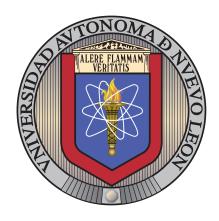# Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado

POR

Beatriz Alejandra García Ramos

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

DOCTORADO EN CIENCIAS EN INGENIERÍA DE SISTEMAS

2024

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Subdirección de Estudios de Posgrado

POR

Beatriz Alejandra García Ramos

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

DOCTORADO EN CIENCIAS EN INGENIERÍA DE SISTEMAS

2024

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

Los miembros del Comité de Tesis recomendamos que la Tesis ¡¡¿¿, realizada por el alumno Beatriz Alejandra García Ramos, con número de matrícula 1550385, sea aceptada para su defensa como requisito parcial para obtener el grado de Doctorado en Ciencias en Ingeniería de Sistemas.

El Comité de Tesis

_____
Dr. Roger Z. Ríos Mercado
Asesor

_____   _____
Dra. Yasmín Águeda Ríos Solís   Nombre del revisor B
Revisor   Revisor

_____   _____
Nombre del revisor C   Nombre del revisor D
Revisor   Revisor

Vo. Bo.

_____
Subdirector de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León, 2024

CONTENTS

LIST OF FIGURES

List of Tables

ACKNOWLEDGMENT

laboratorio en dónde trabajar y salones donde podamos tomar las materias que se nos imparten.

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

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

ABSTRACT

Beatriz Alejandra García Ramos.

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

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: .

Número de páginas: 23.

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

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

One of the contributions is to introduce partial rate coverage to this type of model. Commonly partial coverage is used in deterministic models due to its simplicity. Another contribution is to involve more than one service provider in the system. In Mexico, public and private providers belong to the 911 EMS system, and most of the literature considers only one service provider.

The objective is to improve Mexico's 911 system by coordinating ambulance location and dispatching for all service providers to maximize patient attention at the minimum response time possible.

Firma del director: ——————————————————
<div align="center">Dr. Roger Z. Ríos Mercado</div>

Introduction

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

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

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

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

There exist different types of problems. Various focus on statics systems, where the decisions are fixed after they are taken. Others are dynamic [9], where decisions change throughout time passing or after a change on the system, for example, when a call is received. Many investigations focus their problems on solving optimal locations of their ambulances to improve the system. In contrast, others want to obtain optimal policies to decide which ambulance or ambulances are the best to attend the emergency calls received.

EMS systems have different strategic, tactical, and operational types of planning [11]. Strategic planning focuses on long-term decisions, such as fixed potential locations or the acquisition of resources [1]. Tactical plannings made decisions for a middle time, such as locating ambulances at potential sites or planning which is the best option to dispatch an ambulance for all emergency calls types. Finally, operational planning takes decisions in a short time. These decisions are made frequently when a call enters the system and are divided into online and offline decisions. Online decisions are taken over the full service of a call, and offline decisions are taken when a call is received following the plannings made at tactical plannings, for example, to decide which ambulance will send to attend to the patient [12].

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

The problem in our country is that more than one service provider is involved in the 9-1-1 system. The two most important are CRUM (Centro Regulador de Urgencias Medicas) and Red Cross. There exists in some states the Proteccion Civil and the Green Cross too, but CRUM controls them. Also, other private organizations exist, but they are not involved in the 9-1-1 system. That is one reason why an uncoordinated system exists in our country. CRUM is part of the government, and they control the public ambulances, while Red Cross is a private organization. Each service provider decides where to locate their ambulances without consulting the

other. Furthermore, when an ambulance is needed to attend an emergency call for 9-1-1, CRUM has preferences because Red Cross has its private number where people can call for medical attention. Red Cross dispatch ambulances that are supposed to be available in 9-1-1 systems for its direct emergency calls.We could improve this uncoordinated system if each service provider made its decisions considering others. Moreover, the system has the problem of queues at hospitals, which delays the patients' attention. Since the pandemic for Covid-19 started, some hospitals attend only Covid-19 patients. That causes the rest of the hospitals to have more demand than before the pandemic. So the ambulances from the 9-1-1 system are busy longer than before, and it has to be considered to improve the EMS system in Mexico.

