and the value of the null coverage term in the objective function, for the MEC model (under partial coverage). The remaining columns show the same indicators for the Total-MEC model.

Table 4.1: Comparison between the MEC and the Total-MEC models.

	MEC model			Total-MEC		
$\lvert L\rvert$, $\lvert I\rvert$, $\lvert\Omega\rvert$	Obj. fn. value	Time (CPU sec.)	Null coverage	Obj. fn. value	Time (CPU sec.)	Null coverage
16, 168, 10	4.9	94.8	0	7.0	2342.9	13.8
16, 168, 100	4.9	10804.2	0.07	7.6	10801.9	12.04
16, 168, 200	4.8	10808.7	0.14	7.5	10804.0	12.68
16, 500, 10	7.5	64.8	0	9.3	919.4	13.8
16, 500, 100	7.4	10809.8	1.69	10.6	10804.0	18.8
16, 500, 200	8.1	10810.9	1.93	11.3	10806.6	19.12
16, 1500, 10	9.9	48.1	5.9	11.6	10801.6	27.8
16, 1500, 100	10.6	10808.2	11.82	14.0	10806.2	26.67

Table 4.1: Comparison between the MEC and the Total-MEC models.

	MEC model			Total-MEC		
16, 1500, 200	11.0	10838.3	13.56	14.4	10810.2	27.95
100, 168, 10	7.8	10802.6	0	7.2	10801.3	10.7
100, 168, 100	1.8	10817.4	0	7.5	10809.4	11.63
100, 168, 200	1.9	10843.4	0.24	7.2	10818.3	12.88
100, 500, 10	10.6	10804.2	1.7	9.8	10802.2	18.5
100, 500, 100	2.9	10831.0	1.83	10.7	10815.2	19.6
100, 500, 200	2.7	10905.8	4.36	9.7	10829.8	20.85
100, 1500, 10	13.2	10809.1	4.8	13.1	10803.6	19.6
100, 1500, 100	8.8	10849.6	13.18	14.6	10824.6	28.42
100, 1500, 200	4.2	10910.0	9.7	10.9	10848.9	34.43

From Table 4.1, the most interesting result is the comparison of the null coverage term between both models. As can be seen, the Null coverage values obtained by the MEC model are considerably lower than those obtained by the Total-MEC model. This means that when ignoring the partial coverage terms, more people are left without coverage altogether, which is, of course, undesirable. Moreover, we can also see that there were even some cases in which the null coverage term was zero under the MEC model, indicating the clear benefit of the partial coverage consideration. Overall, the MEC model improves the null coverage by 84% on average, which means more lives can be saved because at least an ambulance will arrive at the emergency. In terms of the location of the ambulances, when contrasting the MEC and Total-MEC solutions, there was more than 15% difference on average. The solution time was almost the same for both models, with the notable exception that in three instances, the MEC ran significantly faster. In summary, the most significant result is that the MEC model has a smaller average null coverage compared to the Total-MEC, indicating the substantial benefit of incorporating partial coverage terms into the objective function.

4.5.2 Evaluation of the MEC, the SBFM, the NLSH, and their comparisons

The objective of this experiment is to conduct a detailed sensitivity analysis of the parameters employed in the MEC, SBFM, and NLSH. A primary goal is to evaluate the extent to which MEC can generate high-quality solutions. Once MEC reaches a point where it can no longer achieve optimal solutions within the designated stopping criteria, a comparison between MEC and SBFM will be made to assess the efficacy of this methodology for large-scale problem instances. Finally, the results obtained from both MEC and SBFM will be compared to those from NLSH to determine if an improvement can be observed from the solutions derived using SBFM. To achieve this, we run the 75 instances mentioned earlier and calculate the objective values.

Experiment 1: Evaluation of the MEC. The goal of this experiment is to evaluate the MEC and verify the quality of the solutions. We run the MEC for the 75 instances and calculate the objective values, which are plotted in Figure 4.4. This figure consists of six plots. The three plots in the left column vary the number of demand points (x-axis) and compare each one to the value of the objective function (y-axis) under different scenarios. The three plots on the right-hand side column vary the tested number of scenarios (x-axis) and show the variation in the solution value for each number of demand points (y-axis). 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 represent the best objective values, while dotted lines represent the best bounds found.

As can be seen on Figure 4.4, 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 location 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 in the plots on the left-hand side, the better the objective function. This implies that a more comprehensive sampling of the emergency demand points benefits the quality of the solution related to 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.

The results are presented in Table 4.2 for detailed values. The first column describes the different demand points and scenario sizes. Objective function values and best bound values for $|L| = 16$ are in columns 2 and 3. For $|L| = 50$, we have objective function values in column 4 and best bound values in column 5. For $|L| = 100$, columns 6 and 7 contain the objective function values and best bound values.

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

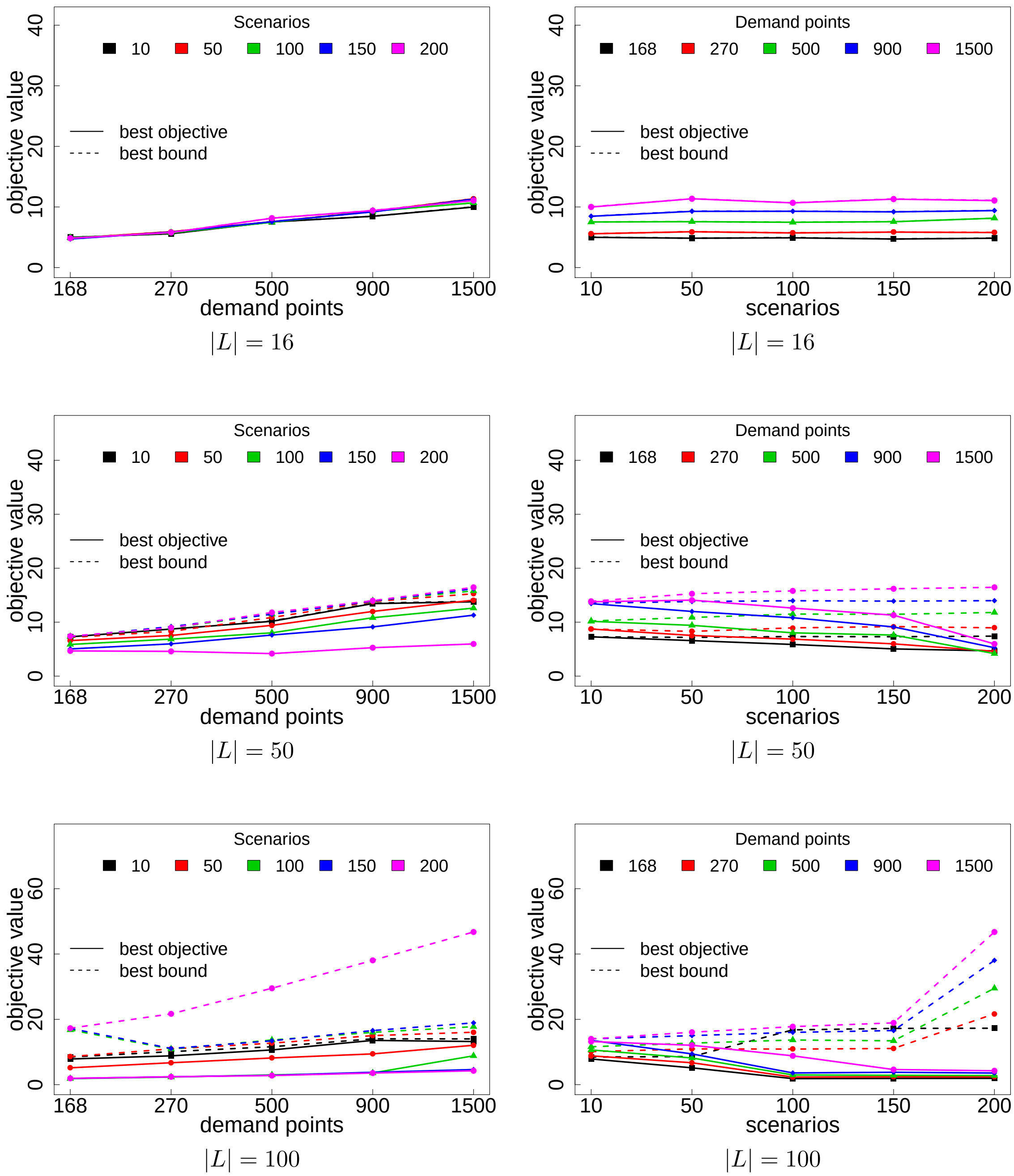


Figure 4.4: Evaluation of the MEC model varying demand points (left-hand side column) and scenarios (right-hand side column) for $|L| = 16, |L| = 50$ and $|L| = 100$.

Table 4.2: Evaluation of the MEC model for 75 instances.

$\|L\|$	16		50		100	
$\|I\|, \|\Omega\|$	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value
168, 10	5.00	5.02	7.29	7.29	7.87	8.46
168, 50	4.85	4.86	6.58	7.18	5.16	8.60
168, 100	4.91	4.94	5.87	7.34	1.85	16.78
168, 150	4.71	4.73	5.04	7.30	1.91	17.23
168, 200	4.84	4.86	4.67	7.41	1.93	17.28
270, 10	5.58	5.59	8.73	8.73	8.83	10.11
270, 50	5.91	5.92	7.54	8.29	6.71	10.96
270, 100	5.72	5.73	6.86	8.94	2.33	10.90
270, 150	5.86	5.87	5.98	9.18	2.41	11.08
270, 200	5.78	5.79	4.59	8.95	2.42	21.69
500. 10	7.53	7.54	10.16	10.22	10.63	11.61
500, 50	7.57	7.58	9.41	10.87	8.17	12.71
500, 100	7.48	7.48	8.02	11.50	2.98	13.68
500, 150	7.57	7.58	7.61	11.45	2.86	13.42
500, 200	8.15	8.17	4.18	11.80	2.79	29.52
900, 10	8.47	8.48	13.44	13.52	13.55	14.04
900, 50	9.28	9.32	11.99	13.82	9.42	15.02
900, 100	9.29	9.30	10.82	13.97	3.59	16.01
900, 150	9.20	9.21	9.12	13.90	3.79	16.57
900, 200	9.40	9.42	5.27	13.98	3.56	38.06
1500, 10	9.99	10.02	13.77	13.87	13.21	14.00
1500, 50	11.36	11.36	14.08	15.27	12.06	16.06

Table 4.2: Evaluation of the MEC model for 75 instances.

$\|L\|$	16		50		100	
$\|I\|, \|\Omega\|$	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value
1500, 100	10.68	10.69	12.61	15.80	8.83	17.76
1500, 150	11.29	11.33	11.30	16.19	4.63	18.92
1500, 200	11.05	11.10	5.97	16.46	4.25	46.75

Experiment 2: Evaluation of the the MEC(SABC). This experiment aims to evaluate the solutions of the MEC(SABC). We run this model for each one of the 75 instances described in the previous section. We have the solution represented in Figure 4.5, which has a similar structure to the previous one. We also show the results in Table 4.3.

Table 4.3: Evaluation of the MEC(SABC) model for 75 instances.

$\|L\|$	16		50		100	
$\|I\|, \|S\|$	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value
168, 10	5.00	5.00	6.49	6.51	7.07	7.08
168, 50	4.85	4.85	6.44	6.44	6.46	6.46
168, 100	4.88	4.88	6.42	6.42	6.70	6.70
168, 150	4.68	4.68	6.38	6.38	6.85	6.85
168, 200	4.82	4.82	6.38	6.38	6.83	6.83
270, 10	5.53	5.53	8.05	8.10	8.26	8.28

Table 4.3: Evaluation of the MEC(SABC) model for 75 instances.

