The challenge of dispatching the right ambulance

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von der

KIT-Fakultät für Wirtschaftswissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Dipl.-Ing. David Alfredo Olave Rojas

Tag der mündlichen Prüfung: 23. Februar 2023

Referent: Prof. Dr. Stefan Nickel

Korreferent: Dr. Roberto Aringhieri

Karlsruhe, 24.03.2023

A la amada familia,
de la cual provengo

“Phantasie ist wichtiger als Wissen, denn Wissen ist begrenzt”
Albert Einstein

Acknowledgements

First, I would like to thank the invaluable opportunity and support given by Prof. Dr. Stefan Nickel. Our conversations were always a source of inspiration and guidance. I will remember his comprehension, especially in the first stages when everything was completely new for me in Germany.

To Prof. Dr. Eduardo Álvarez-Miranda and Prof. Dr. Francisco Saldanha-da-Gama for giving me critical advice about this fantastic world called science through our fruitful conversations. I hope I am capable to go further through this path into the wild.

To my colleagues in the discrete optimization and logistic group (DOL) of the Institute for Operation Research (IOR) at the Karlsruhe Institute for Technology (KIT) for introducing me to the German culture inside and outside of the Institute. Their support and fellowship were critical to feel Karlsruhe as my Home.

Furthermore, I would like to thank "the one and only" Sven Watzinger due his priceless support and patience, especially in the last stages of this journey. I enjoyed every minute of working together *auf Deutsch o en Español*, and I am grateful for his help.

To the Karlsruhe Service Research Institute (KSRI) of the Karlsruhe Institute of Technology for supporting me with a grant and including me in every social activity of this wonderful group.

To my parents, to my sister and her family, since they always support me in every single crazy idea I have, like this Ph.D.

And to every single person in the world, from this side or the another side of the Atlantic Ocean, who had supported, inspired and pushed me like sport fans through these years.

Contents

Zusammenfassung

Das Gesundheitswesen ist eine der wichtigsten Disziplinen, um die Lebensqualität der Menschen zu sichern. In diesem Zusammenhang wurde Rettungsdienst mit dem Hauptzweck, Leben zu retten, geschaffen. Die Verwirklichung dieses Ziels erfordert eine korrekte Reaktion auf lebensbedrohliche Zustände, die sich aus verschiedenen unvorhersehbaren Situationen wie Unfällen, Naturkatastrophen, Terrorismus und Pandemien ergeben. In vielen dieser Fälle geht es um das Sammeln von Informationen über den Notfall und die Koordination verschiedener Ressourcen um den Notfall behandeln zu können. Zum Beispiel sind Ressourcen wie Ambulanzteams, Notärzte, Krankenhäuser und Notrufzentralen innerhalb des Medizinischen Rettungsdienst-Systems durch dynamische Kommunikation miteinander verbunden Diese Ressourcenkommunikation und -koordination sollte so gut wie möglich sein, um die Überlebenswahrscheinlichkeit eines jeden Patienten zu erhöhen. Diese Situationen fördern unterschiedliche Probleme, die analytische, logistische und betriebswirtschaftliche Merkmale umfassen.

Einige dieser Herausforderungen können z. B. durch eine geeignete Einsatzstrategie für Krankenwagen bewältigt werden. Die Auswahl des richtigen Rettungswagens für einen bestimmten Notfall bringt jedoch nicht nur analytische und modelltechnische Herausforderungen für die korrekte Untersuchung des Systems mit sich, sondern erfordert auch genaue und realitätsnahe Entscheidungshilfen, die eine angemessene Modellierung des Systems und der Ressourcen sowie die Koordinierung der entsprechenden medizinischen Notfallsysteme gewährleisten.

Obwohl frühere Arbeiten aus dem Bereich des Operations Research Anstrengungen unternommen haben, um eine Verbesserungen bei der Suche nach der besten Einsatzstrategie zu erzielen, basiert die überwiegende Mehrheit von ihnen auf der Integration von Standortzuweisung und Einsatzplanung. Außerdem, hat keine dieser Arbeiten alle

Akteure in integrierter Weise berücksichtigt und sowohl die Integration als auch die Interaktion zwischen ihnen in Betracht gezogen. Dieser Idee folgend wird in dieser Dissertation ein methodischer Rahmen untersucht und vorgestellt, mit dem Ziel, die Analyse und Reaktion der Rettungsdienste eines bestimmten Gebiets von einer operativen Ebene aus zu verbessern.