We propose a two-stage stochastic programming model with recourses for ambulance location and dispatching, considering two service providers to obtain a coordinated EMS system to solve those problems. The following sections present the background investigations about EMS systems (Chapter 2) and the usually used models. We describe the problem and factors that affect the EMS system in Mexico (Chapter 3). Then, we solve the problem and define the model (Chapter 5) used to do the experiments. Finally, we show conclusions (Chapter 6) that we obtain from experiments that we describe in the previous section.

1.1 Motivation

Our interest is to improve the Emergency Medical Service System in Mexico. This system has different characteristics than others that have been studied. Especially in Mexico, public and private organizations work together, but they do not communicate what others do to make their decisions.

Also, since the Covid-19 pandemic in Mexico started, fewer hospitals are attending general medical emergencies, and ambulances do queues outside the hospitals waiting for patient attention due there are no available rooms for new patients.

1.2 Problem Description

The Uncoordinated Emergency Medical Service System (UEMSS) Problem is focused on improving an EMS system that involves more than one service provider. We have factors that affect service provider decisions for this system that we discuss below. We called patients at people who need some medical attention due to an accident

or a medical condition that affects the person's health. There are some demand points where someone made an emergency call because the patient needs medical attention. Also, some potential sites are spaces in the zones attending to the EMS system where ambulances can be located in the system.

An uncoordinated EMS system has more than one service provider involved in the emergency number system to attend emergency calls. These service providers have a certain number of ambulances available for patients' medical attention. The total ambulance available number may vary depending on the day of the week and the period time of the day. The ambulances need some material resources but also human resources. The EMS system has to consider the paramedics attending to the patients. If there are not enough paramedics for the total ambulances, not all the ambulances are available for the system.

There are two different ambulance types: basic life support (BLS), which attend not severity accidents, and advanced life support, which attend severity accidents. 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 served. Also, hospitals know where patients will move after their attention. Paramedics need to know if a patient has health insurance to decide where a patient will be transferred.

For our UEMSS Problem, we consider scenarios with information about accidents at the demand point. Each element of one scenario has information about how many BLS and ALS need each demand point, and the sum of ambulances needed can be one, two, or three ambulances for point.

1.3 HYPOTHESIS

This investigation hypothesizes that we can model the Uncoordinated Emergency Medical Service System Problem as a stochastic programming model with resources based on different scenarios. These scenarios consider accident types in each demand point; many of them can help know what to do when a situation occurs in the 9-1-1 system. The model includes characteristics of each service provider, and the decisions are taken coordinate considering these characteristics. The ambulance location and dispatching in the system are optimized.

1.4   OBJECTIVES

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

- Various types of accidents and variations on maximal response times depending on accident types;

- Different ambulances types, which are ambulances for basic life support (BLS) and ambulances for advanced life support (ALS);

- Variation in demand points depending on the day of the week and the hour of the day, which can be considered making different scenarios;

- And queues at hospitals for patients' attention that increase the service time for an emergency call.

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

BACKGROUND

When scientists talk about EMS systems, many terms explain the problem. Two of these terms are demand points and potential sites. Demand points are sites where an emergency call is usually done. Commonly, there is a different demand for each point depending on the number of calls made within a period. Potential sites are places where a vehicle (ambulance) could be located if necessary to cover some demand points either statically or dynamically.

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

2.1 STATIC MODELS

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

There are two early models for statics problems: Location Set Covering Model (LSCM) and Maximal Covering Location Problem (MCLP), which are problems focused on covering the maximal demand points in the entire zone.

2.1.1 DETERMINISTIC MODELS

Later, other models were created to solve static problems because emergency calls sometimes need to be attended for different vehicle types. Most of them are covered

once, like the Backup Coverage Problem (BACOP) or the Double Standard Model (DSM), which use two different radii of coverage [14]; these types of models are called deterministic static models. Some deterministic models are the tandem equipment allocation model (TEAM) or the facility-location equipment-emplacement technique (FLEET), which consider two types of vehicles (one for basic life support and another for advanced life support), or the fact that sometimes more than one ambulance has to be located on a potential site to maximize that a demand point is covered twice.

In the thousands was introduced by Berman et al. [5] a decay function to classify coverage as full, none, and partial coverage in a generalized MCLP model. They added a weighted demand for each node covered, considering the distance between facilities and demand points. The objective aims to maximize the total demand weight covered by all facilities when a determined number of facilities are located.