$\lvert L \rvert$	16		50		100	
$\lvert I \rvert, \lvert S \rvert$	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value	Obj. fn. value	Best bound value
270, 50	5.87	5.87	7.38	7.38	8.69	8.69
270, 100	5.70	5.70	7.83	7.83	8.21	8.21
270, 150	5.83	5.84	7.81	7.81	8.38	8.38
270, 200	5.77	5.77	7.50	7.50	8.16	8.16
500. 10	7.31	7.32	9.25	9.26	10.09	10.13
500, 50	7.52	7.53	9.55	9.56	10.10	10.10
500, 100	7.44	7.44	9.95	9.95	10.41	10.41
500, 150	7.51	7.51	9.78	9.78	10.13	10.13
500, 200	8.12	8.12	9.88	9.89	10.30	10.30
900, 10	8.43	8.43	12.54	12.54	12.79	12.82
900, 50	9.24	9.24	12.29	12.29	12.42	12.42
900, 100	9.26	9.26	12.29	12.29	12.18	12.18
900, 150	9.18	9.18	11.84	11.84	12.72	12.72
900, 200	9.39	9.40	11.99	11.99	12.40	12.40
1500, 10	9.94	9.96	12.78	12.80	12.89	12.91
1500, 50	11.31	11.32	13.71	13.78	13.69	13.69
1500, 100	10.64	10.69	13.83	13.83	14.68	14.68
1500, 150	11.28	11.28	14.08	14.08	14.66	14.66
1500, 200	11.06	11.06	14.11	14.11	14.76	14.76

For this methodology, the number of scenarios does not affect the results for

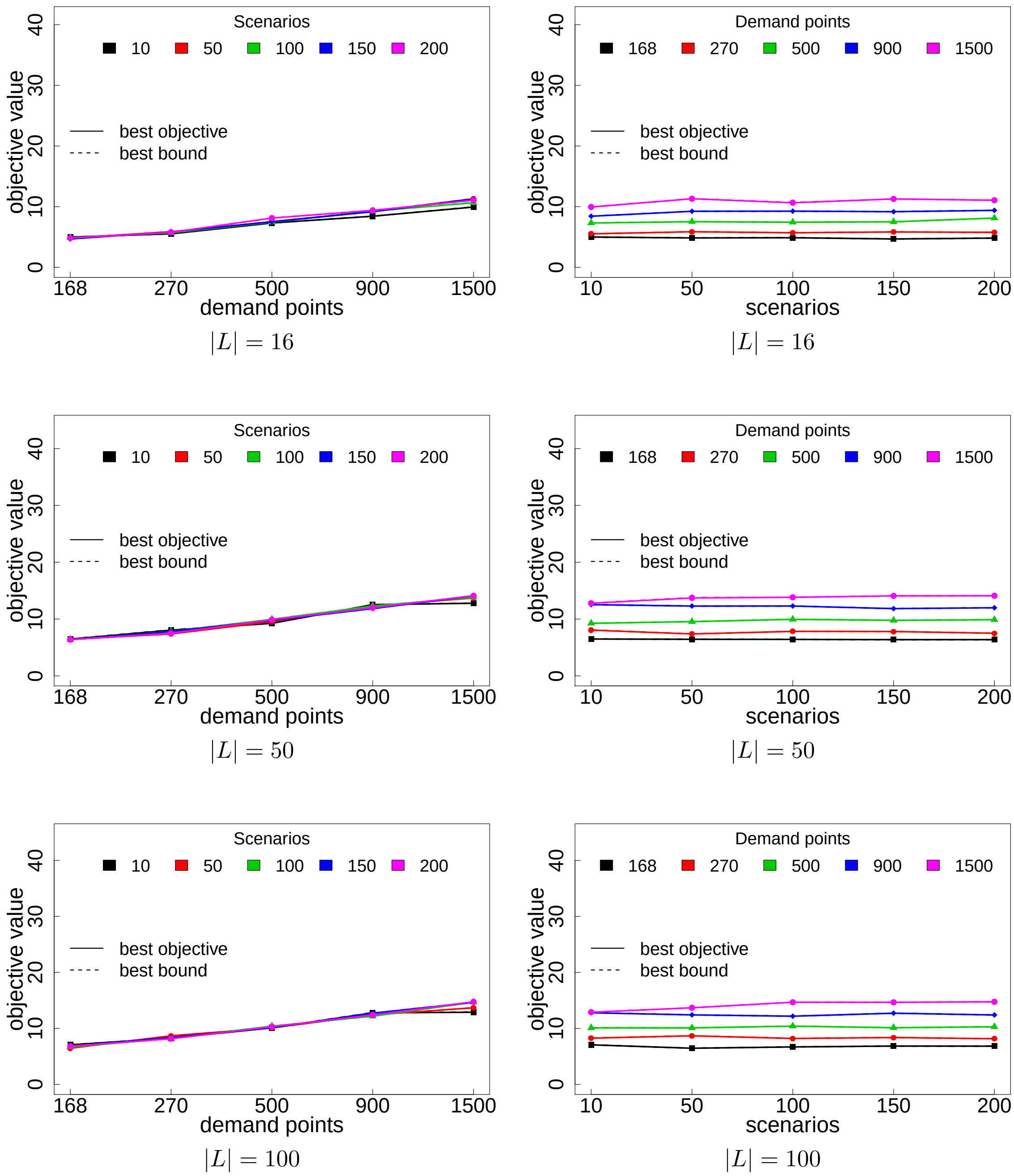


Figure 4.5: Evaluation of the MEC(SABC) model varying demand points (left-hand side column) and scenarios (right-hand side column) for $|L| = 16, |L| = 50$ and $|L| = 100$.

the objective value, because the MEC(SABC) only considers the ambulances serving the accidents. According to this methodology, the optimal is found in an easier and faster way than the MEC, but it is different from the optimal at the MEC, so we need to compare the objective values of both of them.

Experiment 3: Assessment of the surrogate-based feedback method. Naturally, one of the most important aspects to investigate is the value and benefit that the proposed solution method brings to the table. Thus, in this set of experiments, we solved all instances for the different configurations previously discussed under two different methods. We solved the MEC model by directly applying the branch-and-bound method from the solver and compared it with the proposed SBFM. Figure 4.6 consists of six plots. The three plots in the left column consider the same x-axis and y-axis values as we mentioned in the previous figures. Same for the right column figures. 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 for the MEC(SABC) model, while dotted ones are the best objective values for the MEC model.

The results are detailed in Table 4.4. In the first column, we have the demand points and scenario sizes. In the second column, we present the MEC objective function values, and in the third column, we display the SBFM objective function values. These values are for sixteen potential sites. For fifty potential sites, we have the results of the MEC and the SBFM in columns four and five. And finally, in columns six and seven, we have MEC and SBFM objective function values for one hundred potential sites.

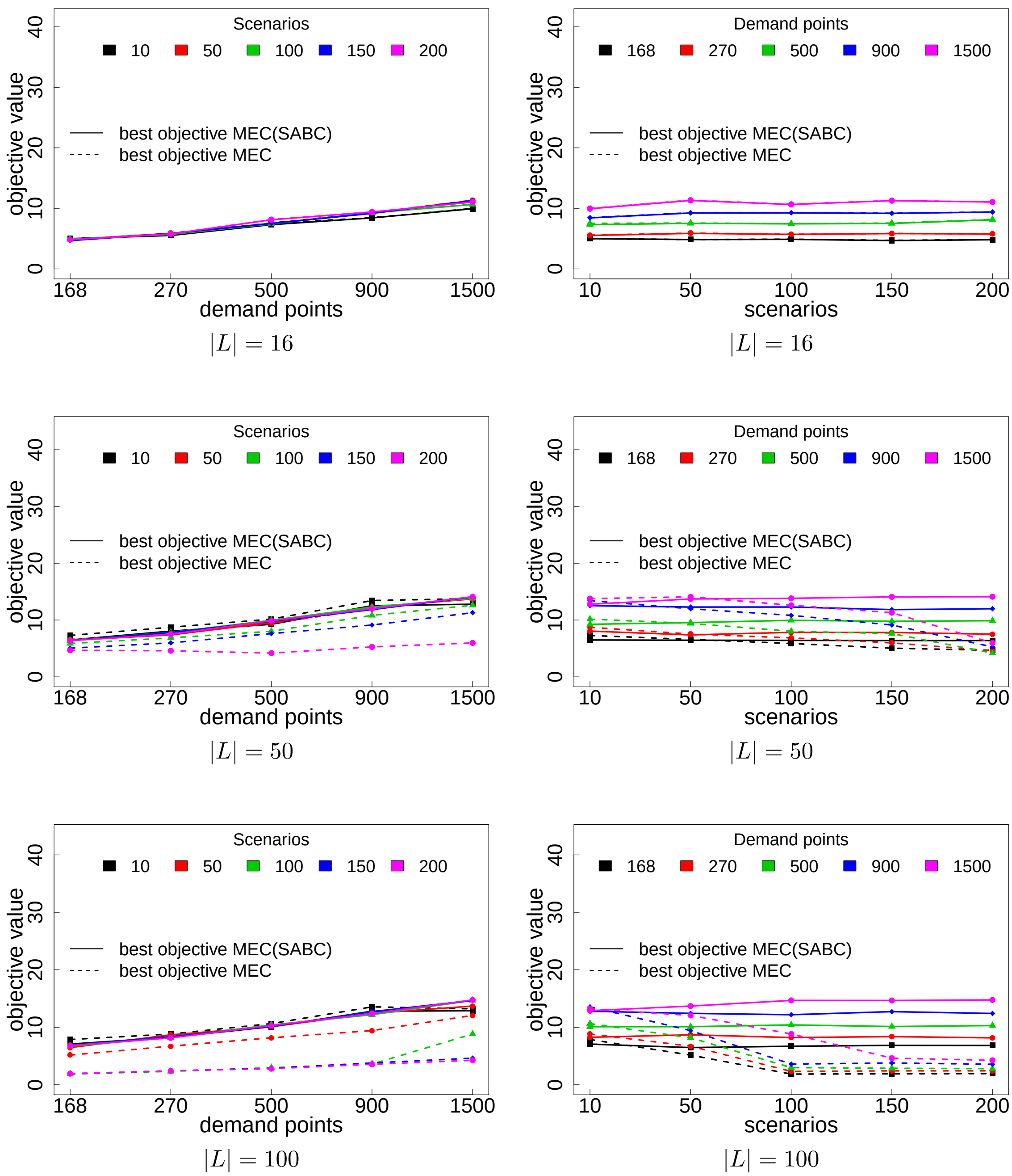


Figure 4.6: Comparison of the MEC and the MEC(SABC) model varying demand points (left-hand side column) and scenarios (right-hand side column) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.

Table 4.4: Assessment of the surrogate-based feedback method.

$\|L\|$	16		50		100	
$\|I\|, \|\Omega\|$	MEC Obj. fn. value	SBFM Obj. fn. value	MEC Obj. fn. value	SBFM Obj. fn. value	MEC Obj. fn. value	SBFM Obj. fn. value
168, 10	5.00	5.00	7.29	6.49	7.87	7.07
168, 50	4.85	4.85	6.58	6.44	5.16	6.46
168, 100	4.91	4.88	5.87	6.42	1.85	6.70
168, 150	4.71	4.68	5.04	6.38	1.91	6.85
168, 200	4.84	4.82	4.67	6.38	1.93	6.83
270, 10	5.58	5.53	8.73	8.05	8.83	8.26
270, 50	5.91	5.87	7.54	7.38	6.71	8.69
270, 100	5.72	5.70	6.86	7.83	2.33	8.21
270, 150	5.86	5.83	5.98	7.81	2.41	8.38
270, 200	5.78	5.77	4.59	7.50	2.42	8.16
500, 10	7.53	7.31	10.16	9.25	10.63	10.09
500, 50	7.57	7.52	9.41	9.55	8.17	10.10
500, 100	7.48	7.44	8.02	9.95	2.98	10.41
500, 150	7.57	7.51	7.61	9.78	2.86	10.13
500, 200	8.15	8.12	4.18	9.88	2.79	10.30
900, 10	8.47	8.43	13.44	12.54	13.55	12.79
900, 50	9.28	9.24	11.99	12.29	9.42	12.42
900, 100	9.29	9.26	10.82	12.29	3.59	12.18
900, 150	9.20	9.18	9.12	11.84	3.79	12.72
900, 200	9.40	9.39	5.27	11.99	3.56	12.40
1500, 10	9.99	9.94	13.77	12.78	13.21	12.89
1500, 50	11.36	11.31	14.08	13.71	12.06	13.69

Table 4.4: Assessment of the surrogate-based feedback method.