Vor diesem Hintergrund wird ein hybrides Simulationsmodell unter Verwendung eines Machine-Learning-Ansatzes für den präklinischen Rettungsdienstprozess vorgestellt. Dieses Modell wird mithilfe von Analytics-Methoden erstellt und validiert, um eine sorgfältig ausgewählte Datenbank zu erstellen, die zwei wichtige Regionen des deutschen Rettungsdienstsystems repräsentiert. Dieses Simulationsmodell enthält wichtige Funktionen, die im Notfallprozess dargestellt werden, wie z. B. die Möglichkeit, den Krankenwagen zu entsenden, wenn dieser zur Wache zurückkehrt, die erneute Disposition im Falle eines schwereren Notfalls, der in das System eingeht, und die Synchronisierung von Notärzten mit Krankenwagen-Teams, usw.

Des weiteren wird eine Methodik zur Bewertung der Einsatzstrategie vorgestellt, die auf dem hybriden Simulationsmodell und der Online-Optimierung basiert. Abschließend wird eine auf Online-Optimierung basierende Studie für das Problem der Einsatzplanung vorgestellt.

Aus den Ergebnissen lässt sich schließen, dass gemischte Strategieansätze in realen Szenarien robuster, zuverlässiger und vorhersehbarer sind als einzelne Dispatch-Strategien. Darüber hinaus wirkt sich die Anzahl der Krankenwagen im System direkt auf die Untersuchung aus, um zu bestimmen, welche Dispatch-Strategie besser oder am Besten abschneidet. Dies ist auch mit der Ressourcenauslastung verbunden, die diesen Indikator als die Stressmenge im System versteht. Obwohl diese Dissertation ein Versuch in die richtige Richtung ist, um neue Möglichkeiten für die Entwicklung eines leistungsfähigeren und dynamischeren Rettungsdienstes zu entdecken, ist es schließlich auch klar, dass es notwendig und dringend ist, mit dieser Doktorarbeit weiter zu gehen, einschließlich neuer Ansätze, wie z. B. die Integration verschiedener Methoden (Simulations-Optimierungsmodelle + künstlichen Intelligenz), um der Komplexität und Nichtlinearität von notfallmedizinischen Systemen zu begegnen.

Stichwörter Simulation, Machine Learning, Rettungsdienst, Online Optimierung

Abstract

Healthcare is one of the most important disciplines to ensure life quality for human beings. In this context, Emergency Medical Systems (EMS) have been created with the principal purpose of saving lives. Addressing this objective implies a correct response to life-threatening conditions resulting from several unpredictable situations such as accidents, natural disasters, terrorism, and pandemics. Many of these cases involve collecting critical information about the emergency and coordinating several resources to attend to it. For instance, resources such as ambulance teams, emergency doctors, hospitals, and call centers are interrelated within the EMS system through dynamic communication. This resource communication and coordination should be the best possible to maximize the survival probability of each patient. These scenarios propose diverse issues from the analytical, logistical, and managerial points of view.

Some of these challenges can be faced, for example, through a proper ambulance dispatch strategy. However, selecting the right ambulance for a specific emergency not only involves analytical and modeling challenges for the correct study of the system but also calls for accurate and real-world-based decision-making tools that ensure appropriate system and resources modeling and coordination of the corresponding emergency medical systems.

Although previous works have devoted efforts from operations research in order to provide improvements towards finding the best dispatch strategy, the vast majority of them are based on the integration of location-allocation and dispatch approaches. Furthermore, none of them has included all the actors in an integrated manner, considering the integration and interaction between them. Following this idea, in this thesis, a methodological framework is studied and presented to enhance the analysis and response of the emergency services of a particular territory from an operational level.