A year later, Karasakal and Karasakal [13] introduced partial coverage to the MCLP problem. This problem aims to maximize coverage level for all demand points deciding where to locate a certain number of facilities within the available potential sites. The model was based on a p-median formulation and classified coverage into three levels: totally covered, partially covered, and not covered. They defined a decay function monotone decreasing according to the distance between the facility and demand point for partial coverage. The distance between a facility and a demand point has to be less or equal to the maximum full coverage distance established to consider total coverage. Demand points are considered not covered for a facility if the distance between it and the demand point is greater or equal to a maximum partial coverage distance. To solve large-size problems, they used a Lagrangian relaxation.

A decade after, Wang et al. [19] used an extension of the MCLP Problem to maximize coverage for fire emergencies establishing a travel cost between potential sites and demand points. This extension considers a partial distance and quantity coverage for multi-type vehicles to locate and dispatch them. Partial distance is calculated with a decay function, which decreases according to the vehicle response time increase. Quantity coverage determines if an emergency is completely served or not, comparing the number of vehicles dispatched with the necessary quantity. For this problem, they have to consider demand priority to know where vehicles must be located and the patient's classification to decide how to dispatch them.

Research that uses a deterministic static model was done for Dibene et al. [10]; these authors use an extension of DSM, which is called the Robust Double Standard Model. They added to original DSM demand scenarios. These scenarios divide weeks into workdays and weekends, divided into four periods: night, morning, afternoon, and evening. They added eight scenarios applied to optimize the Red Cross Tijuana,

Mexico system, increasing the coverage of demand points to more than 95% locating ambulances on different points of the city that are not the original bases.

2.1.2 Probabilistic models

In the eighties, some researchers thought about problems involved with probabilities. One of these probabilities involved in EMS systems is the probability of an ambulance being busy responding to an emergency call. This probability is called the *busy fraction*. The maximum expected covering location problem (MEXCLP) uses this probability. An extension of this model is the TIMEXCLP which considers travel speed variations during the day. Another extension is the adjusted MEXCLP model (AMEXCLP), which considers different busy probabilities for each potential site to locate ambulances. All these models can use the hypercube queueing model to calculate the busy fraction[11].

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

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

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

Laura Albert and Maria Mayorga researched the EMS systems of Hanover, Virginia. All these investigations about Hanover, Virginia, were applied to this county to obtain practical solutions, but all models can be used to any other EMS system changing data inputs. The first research is focused on considering a new approach to calculate the response time threshold (RTT), a class of EMS performance measures [15]. The approach uses the patient survival rate considering that patients

have a cardiac arrest and random response times that depend on the distance between demand points and potential sites instead of patient outcomes, which is most used.

Then, they use these measures on a hypercube model to evaluate different RTTs needed to input a model that considers fire stations and rescue stations to be potential sites where ambulances could be located distributed on Hanover's rural and urban areas. This model optimizes the location of ambulances on potential sites to maximize patient survival.

Later, Albert and Mayorga et al. used the performance measures as an input of the performance measure dispatching problem. According to survival patient rate, they used a Markov decision process that identifies the best and most robust RTT to maximize the covering level, prioritizing patient location. The research concludes that the optimal survival rate is obtained when the system has an eight minutes RTT [16].

However, this time for RTT does not apply to Hanover because of the number of the ambulance that they have, so they started a pilot program called quick-response vehicle to have more vehicles for patient attention obtaining a nine minutes RTT, these new vehicles are as ALS vehicles without transporting patients to the hospital, only attending patients at the scene and BLS ambulances transport patients if is necessary [17]. The idea of including these quick-response vehicles is to minimize the need to use ALS ambulances.

When talking about optimizing EMS systems, one can also speak about dispatching. Bandara et al. [3] 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-Diaz et al. [18] involves location and dispatching decisions for EMS vehicles in the same mathematical model with two focuses, minimizing the mean response time that takes since an emergency call is received and maximizing the expected coverage demand, using a continuous-time Markov process to balancing flow equations needed to control the busy fraction for each ambulance. Balancing these equations takes exponential time, and authors consider a genetic algorithm to obtain some solutions and combine them to create new solutions to reduce the computational time. This genetic algorithm was applied to Hanover, Virginia, and when they have mid-size problems, the nearest dispatch rule is the best solution. It can vary depending on the zone where it is applied.