$\|L\|$	16		50		100	
$\|I\|, \|\Omega\|$	MEC Obj. fn. value	SBFM Obj. fn. value	MEC Obj. fn. value	SBFM Obj. fn. value	MEC Obj. fn. value	SBFM Obj. fn. value
1500, 100	10.68	10.64	12.61	13.83	8.83	14.68
1500, 150	11.29	11.28	11.30	14.08	4.63	14.66
1500, 200	11.05	11.06	5.97	14.11	4.25	14.76

As can be seen in Figure 4.6, 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.

Experiment 4: Evaluation of the NLSH. Once we analyze the MEC(SABC) model, we propose the NLSH to improve the solution we obtained with the SBFM. The objective of this experiment is to evaluate the efficiency of the heuristic. The evaluation includes the effectiveness of the neighborhoods, verifying if all the neighborhoods improve the objective function at some point.

The Figure 4.7 shows the NLSH objective function and the best bound found results. The graphs on the left-hand and the right-hand sides of the figure can be described as previous ones. The y-axis represents the objective values for all graphs. The x-axis represents the demand points and the scenarios for the left-hand and right-hand graphs, respectively. Straight lines indicate the best objective value found throughout the heuristic, i.e., the solution for the best neighbor found from the four neighborhoods explored. The results are also in Table 4.5.

Table 4.5: Evaluation of the NLSH for 75 instances.

$\|L\|$	16		50		100	
$\|I\|, \|\Omega\|$	Obj. fn. value	Best bound	Obj. fn. value	Best bound	Obj. fn. value	Best bound
168, 10	5.00	5.01	6.97	6.97	7.39	7.42
168, 50	4.85	4.85	6.56	6.56	6.47	6.47
168, 100	4.91	4.91	6.45	6.45	6.71	6.71
168, 150	4.70	4.71	6.40	6.40	6.85	6.85
168, 200	4.84	4.84	6.40	6.40	6.83	6.83
270, 10	5.58	5.58	8.35	8.35	8.55	8.58
270, 50	5.89	5.89	7.46	7.47	8.75	8.76
270, 100	5.71	5.71	7.85	7.85	8.21	8.21
270, 150	5.85	5.85	7.81	7.81	8.38	8.38
270, 200	5.77	5.77	7.50	7.50	8.20	8.20
500, 10	7.49	7.54	9.64	9.70	10.18	10.19
500, 50	7.55	7.55	9.61	9.62	10.16	10.17
500, 100	7.46	7.46	9.98	9.99	10.41	10.41
500, 150	7.54	7.55	9.80	9.80	10.18	10.18
500, 200	8.14	8.14	9.90	9.91	10.32	10.32
900, 10	8.45	8.51	12.93	12.93	13.05	13.06
900, 50	9.26	9.27	12.30	12.30	12.50	12.51
900, 100	9.27	9.27	12.29	12.29	12.19	12.19
900, 150	9.18	9.18	11.84	11.84	12.73	12.74

Table 4.5: Evaluation of the NLSH for 75 instances.

$\|L\|$	16		50		100	
900, 200	9.39	9.40	12.02	12.03	12.44	12.44
1500, 10	9.98	9.98	13.27	13.27	13.21	13.24
1500, 50	11.32	11.33	13.73	13.77	13.69	13.69
1500, 100	10.64	10.69	13.84	13.84	14.68	14.68
1500, 150	11.28	11.28	14.08	14.08	14.69	14.69
1500, 200	11.06	11.06	14.11	14.11	14.76	14.80

As we can see, even for the larger scenarios, we can obtain a solution close to the best bound, which is represented by no straight lines. This means that fixing locations facilitates exploration in the second stage of the MEC, and heuristic solutions can be found efficiently. However, we need to compare this solution with those from the MEC and MEC(SABC) to evaluate this heuristic more effectively.

Experiment 5: Assessment of the neighborhood local search heuristic. To verify if the heuristic shows better results than the MEC(SABC) model, we compare the results between the objective values obtained from the two of them. Figure 4.8 shows these results, where the axes and the left-hand and right-hand graphs are described as the previous ones. MEC(SABC) objective function values are represented by the straight lines, and the dotted lines represent the NLSH objective function values.

Results are also in Table 4.6, where the first column indicates the demand points and scenarios' sizes. Columns two, three, and four show the SBFM objective function value, the NLSH objective function value, and the improvement percentage found in NLSH, respectively, for 16 potential sites. Similarly, columns five, six, and seven display the values for 50 potential sites, while columns eight, nine, and ten

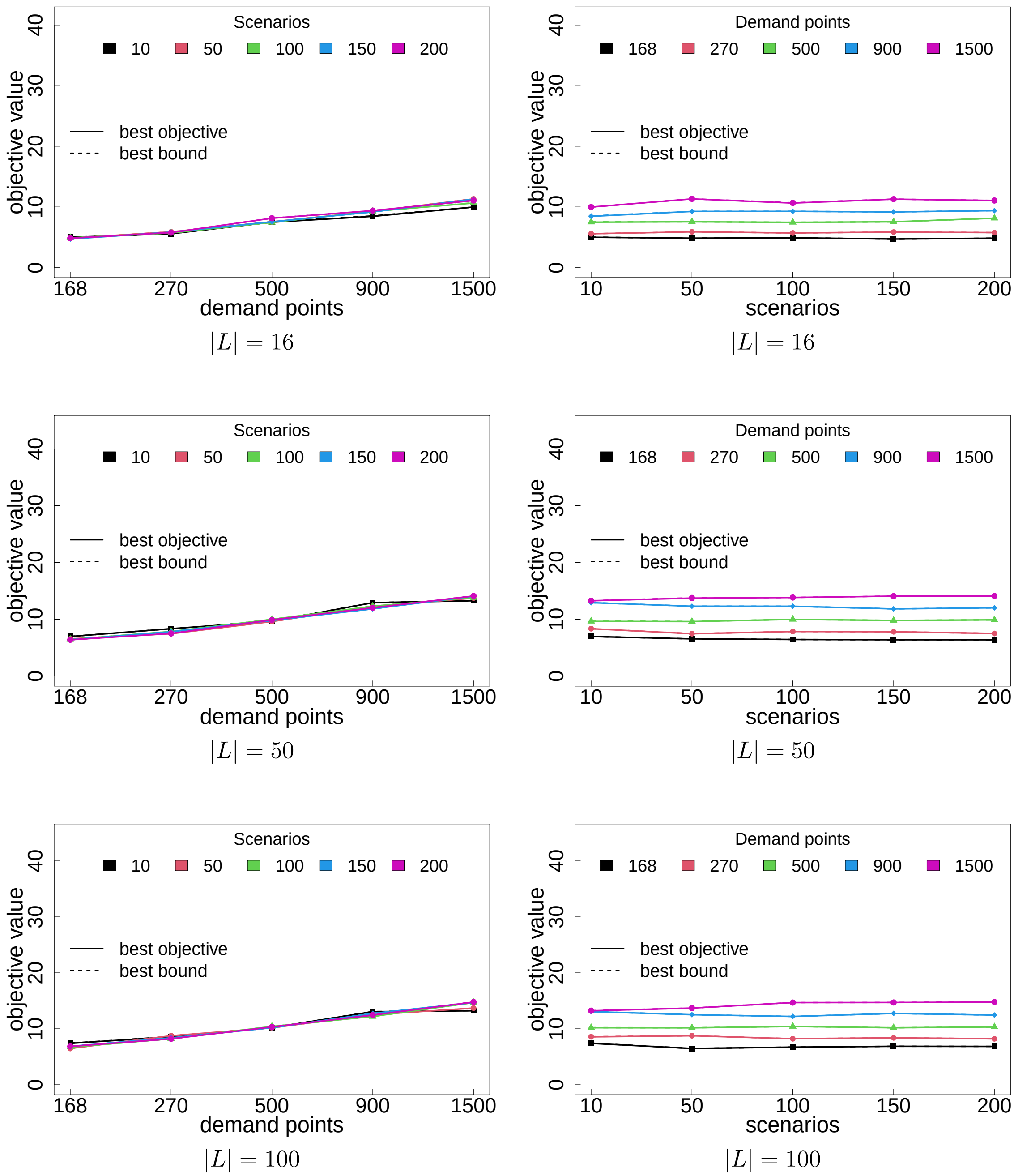


Figure 4.7: Evaluation of the NLSH varying demand points (left-hand side column) and scenarios (right-hand side column) for $|L| = 16, |L| = 50$ and $|L| = 100$.

show the results for 100 potential sites. The improvement percentage is calculated as $\frac{\text{(NLSH objective function value} - \text{SBFM objective function value)}}{\text{SBFM objective function value}} \times 100$.

Although there is no significant difference between MEC(SABC) and NLSH, the instances that yielded improved solutions in NLSH suggest that we can experiment with modifying certain aspects of NLSH to achieve better solutions in future investigations. The small improvements may be due to the computational time limit that the solver had to solve each neighbor for each neighborhood in every instance. NLSH depends on the time we allow for the neighborhoods' exploration. As the number of demand points, potential sites, and scenarios increases, the number of neighbors in each neighborhood also increases. Considering that each neighbor is verified for the MEC(SABC), the computational time required is very large if we let NLSH do a complete procedure.

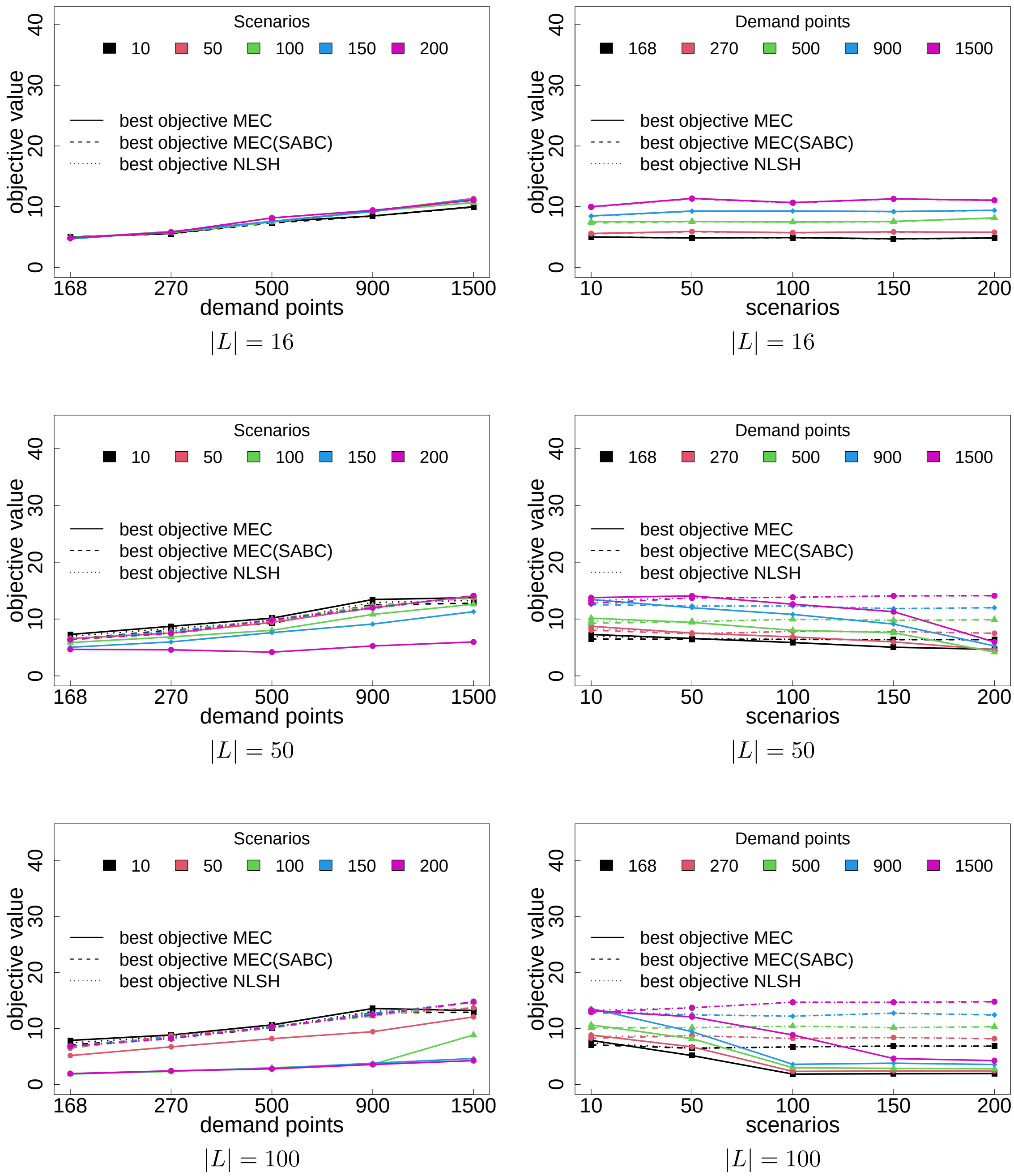


Figure 4.8: Comparissons between the objective values of the MEC, MEC(SABC) and NLSH varying demand points (left-hand side column) and scenarios (right-hand side columns) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.