In particular, a hybrid simulation model using a machine learning approach is presented for the pre-hospital emergency process. This model is built and validated using analytics tools to produce a carefully-curated database representing two important regions of the German emergency medical system. This simulation model includes important features presented in the emergency process, such as the ability to dispatch the ambulance when returning to the base, re-dispatching in case of a more severe emergency entering the system, and synchronizing emergency doctors with ambulance teams, among others. After that, a dispatch strategy evaluation methodology is presented based on the hybrid simulation model and Online optimization. Finally, an online optimization-based study is presented for the ambulance dispatch problem.

The results can conclude that mixed strategy approaches are more robust, reliable, and predictable than single dispatch strategies in real-world scenarios. Furthermore, the number of ambulances in the system directly affects the analysis to define which dispatch strategy performs better. This is also associated with the resource utilization understanding this indicator as the amount of stress presented in the system.

Finally, although this thesis is an effort in the right direction in order to discover new opportunities for developing a more capable and dynamic emergency medical service, it is also clear that there exists a need, an urgency, to go further with this doctoral thesis, including new approaches, such as combinations of different methods or Artificial Intelligence Simulation-Optimization models in order to face the complexity and non-linearity of emergency medical systems.

Keywords Ambulance dispatch, Simulation, Machine Learning, Emergency Medical Service, Online Optimization.

Chapter 1

Introduction

In this chapter we present an introduction to the principal challenges related to Emergency Medical Systems which motivate this thesis. Furthermore, the research goals and the structure of the thesis are also presented.

1.1 Motivation

According to World Health Organization [2020] (WHO), cardiovascular disease (CVD) is the first cause of death worldwide, taking an estimated 17.9 million lives each year. Four of five CVD deaths are related to heart attacks and strokes. In Europe, CVD causes more than every second disease, according to WHO Europe [World Health Organization regional office for Europe, 2020]. Meanwhile, in the United States, Coronary Heart Diseases (CHD) are the leading cause of CVD deaths, followed by strokes, according to Virani et al. [2020]. The vast majority of these cases are time-critical, i.e., survival expectations depend directly on how fast patients receive the proper treatment. In these cases, Emergency Medical Services (EMS) are the first to respond and take care of the patients, many of them in life-threatening conditions. However, EMS also have to attend to calls from different situations such as accidents, natural disasters, terrorism, pandemics, and patient transport. These present complex scenarios from a logistical point of view. The dispatch and the use of resources should be resolved efficiently to achieve a satisfactory response to each situation.

Facing these scenarios involves resolving logistic challenges, such as ambulance location, allocation, workforce scheduling (ambulances and call centers), and ambulance dispatch problems. Within the EMS system, resources such as ambulance teams, emergency doctors, hospitals, and call centers are interrelated through dynamic communication. For instance, in a heart attack situation, the call-taker attends the emergency call. After collecting all relevant information, an emergency resource is dispatched according to the available options by a dedicated dispatcher or the same call-taker. Subsequently, the emergency resource, commonly an ambulance team and an emergency doctor, drives to the emergency place. Meanwhile, the call-taker assists the cardiopulmonary resuscitation by phone until the emergency resources arrive at the patient. This event sequence takes place just in minutes, forcing the system to a coordinated response in just seconds.

In this context, resource coordination should be the best possible to maximize the survival probability of each patient. Proposing solutions for the ambulance dispatch problem could be challenging since sometimes the information about the emergency is incomplete, and it is hard to predict when and where the next emergency will occur.

Two essential works in this field, such as Aringhieri et al. [2017a] and Singer and Donoso [2008] , present simulation techniques as one the best in terms of reality process representation for complex systems. Such an approach is one of the best to evaluate implementable solutions. In fact, Aringhieri et al. [2017a] point out that establishing some assumptions for better computational tractability could affect the solution quality. According to the authors, these assumptions are required to use classical operational research approaches such as stochastic and mixed-integer programming or queue theory.

Furthermore, Aboueljinane et al. [2013] indicate that integrated approaches such as optimization and simulation lead us to better decisions and better analysis since simulation would assess the impacts of the solutions proposed by the optimization models in a more realistic context.