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

2.1.3   Stochastic models

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

In 2017, Boujemaa et al. [8] proposed a two-stage stochastic model with recourse. The model's first stage determines where to open ambulance stations with a fixed cost for opening them. For the second stage, allocation is determined considering the expected traveling cost from ambulance stations to demand points. A demand point is considered covered if an ambulance station is within a threshold value. And some important factors that they included are two different demand types: life-threatening calls and non-life-threatening calls; two ambulance types: ALS and BLS; and scenarios structured by two data for each demand point: number of life-threatening calls and number of non-life-threatening calls, respectively. This problem minimizes the ambulance location-allocation cost and is solved by a Sample Average Approximation (SAA) algorithm that allows computing lower and upper bounds for problem solutions and providing the corresponding optimality gaps.

Later, Bertsimas and Ng [6] implemented stochastic and robust formulation for ambulance deployment and dispatch for a problem constructed as a graph. These formulations were compared with MEXCLP and MALP problems and aimed to minimize the fraction of late-arrivals without requiring ambulances to be repositioned, sending to demand points the closest available ambulance, and maintaining a call at a queue if there are no ambulances available at the system. The demand has the problem's uncertainty, which was constructed by four demand types: single for each demand point, local for the demand point and the nearest points, regional for a region of the entire zone, and global for the whole area. They determined a deterministic equivalent model to solve the stochastic formulation, and for the robust formulation, they did a column and constraint algorithm.

Recently, Yoon et al. [20] studied a two-stage stochastic problem for locating and dispatching two types of emergency vehicles: ALS and BLS. The first stage

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

2.2   DYNAMIC MODELS

A dynamic model on EMS systems is used to locate and relocate ambulances from their base to another if necessary for real-time. These models are called dynamic because the system is changing over time. Usually, the *real time* that is used in an EMS system is when an emergency call is received.

One of these models is the Dynamic Double Standard Model at time t, an extension of DSM [4]. Some differences for this model are that every time t, ambulances are moved to another base to cover once demand points at a minimum probability α. There is a cost to move the ambulances from its base to another potential site. This cost is a penalty in the objective function.

Another model of these models is the dynamically available coverage location (DACL). For this model, the system changes when the demand of demand points changes during the day and uses the hypercube theory to calculate the busy probability for potential sites every time.

2.3   CONTRIBUTION

Our opportunity area for this thesis is to analyze a problem for maximizing coverage considering null, partial and full coverage where more than one service provider exists in the EMS system. We use different levels to evaluate a demand point as partially covered. It depends on the distance between the potential site and the demand

point, calculated with a decay function. Also, we calculate how many ambulances of each type were sent to service an accident.

We propose a two-stage stochastic problem with recourse where the first stage is for locating ambulances at potential sites. The second stage determines how to dispatch them to demand points when an accident occurs.

ALS and BLS vehicles are considered for this problem. Any ambulance type can serve accidents that need BLS ambulances, but accidents that need ALS ambulances can be only served by ALS ambulance type. For our scenarios, we have two data: the first is for the number of BLS vehicles needed at the demand point, and the second is the quantity of ALS vehicles required for the same demand point.

In the literature, we do not find a system where exists more than one service provider and considers all factors that we assume. However, it is essential to study it because public and private service providers are involved in Mexico. Sometimes, there is a conflict with them when dispatching is done for an accident. So we want to consider both of them to obtain a coordinated EMS system.

THE UNCOORDINATED EMS SYSTEM PROBLEM

For a preliminar model, we classified if an accident needs one, two, or three ambulances for patients' attention. And we consider the response time between every potential site to every demand point.

The following is a preliminar model for UEMSS Problem. We will describe the sets, parameters, and variables we used and the assumptions we took to model this problem.

SOLUTION DESCRIPTION

The methodology we will follow is:

- Do models construction to represent the thesis' problem proposed. We will consider some different versions to decide which is the best to describe the problem exactly. Versions model will include the following factors: different vehicle types, partial coverage, partial distance, more than one service provider, and hospitals capacity.

- Evaluate each model as an exact model version for small instances. With these results, we can discuss if the models are well implemented or not according to the problem. We will use Gurobi Optimizer to program models' codes.

- Study stochastic programming in Julia, which is recently in the literature. We will see if Julia can solve our models, or we will study Branch and Benders cut solution used in a recent problem.

- Solve small and large-scale problems and evaluate the EMS system to know if we can improve patients' attention.