Table 4.6: Comparison between the SBFM and the NLSH for 75 instances.

$\|L\|$	16			50			100		
$\|I\|, \|\Omega\|$	SBFM Obj. fn. value	NLSH Obj. fn. value	Imp. %	SBFM Obj. fn. value	NLSH Obj. fn. value	Imp. %	SBFM Obj. fn. value	NLSH Obj. fn. value	Imp. %
168, 10	5	5.01	0.00	6.49	6.97	7.40	7.07	7.42	4.96
168, 50	4.85	4.85	0.00	6.44	6.56	1.86	6.46	6.47	0.07
168, 100	4.88	4.91	0.61	6.42	6.45	0.47	6.70	6.71	0.15
168, 150	4.68	4.71	0.43	6.38	6.40	0.31	6.85	6.85	0.01
168, 200	4.82	4.84	0.41	6.38	6.40	0.31	6.83	6.83	0.01
270, 10	5.53	5.58	0.90	8.05	8.35	3.73	8.26	8.58	3.96
270, 50	5.87	5.89	0.34	7.38	7.47	1.08	8.69	8.76	0.81
270, 100	5.7	5.71	0.18	7.83	7.85	0.26	8.21	8.21	0.00
270, 150	5.83	5.85	0.34	7.81	7.81	0.00	8.38	8.38	0.01
270, 200	5.77	5.77	0.00	7.5	7.50	0.00	8.16	8.20	0.55
500, 10	7.31	7.54	2.46	9.25	9.70	4.22	10.09	10.19	1.00
500, 50	7.52	7.55	0.40	9.55	9.62	0.63	10.10	10.17	0.70

Table 4.6: Comparison between the SBFM and the NLSH for 75 instances.

$\|L\|$	16			50			100		
500, 100	7.44	7.46	0.27	9.95	9.99	0.30	10.41	10.41	0.00
500, 150	7.51	7.55	0.40	9.78	9.80	0.20	10.13	10.18	0.49
500, 200	8.12	8.14	0.25	9.88	9.91	0.20	10.30	10.32	0.23
900, 10	8.43	8.51	0.24	12.54	12.93	3.11	12.79	13.06	2.12
900, 50	9.24	9.27	0.22	12.29	12.30	0.08	12.42	12.51	0.72
900, 100	9.26	9.27	0.11	12.29	12.29	0.00	12.18	12.19	0.05
900, 150	9.18	9.18	0.00	11.84	11.84	0.00	12.72	12.74	0.13
900, 200	9.39	9.40	0.00	11.99	12.03	0.25	12.40	12.44	0.31
1500, 10	9.94	9.98	0.40	12.78	13.27	3.83	12.89	13.24	2.70
1500, 50	11.31	11.33	0.09	13.71	13.77	0.15	13.69	13.69	0.00
1500, 100	10.64	10.69	0.00	13.83	13.84	0.07	14.68	14.68	0.01
1500, 150	11.28	11.28	0.00	14.08	14.08	0.00	14.66	14.69	0.18
1500, 200	11.06	11.06	0.00	14.11	14.11	0.00	14.76	14.80	0.28

4.5.3 Response time for the MEC, MEC(SABC) and NLSH

Experiment 1: Runtime for MEC. The objective of this experiment is to verify if the MEC model can find solutions within an established computational time for each instance. Figure 4.9 shows the runtime in seconds needed to obtain MEC solutions. Left-hand side graphs have variations in demand points, while the right-hand side graphs have variations in scenarios. The graphs are for 16, 50, and 100 potential sites. As we can see, 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. This is why the MEC model often fails to find optimal solutions for most instances.

Experiment 2: Runtime for MEC(SABC). The main goal for this experiment is to verify the computational time for the SBFM. As we can see in Figure 4.10, structured as the previous one, the SBFM 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 to be solved faster than the MEC model. 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 EVCPP. The MEC(SABC) model is an approximate methodology, but it provides solutions that are as good as those of the MEC and even better when the MEC instances do not reach optimality and their gaps are large. The SBFM solves most instances in under a minute.

Experiment 3: Runtime for NLSH. The objective of this experiment is to evaluate the performance of the runtime in seconds for the NLSH. The results

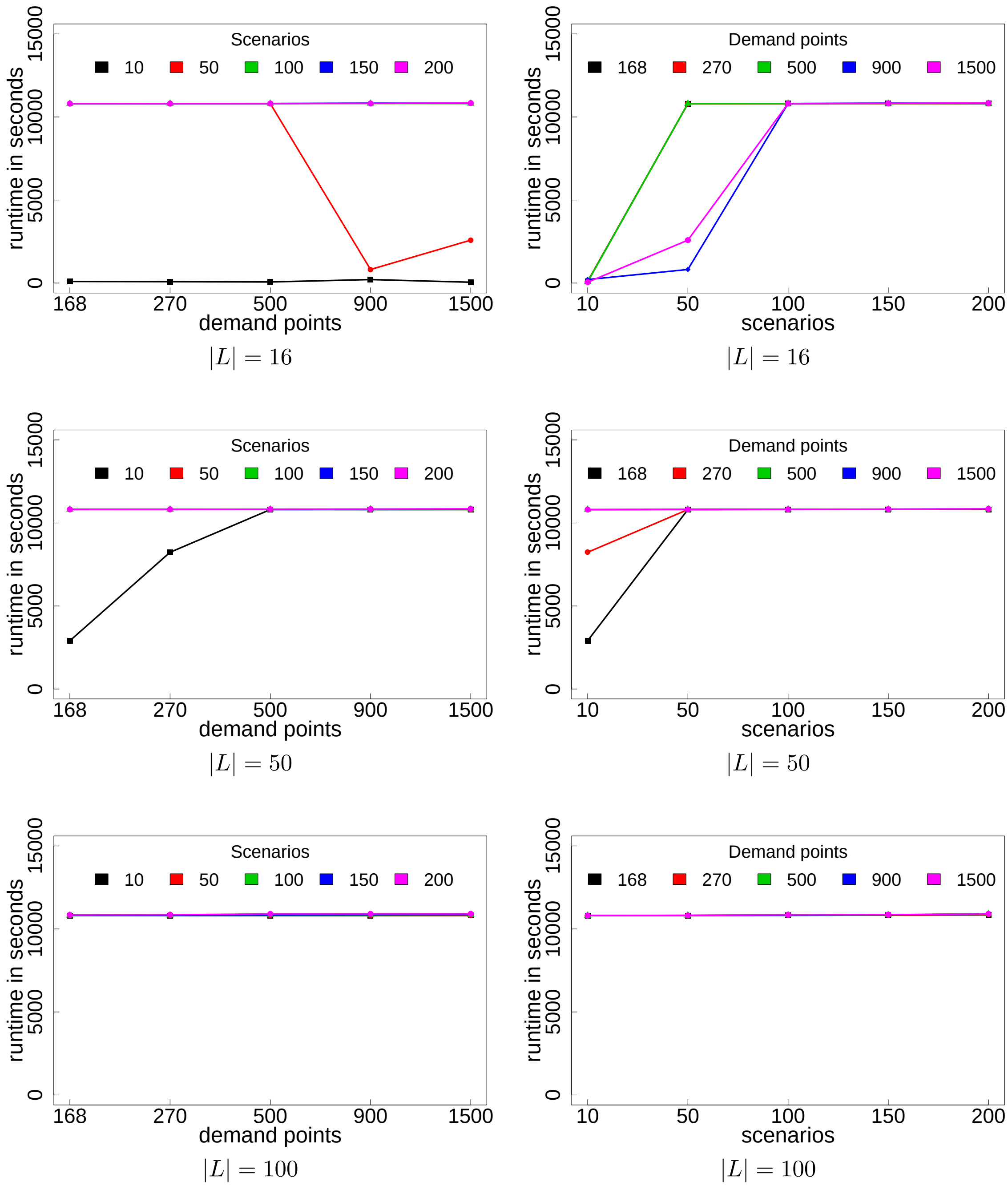


Figure 4.9: CPU time in seconds for the MEC varying demand points (left-hand side column) and scenarios (right-hand side columns) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.

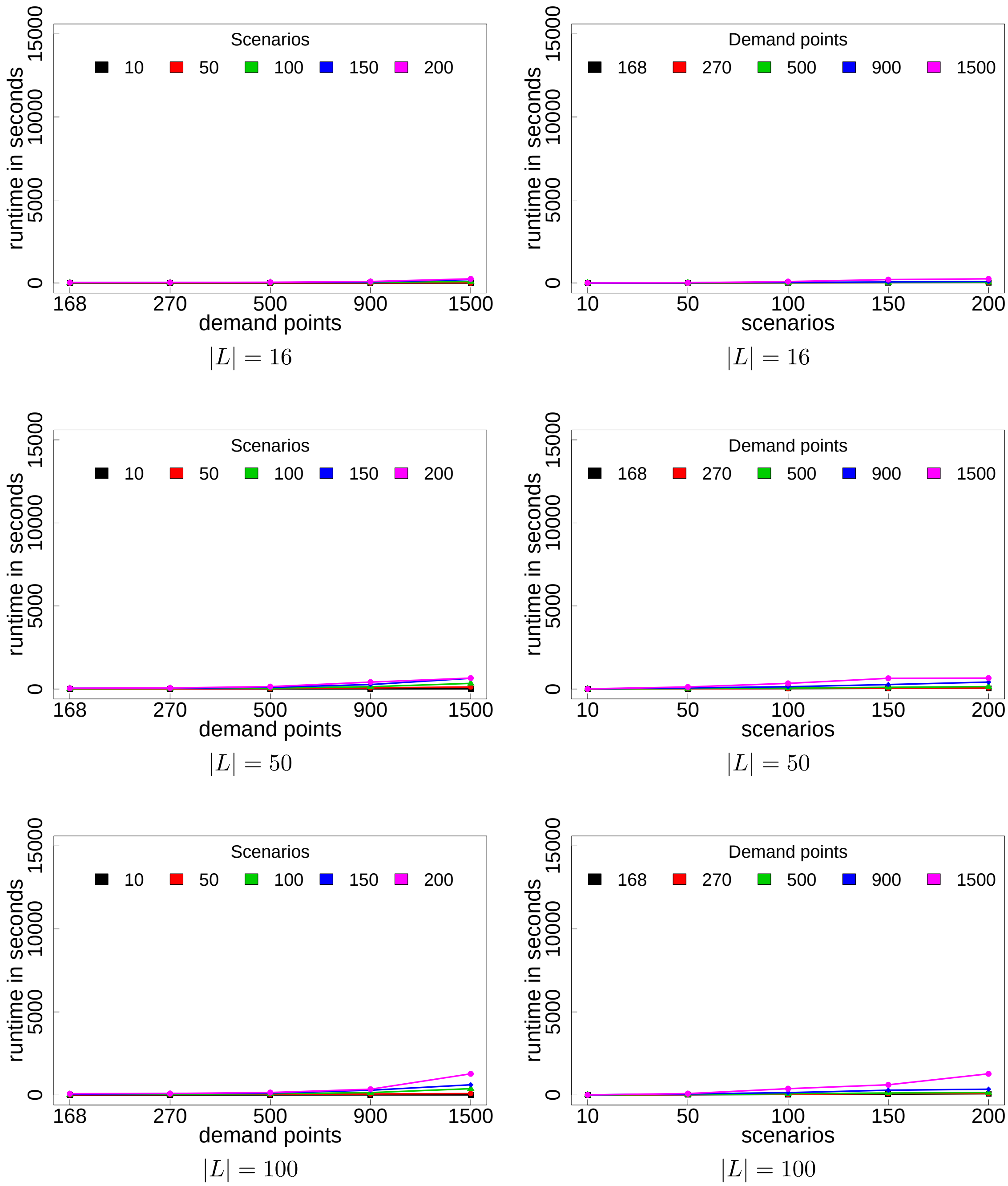


Figure 4.10: CPU time in seconds for the MEC(SABC) varying demand points (left-hand side column) and scenarios (right-hand side columns) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.

are shown in Figure 4.11, which is described same way as the previous graphics. Results show that NLSH computational time increases while the potential sites' size increases. This increment is expected due to the quantity of neighbors created at each neighborhood, which is more for 50 and 100 potential sites compared with 16 potential sites. The reason we obtained high computational times is that the MEC model is evaluated for every neighbor constructed.