4.1 PROPOSED SOLUTION

Our problem is focused on two decisions. The first is where to locate available ambulances within potential sites, and the second is how to dispatch them to accidents in scenarios. For these scenarios we have a sample set $S = \{s_1, s_2, ..., s_n\}$ taken from infinite scenarios possibles. Each scenario determines the number of ambulances type 1 and type 2 needed for each demand point. A scenario can be seen as follow: We will consider a demand point set $I = \{1, 2, 3, 4\}$. A possible scenario for these

demand points will be $s_1 = \{(0,1),(1,2),(3,0),(1,1)\}$, which means that demand point 1 needs zero type 1 ambulances and one type 2 ambulances, demand point 2 needs one type 1 ambulances and two type 2 ambulances, and so on.

Ambulance location is a fixed decision. Once the EMS system locates its ambulances at potential sites, these ambulances attend emergency calls and return to the potential sites where they were located. Considering this location, operators receiving emergency calls will have policies to determine what ambulance (or ambulances) is (are) best to attend an emergency at the demand call point. The mean idea for the EMS system's solution is to provide information about where to locate their ambulances depending on the number of ambulances needed at demand points in all scenarios.

4.2 COMPUTING TOOLS DESCRIPTION

We will use the latest version of the Gurobi Optimizer modeling with the Gurobi Python Interface to solve the exact small-scale problems.

For large-scale problems, we will try the latest version of the Julia Programming Language for stochastic problems, which suggests some model decompositions into a master problem and a set of subproblems to solve stochastic models quickly and effectively [7]. If these decompositions can not solve our problem, we will use a Branch and Benders cut to decompose our model, and we will code in Python.

EXPERIMENTAL WORK

We first study solutions to the preliminar model for the UEMSS Problem. In this preliminar model, we obtain results about where to locate available ambulances of each type of ALS and BLS and how the dispatches were in each scenario, knowing how was each coverage. We do not include different service providers in this model yet.

We consider random instances containing information about the total available ambulances for ALS and BLS types, the distance between each potential site and demand point, the possible scenarios, and the value of response time's decay function. For preliminar results, which have information about how the model is solved, we consider 50, 200, and 350 demand point sizes, 20, 40, and 55 potential sites' sizes, and 15, 50, and 75 scenarios' sizes.

The numerical experiments were carried out on a dual Intel Core i7-7500U processor with 2.7 GHz and 12 GB RAM. The models were implemented in the Scientific Python Development Environment Spyder 4.1.5 and linked with Gurobi Optimizer 9.1.1. Each instance had a time limit equal to 60 seconds to be solved.

We will show an instance's solution when 50 demand points, 20 potential sites, and 15 scenarios were considered to present a solution. The total number of ALS and BLS ambulances for these experiments was 11 and 20, respectively. Potential sites where ambulances were located are shown in Table 5.1; for each potential site, we know how many BLS and ALS ambulances are in there.

To know how the dispatches were done, we have an example in Table 5.2 where is shown one of the 15 instance's scenarios. We only have demand points where an accident exists. For each demand point, we have information about the number of ambulances needed for each type, and then we have how many ambulances were sent for each. We define a demand point as partially covered when not all

Potential site	BLS ambulance	ALS ambulance
1	4	0
2	2	0
3	1	1
4	0	0
5	1	1
6	3	2
7	0	0
8	0	0
9	0	0
10	0	0
11	1	1
12	2	1
13	0	0
14	1	1
15	2	1
16	0	0
17	2	2
18	0	0
19	0	0
20	1	1

Table 5.1: Ambulance location for instance of 50 demand points, 20 potential sites, and 15 scenarios.

Demand point	BLS ambulance needed	ALS ambulance needed	BLS ambulance sent	ALS ambulance sent	Coverage type
1	0	2	0	1	Partial
2	0	1	0	0	Null
4	3	0	1	0	Partial
11	2	0	1	0	Partial
19	1	0	0	0	Null
22	1	0	0	0	Null
24	0	1	0	1	Full
27	0	1	0	0	Null
29	0	1	0	1	Full
33	0	1	0	0	Null
35	2	0	1	0	Partial
37	1	0	0	0	Null
39	2	0	1	0	Partial
40	0	1	0	1	Full
42	0	1	0	0	Null
45	2	1	1	0	Partial
49	1	0	1	0	Full
50	0	1	0	1	Full

Table 5.2: Ambulance dispatches for instance of 50 demand points, 20 potential sites, and 15 scenarios.

the required ambulances were dispatched at the demand point. If all necessary ambulances were sent, we could determine that demand point as full coverage. And finally, if no ambulances were sent to the demand point, it is called null coverage. This information is known for all scenarios, and we calculate the maximum expected coverage as an average scenario objective value.