MEC, SBFM, and NLSH computational times are compared in Table 4.7. MEC requires a larger computational time for most instances compared to the other two methodologies. As we can see, SBFM computational time is always less than or equal to NLSH computational time. This occurs because NLSH repeats the process SBFM does to evaluate every neighbor for each neighborhood. Still, it is not a time that should alarm us, which indicates that it is a good procedure.

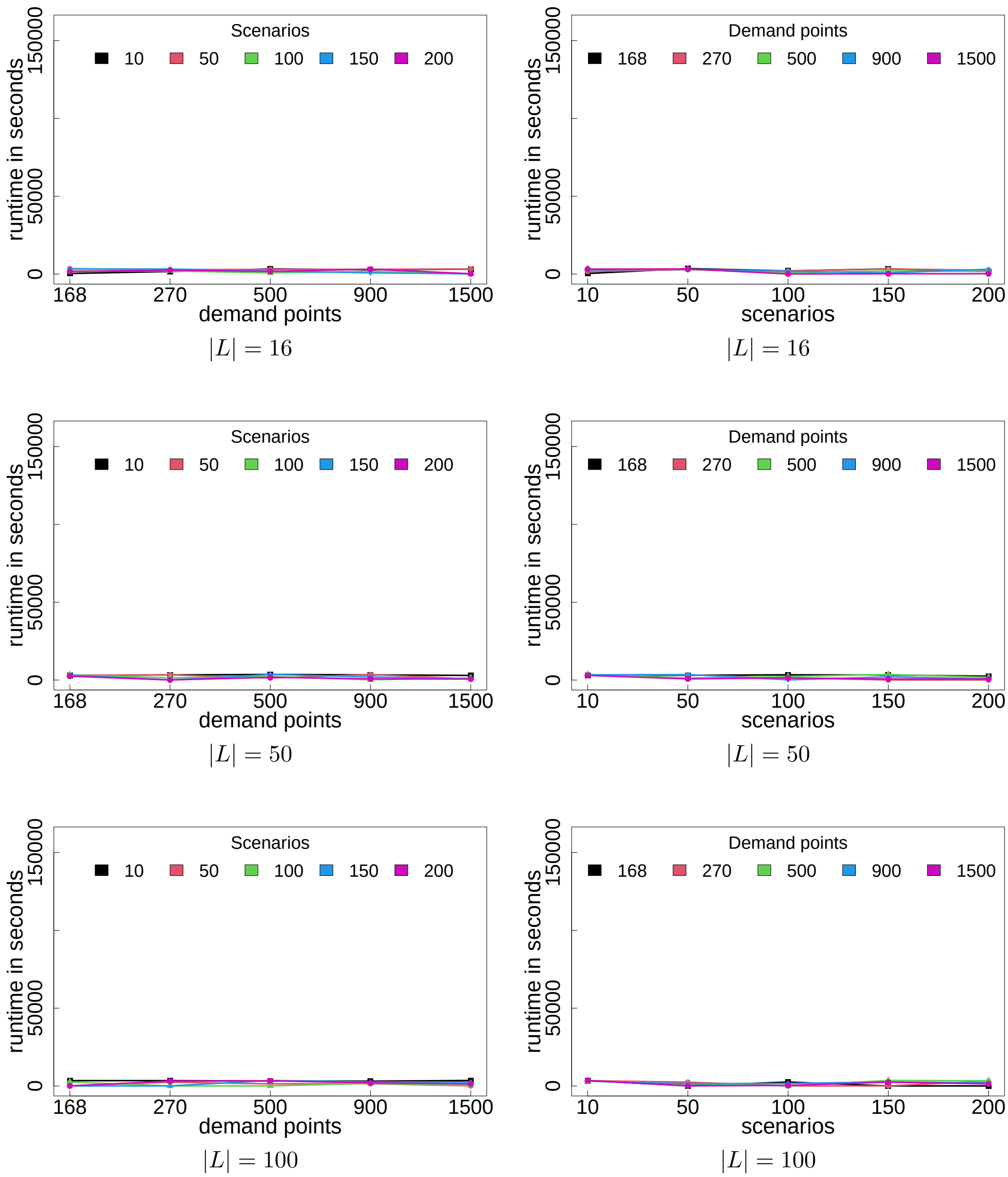


Figure 4.11: CPU time in seconds for the NLSH varying demand points (left-hand side column) and scenarios (right-hand side columns) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.

Table 4.7: Comparison of runtime in seconds between the MEC, the SBFM, and the NLSH.

$\lvert L \rvert$	16			50			100		
$\lvert I \rvert, \lvert \Omega \rvert$	MEC RT in sec.	SBFM RT in sec.	NLSH RT in sec.	MEC RT in sec.	SBFM RT in sec.	NLSH RT in sec.	MEC RT in sec.	SBFM RT in sec.	NLSH RT in sec.
168, 10	94.89	1.35	399.30	2908.63	3.12	3069.81	10802.61	4.327034	3460.58
168, 50	10801.68	6.07	3532.62	10806.89	11.86	3088	10807.96	15.60854	110.1959
168, 100	10804.23	14.18	2072.88	10810.25	24.75	3356.88	10817.49	34.00788	2553.664
168, 150	10809.77	20.35	3333.29	10813.55	41.31	3242.55	10834.65	54.08641	3627.168
168, 200	10808.75	24.97	1685.21	10817.94	47.44	2487.17	10843.44	77.58573	3639.192
270, 10	80.96	2.85	1691.55	8233.9	3.95	3287.56	10803.15	5.688764	3445.998
270, 50	10802.57	8.72	2854.17	10806.19	13.27	3330.74	10810.63	25.0612	2462.174
270, 100	10806.73	18.26	2017.38	10810.57	33.12	1390.74	10821.06	46.6905	3642.137
270, 150	10807.92	23.97	3256.04	10817.77	50.45	3615.61	10829.89	84.37731	3609.065
270, 200	10808.73	31.34	2555.05	10821.08	64.36	3625.3	10854.43	90.15973	3171.322
500, 10	64.82	3.07	3315.45	10805.04	7.58	3590.1	10804.23	7.620661	3248.521
500, 50	10804.89	10.19	3021.70	10808.06	28.69	1280.2	10813.28	32.5936	1402.964

Table 4.7: Comparison of runtime in seconds between the MEC, the SBFM, and the NLSH.

$\lvert L \rvert$	16			50			100		
500, 100	10809.84	20.29	3601.21	10814.43	57.38	2298.09	10831.04	76.93659	3632.826
500, 150	10810.07	33.04	1998.46	10821.56	103.02	3588.91	10838.22	119.9929	3581.681
500, 200	10810.94	46.16	1743.65	10828.09	144.45	1760.14	10905.89	149.2151	3336.277
900, 10	208.30	4.49	2757.75	10803.26	9.12	3334.56	10805.44	10.43479	3098.227
900, 50	815.47	15.38	3096.01	10813.11	69.66	3534.5	10819.42	64.37896	1471.902
900, 100	10809.53	42.07	1764.31	10818.95	136.54	3635.14	10834.91	142.3096	1671.386
900, 150	10830.55	64.30	3651.74	10825.67	268.54	2024.45	10850.71	289.0348	2580.19
900, 200	10814.37	90.56	3134.33	10835.75	416.23	3833.63	10907.72	347.598	2015.541
1500, 10	48.14	4.77	3204.10	10804.24	9.83	2968.26	10809.18	8.910593	3546.228
1500, 50	2582.05	18.12	3288.65	10810.78	123.77	3712.32	10819.47	84.84024	3651.345
1500, 100	10808.24	88.14	3625.25	10820.77	339.85	1391.06	10849.67	381.4811	3594.359
1500, 150	10818.30	207.51	3689.78	10832.53	644.85	3612.54	10861.18	614.6643	2480.451
1500, 200	10838.33	252.37	3797.49	10847.24	657.23	3737.14	10910.06	1276.197	1383.388

4.5.4 Coverage for the MEC, MEC(SABC) and NLSH

The objective values and running times are crucial for evaluating the models' performance. However, the most critical objective of the EVCPP is to cover the largest number of demand points within a fixed response time.

Experiment 1: Coverage for MEC. The main goal of this experiment is to evaluate the performance of the MEC model to cover the accidents that we need to cover in the EMS system. The percentage of emergency coverage for all instances for the MEC model is presented in Figure 4.12.

The figure mentioned earlier displays two columns, each with three plots, varying the number of scenarios (on the left-hand side) and the number of demand points (on the right-hand side). Each plot shows the type of ambulance percentage coverage obtained: 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 4.12 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 and Partial-late for the MEC model, which is better than not covering the emergencies. This behavior is likely related to the large gaps obtained by the MEC model for large instances; however, the number of null coverage is still remarkably low.

Experiment 2: Coverage for MEC(SABC). The objective of this experiment is to evaluate the coverage percentage for the MEC(SABC) methodology. Figure 4.13 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 timely first-aid medical care, increasing the probability of saving lives. With these results, we can conclude that SBFM performance for covering emergency calls is equally good as MEC coverage performance.

Experiment 3: Coverage for NLSH. The goal of this experiment is to evaluate the coverage percentage for NLSH. The results of this experiment are presented in Figure 4.14, which is structured similarly to the previous figures.

As we saw in Figure 4.13, the NLSH tends to have large Partial and Partial-late coverages. The results are for the best neighbor found throughout the heuristic, which are similar to the MEC(SABC) results. This means that NLSH's performance in covering emergency calls is comparable to that of the MEC and MEC(SABC) models. It even performs better as the number of scenarios and demand points increases, which was expected because the heuristics should improve the solutions of the other models.

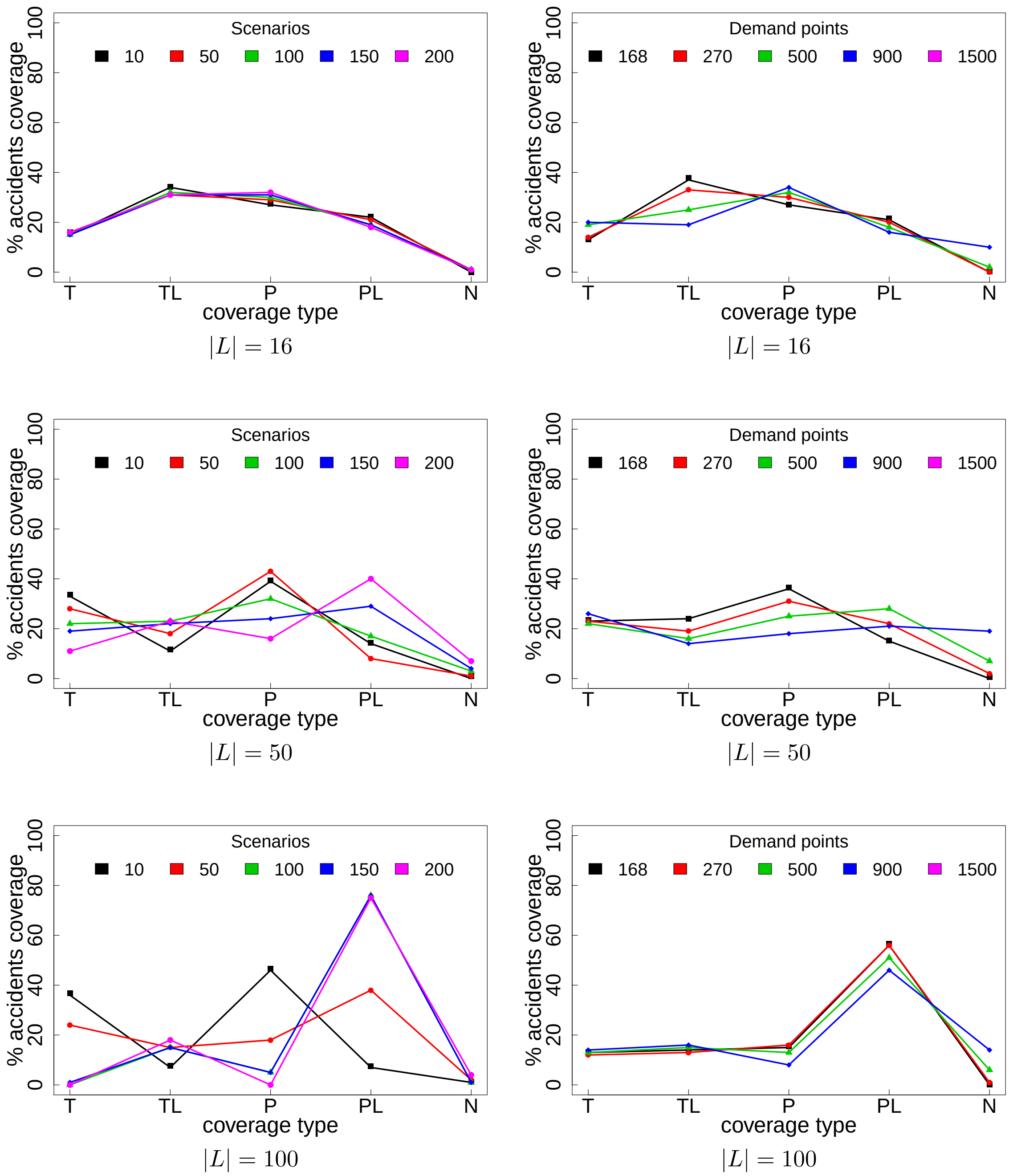
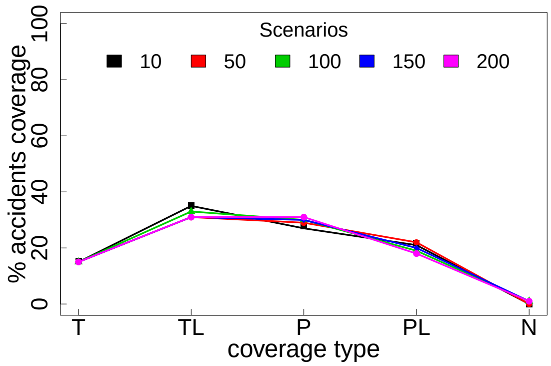
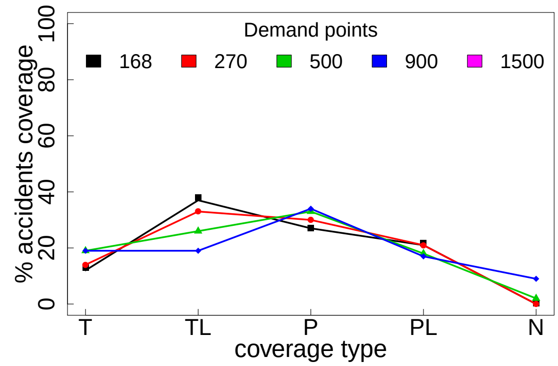


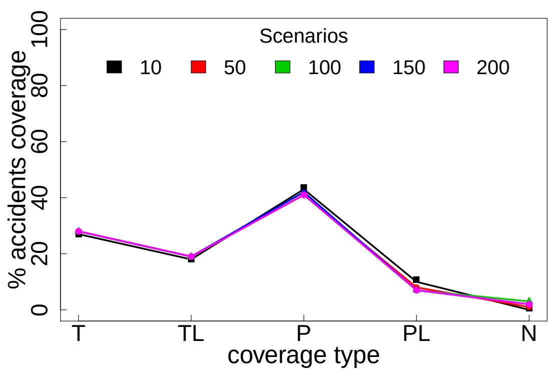
Figure 4.12: Percentage coverage type for the MEC varying demand points (left-hand side column) and scenarios (right-hand side columns) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.



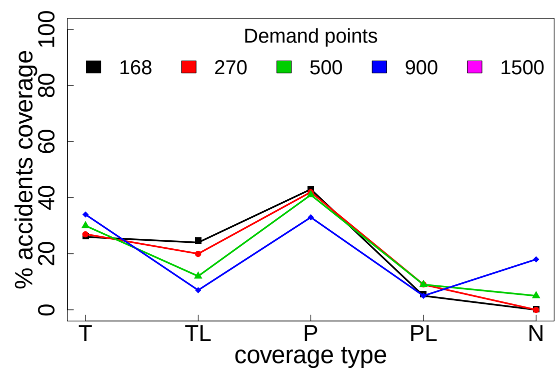
$|L| = 16$



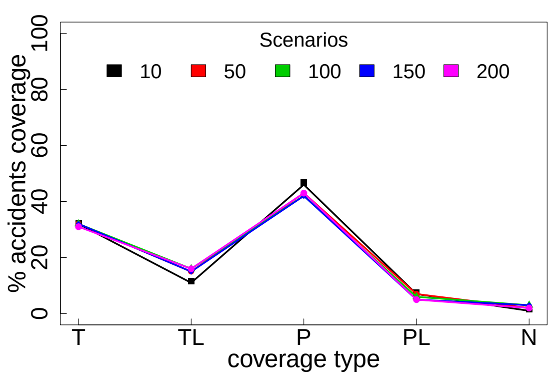
$|L| = 16$



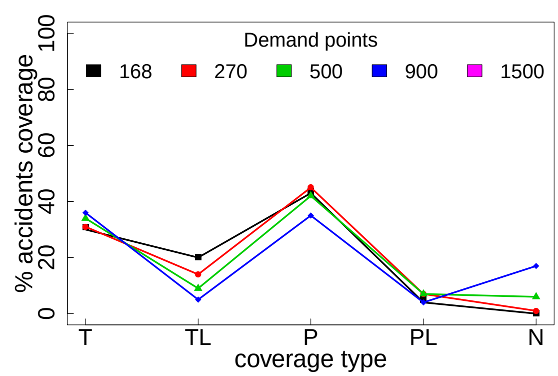
$|L| = 50$



$|L| = 50$



$|L| = 100$



$|L| = 100$

Figure 4.13: Percentage coverage type for the MEC(SABC) varying demand points (left-hand side column) and scenarios (right-hand side columns) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.

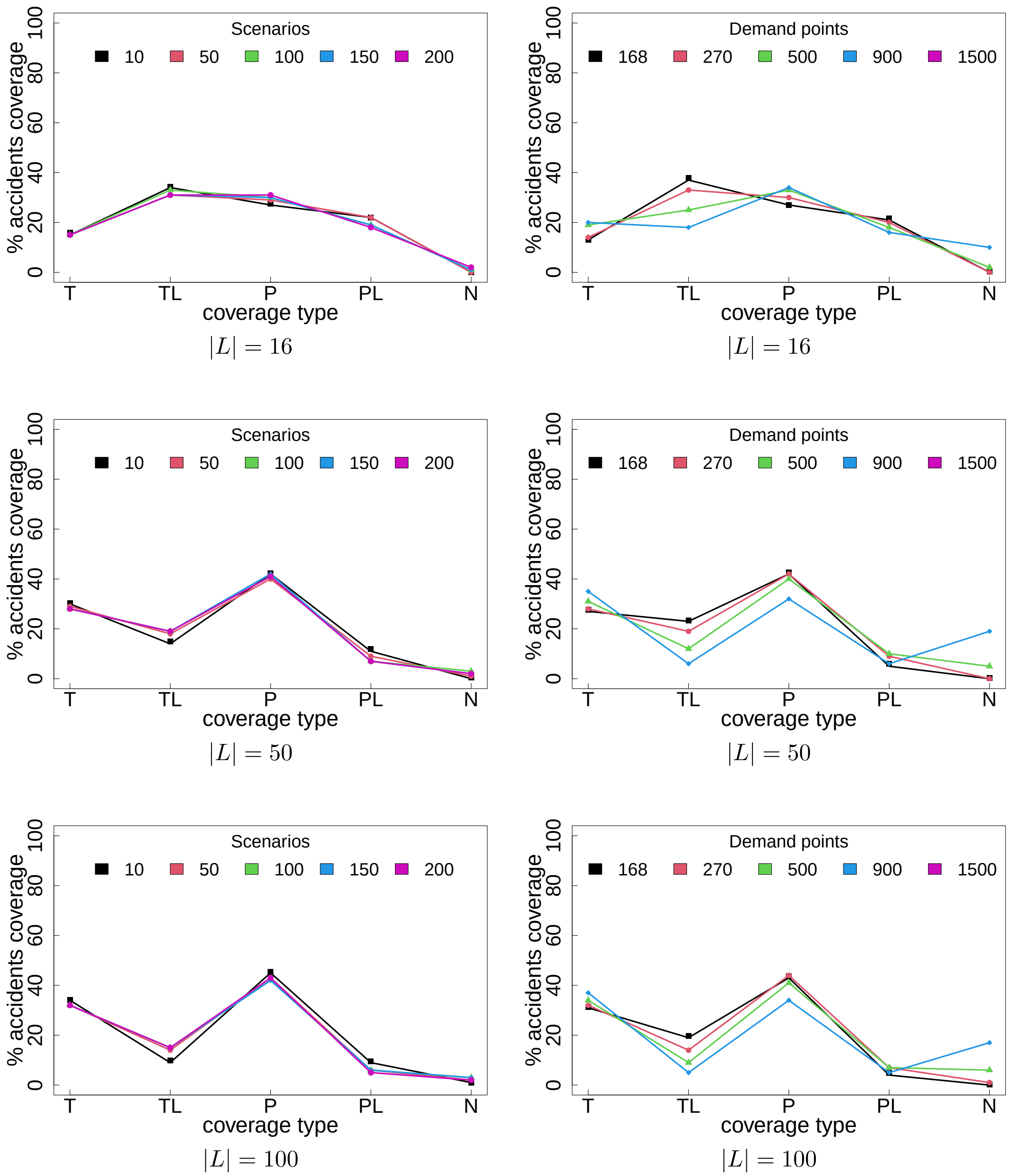


Figure 4.14: Percentage coverage type for the NLSH varying demand points (left-hand side column) and scenarios (right-hand side columns) for $|L| = 16$, $|L| = 50$ and $|L| = 100$.

4.5.5 Measures of the value of information and modeling

In this experiment, we calculate the Expected Value of Perfect Information (EVPI) and the Value of the Stochastic Solution (VSS), two concepts used in stochastic programming to evaluate the accuracy of the model [16]. The EVPI measures the maximum amount a decision maker would be ready to pay in return for complete (and accurate) information about the future. For a maximization problem, this is computed as EVPI = WS − RP, where RP is the value of the Recourse Program (our MEC model) and WS is the wait-and-see solution. In our case, we have an infinite number of scenarios for which the MEC model gives optimal solutions; thus, it is difficult to compute an exact expression. However, for instances with 900 demand points, 16 possible location points, and only 10 scenarios, we obtain small but positive EVPI values of 0.03 on average, confirming the value of the MEC model, especially as the number of scenarios will increase.

The Value of the Stochastic Solution (VSS) measures how good or, more frequently, how bad a decision based on solving the deterministic case for the average scenario with respect to the RP solution is. The VSS is computed as RP − EVV, where EVV is the expected result of using the EV solution in the MEC model. Table 4.8 shows in the first column the size of the instances for which we could obtain optimal solutions with the MEC model (demand points, location points, scenarios). The second column displays the RP value, followed by the EVV and then the VSS.

$\|I\|, \|L\|, \|S\|$	RP value	EVV	VSS
168, 16, 5	5.36	5.27	0.09
270, 16, 5	5.32	5.14	0.18
500, 16, 5	7.00	6.68	0.32

$\lvert I\rvert, \lvert L\rvert, \lvert S\rvert$	RP value	EVV	VSS
900, 16, 5	8.87	8.37	0.50
1500, 16, 5	9.29	8.44	0.85
168, 16, 10	5.00	4.64	0.36
270, 16, 10	5.58	5.2	0.38
500, 16, 10	7.53	7.13	0.39
900, 16, 10	8.47	7.87	0.59
1500, 16, 10	9.99	8.75	1.24

Table 4.8: The value of the stochastic solution for the MEC model.