We know the time spent to solve the instances, the best objective value, the best-bound value, and the gap obtained for all experiments. We show the results of the instances described before in Table 5.3, which are an example of how a solution is proportioned. In this table, we can notice that gaps are between zero and one except for two instances where gaps are more significant than 100% because instances are big enough to can not be solved correctly within 60 seconds.

With these experiments, we are sure our model has to be solved by a model decomposition, especially when dealing with large-scale problems, which is our next step.

Instance name	I size	L size	S size	Computer time	Best objective	Best bound	Gap %
Instance_50_20_15	50	20	15	1.013365	39.68222	39.68222	3.58E-14
Instance_50_20_50	50	20	50	5.24326	40.5165	40.5165	3.51E-14
Instance_50_20_75	50	20	75	9.453496	40.31611	40.31611	2.11E-13
Instance_50_40_15	50	40	15	3.710979	42.98478	42.98478	3.31E-14
Instance_50_40_50	50	40	50	12.17611	42.836	42.836	0
Instance_50_40_75	50	40	75	18.94751	42.90256	42.90256	7.45E-13
Instance_50_55_15	50	55	15	4.99372	46.19278	46.19278	7.69E-14
Instance_50_55_50	50	55	50	17.27622	45.7125	45.7125	0
Instance_50_55_75	50	55	75	26.82138	45.93282	45.93282	9.28E-13
Instance_200_20_15	200	20	15	12.42387	142.1911	142.1911	0
Instance_200_20_50	200	20	50	70.23259	144.15	144.185	0.02428
Instance_200_20_75	200	20	75	74.83126	145.5244	145.6672	0.098137
Instance_200_40_15	200	40	15	8.413328	143.2731	143.2731	0
Instance_200_40_50	200	40	50	79.12349	145.194	145.269	0.051655
Instance_200_40_75	200	40	75	89.60279	144.5373	144.5894	0.036018
Instance_200_55_15	200	55	15	11.50527	140.2724	140.2724	0
Instance_200_55_50	200	55	50	53.80193	144.9985	144.9985	1.76E-13
Instance_200_55_75	200	55	75	113.5206	114.6637	327.8349	185.91
Instance_350_20_15	350	20	15	7.783095	220.3283	220.3283	0
Instance_350_20_50	350	20	50	27.36259	230.444	230.444	4.93E-14
Instance_350_20_75	350	20	75	43.73432	232.43	232.43	0
Instance_350_40_15	350	40	15	14.60521	225.4009	225.4009	0
Instance_350_40_50	350	40	50	57.79829	229.3985	229.3985	0
Instance_350_40_75	350	40	75	99.36153	230.1897	230.1897	0
Instance_350_55_15	350	55	15	20.93256	226.7583	226.7583	0
Instance_350_55_50	350	55	50	86.65615	232.371	232.371	0
Instance_350_55_75	350	55	75	124.7983	200.6852	575.8865	186.9602

Table 5.3: Experimental results for preliminar UEMSS Problem model.

CHAPTER 6

CONCLUSIONS

For now, we have only conclusions about our preliminar model.

6.1 MAIN CONTRIBUTIONS AND CONCLUSIONS

Our preliminar model experiments show us that large-scale problems can not be solved because the solver has not had sufficient memory to support them. Nevertheless, small instances results generate feasible solutions for ambulance location and dispatching decisions in different scenarios. Partial rate coverage allows sending ambulances even if there are not enough ambulances to cover an accident totally or if an ambulance (or more than one ambulance) is (are) out of the desired response time for patient attention.

These results help us cover more demand points in the system, allowing beginning attention for patients, which can be finished after providing the first aid to them in those demand points where not all ambulances were sent.

6.2 FUTURE WORK

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

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

literature.

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

Bibliography

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

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

[3] 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.

[4] 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.

[5] 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.

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

[7] M. Biel and M. Johansson. Efficient stochastic programming in julia. *INFORMS Journal on Computing*, 2022.

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

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

[10] 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.

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

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

[13] 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.

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

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

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

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

[18] 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.

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

[20] 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.

Resumen autobiográfico

Beatriz Alejandra García Ramos

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

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica

Tesis:

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