Table 4.8 shows that as the size of the instance increases, the VSS also increases. Note that the small number of scenarios for which we could obtain optimal solutions is relatively small. However, the behavior of the VSS shows the benefit of considering a stochastic setting even for instances with few scenarios.

4.5.6 Number of ambulances

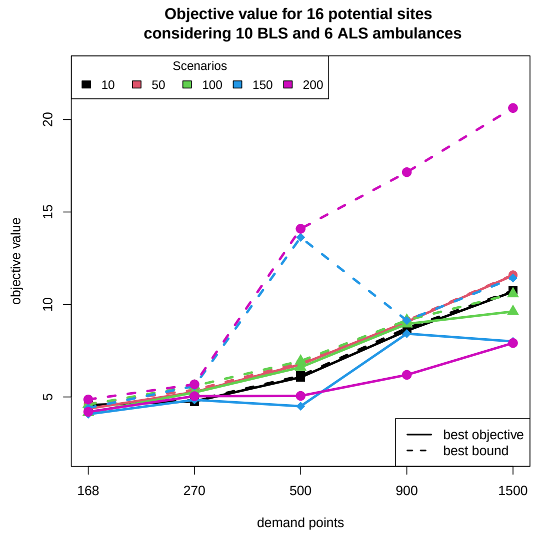
All previous experiments were conducted with the number of ambulances equal to $(\eta_1, \eta_2) = (35, 20)$. A central feature of the EVCPP is that an ALS ambulance can be sent instead of a BLS one, which provides a more flexible setting but may introduce difficulty when solving the models.

For a preliminary evaluation, we conducted experiments varying the number of ambulances for the MEC problem, considering $|L| = 16$ potential sites for all demand point sizes and scenarios. Figure 4.15 presents objective values. As we expected, the number of ambulances increases the gap between best objective and best bound found in some cases, mostly when the number of scenarios increases for

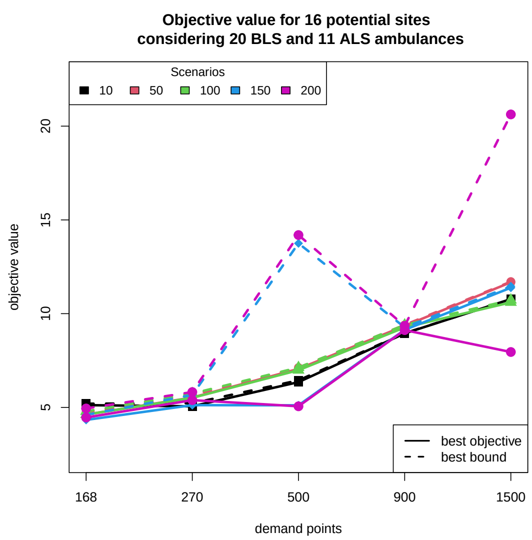
(η_1, η_2)=(10,6) and (20,11). Although the number of ambulances increases, the gap decreases. We decided to reduce our instance set to (η_1, η_2)=(35,20) to obtain more realistic results, considering the number of ambulances in our case of the EMS system in Nuevo León. In addition, as shown in Figure 4.16, there is no significant difference in runtime when varying the number of ambulances. Therefore, considering a small number of them does not significantly reduce the model's computational time.

We solved 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 4.17 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 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.

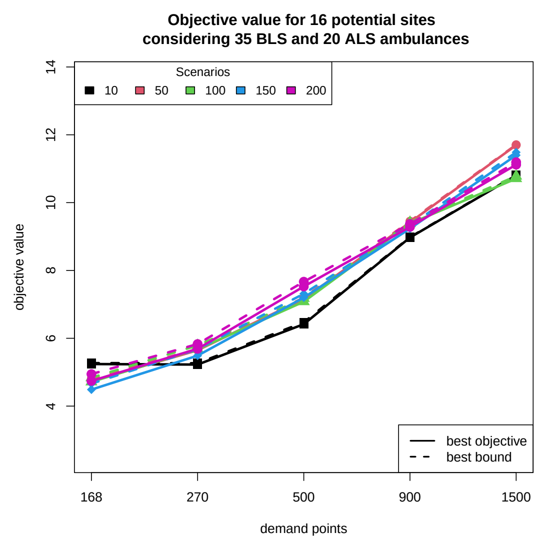
From Figure 4.17a, we observe that the difference between the best objective and the best bound for the MEC model (left-hand side plots) increases slightly with the number of ambulances. Thus, the larger the number of ambulances, the harder the instances for the MEC model. Furthermore, the time limit is reached for every instance of the MEC model that is tested. For the SBFM (right-hand side plots), the relative optimality gaps are equal to 0 for all instances. In addition, the objective values are comparable to characteristic. Furthermore, under the SBFM, all instances are solved in less than one minute, and this time is not affected by the number of ambulances.



Considering (η_1, η_2)=(10,6)

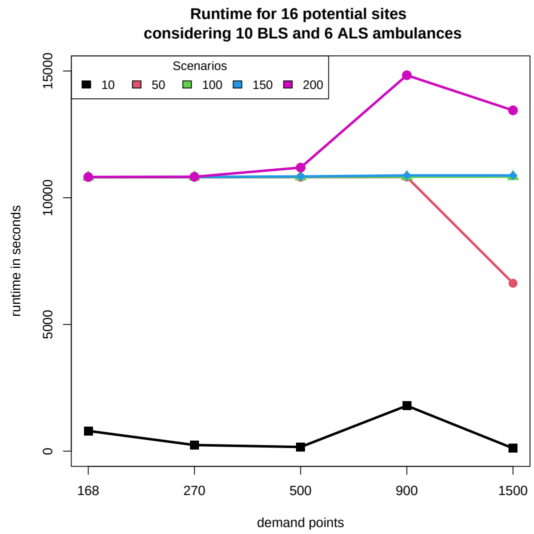


Considering (η_1, η_2)=(20,11)

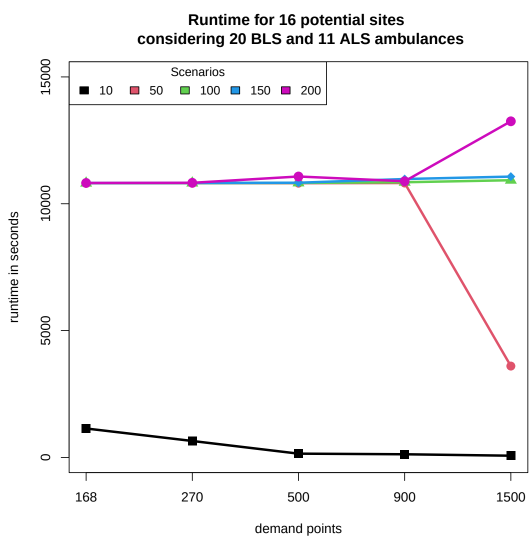


Considering (η_1, η_2)=(35,20)

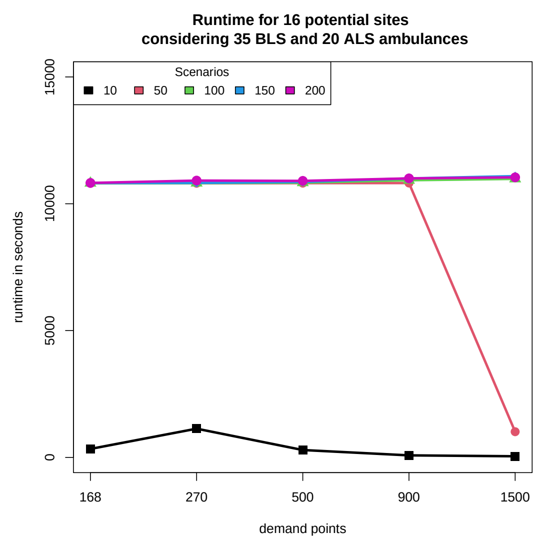
Figure 4.15: Objective function values for preliminary experiments to evaluate different numbers of ambulances.



Considering (η_1, η_2)=(10,6)



Considering (η_1, η_2)=(20,11)



Considering (η_1, η_2)=(35,20)

Figure 4.16: CPU time in seconds for preliminary experiments to evaluate different numbers of ambulances.

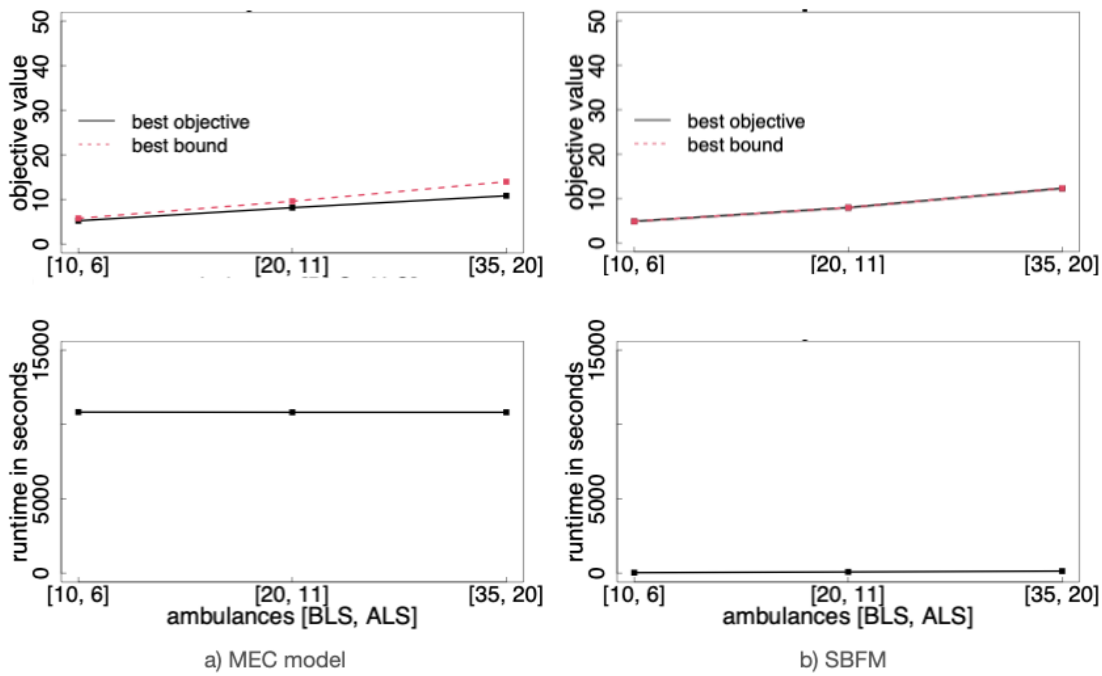


Figure 4.17: Objective value and running time versus the number of ambulances for a) MEC model and b) SBFM.

Chapter 5

Conclusions

5.1 Main findings

EMS systems in developing countries, such as Mexico, face a significant shortage of ambulances. Therefore, one of the main objectives of this work is to investigate and develop tools that facilitate the optimal location of ambulances. This would improve the performance of EMS systems and ensure that all emergency calls are fully addressed, whether they are currently uncovered or only partially covered.

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

We propose a two-stage stochastic integer program for the EVCPP that can be solved utilizing the MEC formulation 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 solving 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 programming solver, without requiring the use of complex decomposition techniques.

Besides, a local search heuristic was proposed to improve the quality of the solutions. This heuristic involves improving SBFM solutions, treating them as initial solutions for the NLSH. The heuristic is structured to perform a local search around four neighborhoods that perturb the ambulance location, which are the neighbors, to improve the objective function value of the MEC problem, which is evaluated by fixing the ambulance location variables. There were slight improvements in the solutions.

In this manner, we have proposed models and heuristic methodologies to solve the ECVP problem. Moreover, we have tackled a real-world problem, and now we can test it in practice.

The development of the SBFM heuristic is very important. Large-scale instances proved intractable by solving the MEC model. In those experiments, it was clear that the branch-and-bound method ran out of memory. The proposed heuristic was able to improve solution quality of a number of instances. In many cases this improvement was relatively small.

In terms of modeling, the introduction of partial rate coverage into the objective function allows sending ambulances even if there are not enough ambulances to cover an accident totally or if an ambulance or more than one ambulance are far away from the desired response time for patient attention. These results help us cover more demand points in the system, allowing us to start giving attention to patients, which can be finished after providing first aid to them at those demand points where not all ambulances were sent.

5.2 Future work

Our future work involves collaboration with more than one service provider in the system, taking into account the differences between them and the preferences that public ambulances may have compared to private ambulances. In fact, in Monterrey, there are at least three emergency service providers that cooperate and compete with each other: *Cruz Roja, Protección Civil, and CRUM.* Thus, game theory or bi-level programming could be used to determine the best policies in such a way that the population is benefited.

Naturally, several lines of work can be further investigated. For example, another interesting aspect we observed is that there are some private EMS services that also dispatch vehicles to accident sites. Some of these are neither regulated nor coordinated by the state. In some cases, this provokes a conflict as too many ambulances arrive at the site, leaving other points unattended. This situation could, of course, benefit from coordinated decision-making tools as those developed here.

Regarding the methodology, to solve the preliminary model and future models, one could benefit from developing appropriate valid inequalities. The difficulty lies in that our problem involves integer and binary variables in the second stage of the stochastic programming; thus, we must evaluate whether considering real variables yields feasible solutions and how much the reduction in quality is.

Also, one could consider including queues at hospitals. During the early stages of the COVID-19 pandemic, some hospitals focused solely on treating COVID-19 patients, resulting in other hospitals experiencing ambulance queues due to an overwhelming demand. This wasted time waiting for attention affects ambulance availability and should be accounted for in the EMS system.

Another approach is to consider scenarios by clusters, as done by Hewitt et al. [37]. It could reduce the size of the stochastic program, and it may not be necessary to use the Benders decomposition; instead, it may be competitive with heuristic methods.

Finally, the proposed heuristic is a simple local search that uses four neighborhood structures. These components could be further enhanced if cast into a more sophisticated metaheuristic framework such as tabu search or scatter search.

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. Ahlenius, V. Lindström, and V. Vicente. Patients' experience of being badly treated in the ambulance service: A qualitative study of deviation reports in Sweden. *International Emergency Nursing*, 30:25–30, 2017.

[3] A. Akıncılar and E. Akıncılar. A new idea for ambulance location problem in an environment under uncertainty in both path and average speed: Absolutely robust planning. *Computers & Industrial Engineering*, 137:106053, 2019.

[4] 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.

[5] S. Ansari, L. Albert McLay, and M. E Mayorga. A maximum expected covering problem for district design. *Transportation Science*, 51(1):376–390, 2017.

[6] 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.

[7] 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.

[8] M. O. Ball and F. L. Lin. A reliability model applied to emergency service vehicle location. *Operations Research*, 41(1):18–36, 1993.

[9] 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.

[10] R. Batta, M. Dolan, J, and N. N. Krishnamurthy. The maximal expected covering location problem: Revisited. *Transportation science*, 23(4):277–287, 1989.

[11] 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.

[12] 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.

[13] P. L. V. D. Berg and J. T. V. Essen. Comparison of static ambulance location models. *International Journal of Logistics Systems and Management*, 32(3-4): 292–321, 2019.

[14] 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.

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

[16] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, USA, 2nd edition, 2011.

[17] T. H. Blackwell and J. S. Kaufman. Response time effectiveness: comparison of response time and survival in an urban emergency medical services system. *Academic Emergency Medicine*, 9(4):288–295, 2002.

[18] 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*, 30:123–152, 2018.

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

[20] L. Brotcorne, G. Laporte, and F. Semet. Ambulance location and relocation models. *European Journal of Operational Research*, 147(3):451–463, 2003.

[21] S. Chanta, M. Mayorga, and L. McLay. A hybrid tabu search for locating and dispatching emergency response units using an imbedded queuing model. In *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*, Orlando, USA, May 2012.

[22] S. Chanta, M. E Mayorga, and L. A. McLay. Improving emergency service in rural areas: a bi-objective covering location model for EMS systems. *Annals of Operations Research*, 221:133–159, 2014.

[23] R. Church and C. R. Velle. The maximal covering location problem. *Papers in Regional Science*, 32(1):101–118, 1974.

[24] N. Cook. Multiple operational resource on scene procedure. Technical report, South Central Ambulance Service NHS Foundation Trust, Oxfordshire,

UK, November 2018. Available at: https://www.scas.nhs.uk/wp-content/uploads/2022/02/Multiple-Operational-Resource-on-Scene-Policy-Accessible.pdf.

[25] M. S. Daskin. A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation Science*, 17(1):48–70, 1983.

[26] Dirección de Comunicación Social. La secretaría de movilidad tiene un registro de 487 ambulancias en la ciudad de méxico. https://infocdmx.org.mx/index.php/2-boletines/5754-dcs-013-18.html#:~:text=La%20Secretar%C3%ADa%20de%20Movilidad%20de,por%20el%20Pleno%20del%20INFODF., 28/02/2018. Last accessed: 10-June-2025. In Spanish.

[27] 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.

[28] S. Enayati, M. E. Mayorga, H. Toro-Díaz, and L. A. Albert. Identifying trade-offs in equity and efficiency for simultaneously optimizing location and multipriority dispatch of ambulances. *International Transactions in Operational Research*, 26(2):415–438, 2019.

[29] Excelsior. Así opera el 911 ante una llamada de emergencia. News note, January 2017. In Spanish. Available at: https://www.excelsior.com.mx/nacional/2017/01/19/1140428. Last accessed: 04-Mar-2025.

[30] E. Fadda, D. Manerba, G. Cabodi, P. Camurati, and R. Tadei. Comparative analysis of models and performance indicators for optimal service facility location. *Transportation Research Part E: Logistics and Transportation Review*, 145:102174, 2021.

[31] E. Fadda, D. Manerba, and R. Tadei. How to locate services optimizing redundancy: A comparative analysis of K-covering facility location models. *Socio-Economic Planning Sciences*, 94:101938, 2024.

[32] F. García-Alfranca, A. Puig, C. Galup, H. Aguado, I. Cerdá, M. Guilabert, V. Pérez-Jover, I. Carrillo, and J. J. Mira. Patient satisfaction with pre-hospital emergency services. a qualitative study comparing professionals' and patients' views. *International Journal of Environmental Research and Public Health*, 15 (2):233, 2018.

[33] M. Gendreau, G. Laporte, and F. Semet. Solving an ambulance location model by tabu search. *Location Science*, 5(2):75–88, 1997.

[34] 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.

[35] W. J. Gutjahr. Recent trends in metaheuristics for stochastic combinatorial optimization. *Central European Journal of Computer Science*, 1:58–66, 2011.

[36] R. D. Gálvao 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.

[37] M. Hewitt, J. Ortmann, and W. Rei. Decision-based scenario clustering for decision-making under uncertainty. *Annals of Operations Research*, 315(2):747–771, 2022.

[38] K. Hogan and C. ReVelle. Concepts and applications of backup coverage. *Management Science*, 32(11):1434–1444, 1986.

[39] 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*, 1:129–175, 2012.

[40] PRONAMED Salud Integral. Ambulancia para atención de urgencias prehospitalarias 24/7. `https://bibtex.eu/es/faq/how-can-i-use-bibtex-to-cite-a-website/`, 2018. Last accessed: 20-June-2025. In Spanish.

[41] W Jian, L. Han, A. Shi, and C. Na. A new partial coverage locating model for cooperative fire services. *Information Sciences*, 373:527–538, 2016.

[42] S. M. Kabene, C. Orchard, J. M. Howard, M. A. Soriano, and R. Leduc. The importance of human resources management in health care: a global context. *Human Resources for Health*, 4:1–20, 2006.

[43] 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.

[44] F. Khosgehbari and M. J. Mirzapour Al-e Hashem. Ambulance location routing problem considering all sources of uncertainty: Progressive estimating algorithm. *Computers & Operations Research*, 160:106400, 2023.

[45] S. S. W. Lam, J. Zhang, Z. C. Zhang, H. C. Oh, J. Overton, Y. Y. Ng, and M. E. H. Ong. Dynamic ambulance reallocation for the reduction of ambulance response times using system status management. *The American Journal of Emergency Medicine*, 33(2):159–166, 2015.

[46] J. Layton. How 9-1-1 works. HowStuffWorks web site `https://people.howstuffworks.com/9-1-1.htm`, July 2006. Last accessed: 04-Mar-2025.

[47] 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*, 74:281–310, 2011.

[48] M. B. Mandell. Covering models for two-tiered emergency medical services systems. *Location Science*, 6(1-4):355–368, 1998.

[49] V. Marianov and C. ReVelle. The queueing maximal availability location problem: a model for the siting of emergency vehicles. *European Journal of Operational Research*, 93(1):110–120, 1996.

[50] M. E. Mayorga, D. Bandara, and L. A. McLay. Districting and dispatching policies for emergency medical service systems to improve patient survival. *IIE Transactions on Healthcare Systems Engineering*, 3(1):39–56, 2013.

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

[52] L. A. McLay and M. E. Mayorga. Evaluating emergency medical service performance measures. *Health Care Management Science*, 13:124–136, 2010.

[53] N. Megiddo and A. Tamir. On the complexity of locating linear facilities in the plane. *Operations Research Letters*, 1(5):194–197, 1982.

[54] 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.

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

[56] Nuevo Leon Government. Fortalece estado cobertura prehospitalaria con 30 nuevas ambulancias. Informative bulletin, 12 December 2024.

In Spanish. Available at: https://www.nl.gob.mx/es/boletines/fortalece-estado-cobertura-prehospitalaria-con-30-nuevas-ambulancias#:~:text=en%20la%20entidad.-,En%20la%20administraci%C3%B3n%20del%20Gobierno%20del%20nuevo%20Nuevo%20Le%C3%B3n%20la,de%205%20a%2047%20. Last accessed: 11-Mar-2025.

[57] 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.

[58] D. Olave-Rojas and S. Nickel. Modeling a pre-hospital emergency medical service using hybrid simulation and a machine learning approach. *Simulation Modelling Practice and Theory*, 109:102302, 2021.

[59] J. Ong, D. Kulpanowski, Y. Xie, E. Nikolova, and N. M. Tran. OpenEMS: An open-source package for two-stage stochastic and robust optimization for ambulance location and routing with applications to Austin-Travis County EMS data. ArXiv preprint arXiv:2201.11208, 2022. URL https://arxiv.org/abs/2201.11208.

[60] 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.

[61] C. ReVelle and K. Hogan. The maximum availability location problem. *Transportation Science*, 23(3):192–200, 1989.

[62] C. ReVelle, C. Toregas, and L. Falkson. Applications of the location set-covering problem. *Geographical Analysis*, 8(1):65–76, 1976.

[63] S. A. Rodriguez, A. Rodrigo, and M. M. Aguayo. A facility location and equipment emplacement technique model with expected coverage for the

location of fire stations in the concepción province, chile. *Computers & Industrial Engineering*, 147:106522, 2020.

[64] 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.

[65] 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, 2016.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] Y. Yuan, Q. Song, and B. Zhou. A Wasserstein distributionally robust chance constrained programming approach for emergency medical system planning problem. *International Journal of Systems Science*, 53(10):2136–2148, 2022.

[73] M. A. Zaffar, H. K. Rajagopalan, C. Saydam, M. Mayorga, and E. Sharer. Coverage, survivability or response time: A comparative study of performance statistics used in ambulance location models via simulation–optimization. *Operations Research for Health Care*, 11:1–12, 2016.

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

[75] 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

Candidata para obtener el grado de

Doctora en Ciencias en Ingeniería de Sistemas

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Tesis:

Locating and Dispatching Two Types of Ambulances Considering Partial Coverage: Stochastic Integer Programming Models and Heuristics

Desde mi nacimiento, siendo la fecha del mismo un 20 de enero de 1995, mis padres Beatriz Ramos Larralde y Jesús García Gámez han estado para mí en todo momento. Ambos nos han educado y guiado por el camino de la fe católica a mi hermana Karina Guadalupe García Ramos y a mí hasta la fecha. En el año de 2017 concluí mis estudios en la Facultad de Ciencias Físico Matemáticas de la Universidad Autónoma de Nuevo León como Licenciada en Matemáticas. Posteriormente, en el año 2019, obtuve mi grado como Maestra en Ciencias en Ingeniería de Sistemas en la Facultad de Ingeniería Mecánica y Eléctrica, la cual pertenece a la Universidad

Autónoma de Nuevo León. En esta misma facultad comencé a estudiar en el programa de Doctorado en Ingeniería de Sistemas en el año 2020.