In this context, online optimization approaches are perfect since the ambulance dispatch problem is defined at the operational level. Moreover, ambulance dispatching is highly complex due to the number of variables and players, and the findings could be easy to implement in the real world. Online optimization approaches are pretty close to reality and can address the difficulties presented in the real world.

Several previous studies in the area of ambulance dispatch compare many strategies to find the best in terms of response time, survivability or coverage. The most common policies are the nearest idle ambulance, some strategy based on a covering indicator, or some policy based on maximizing survivability [see Aboueljinane et al., 2014, Bandara et al., 2014, Bélanger et al., 2020, as examples in the literature]. Furthermore, many studies test a combination of location-allocation strategy with a dispatch strategy. This approach is efficient since there are synergies between both strategies working together, as pointed out by Aringhieri et al. [2018], Bélanger et al. [2020], among others.

However, these approaches are challenging to implement without incurring negotiations with paramedics or ambulance teams. They claim these strategies force them to be in constant movement, increasing travel times and workload but decreasing rest time. Moreover, an allocation-dispatch strategy does not allow us to know how effective the dispatch strategy is.

Additionally, the works including dispatch policies tested using simulation, do not represent EMS in a detailed manner since they do not include the interaction between resources, ambulance dispatching when the ambulance is returning to the base, realistic travel speeds according to the moment of the day or distances, decision protocols of ambulance teams, re-dispatching among others (see section 4.1 and 6.1 for more details).

This scenario proposes the following research questions(RQ):

RQ 1. How is the interaction presented between the emergency resources in an EMS process?

RQ 2. How could affect the territory orography and external parameters the ambulance dispatch performance?

RQ 3. What are the relations between capacity, demand, and dispatch strategy?

RQ 4. Which mathematical characteristics shall be present in a model which includes the requirements from practitioners, patients, and the EMS?.

Finally, these questions lead the research effort, where the corresponding results we present in this doctoral thesis. Furthermore, these RQs inspire the goals of this work.

1.2 Research goals

1.2.1 General objective

This thesis aims to develop and present a model and an evaluation methodology framework for assessment and testing ambulance dispatch strategies to aid decision-makers in real-life situations.

1.2.2 Specific objectives

- To understand and characterize the existing emergency medical service process in a real-world context.

- To characterize the elements presented in the emergency medical service process by representing functional and topological properties and characteristics, resource availability, the interrelation between these resources, and the emergency demand.

- To propose simulation models that allow us to understand the relationship between dispatch strategies and a particular territory.

- To establish the applicability of the methodological framework for defining the best ambulance dispatch policy for robust system response.

1.3 Overview of the thesis

The presented doctoral thesis is organized as the following:

Chapter 2: We discuss Emergency Medical Service (EMS) systems in this chapter, presenting the main philosophies and focusing on the German EMS.

Chapter 3: It includes the methodological background presented in this thesis, which includes simulation, online optimization, and machine learning approaches.

Chapter 4: We present a detailed and realistic simulation model for the pre-hospital emergency medical process based on the German EMS system.

Chapter 5: It presents an evaluation methodology for ambulance dispatch strategies. We also present theoretical analysis to understand how some spatial and resource variations could affect dispatch strategy performance.

Chapter 6: We present our study to answer which ambulance dispatch strategy is the best. This study includes realistic characteristics using the methodology presented in Chapter 5.

Chapter 7: Finally, we present the conclusions, further research, and some discussions about this thesis.

Chapter 2

The Emergency Medical Service system

This chapter presents a definition of Emergency Medical Service (EMS) Systems through a historical overview. Furthermore, we present the main EMS models and philosophies from the point of view of the resources. Finally, we describe the German EMS system's organization and resources. Some parts of this chapter are published in Olave-Rojas and Nickel [2021].

2.1 Towards a modern Emergency Medical Service

Emergency Medical Services (EMSs) are one of the three emergency services presented the modern society, besides Police and Fire and Rescue Services. Their principal task is to provide Emergency Medicine.

Nevertheless, the question remains about what EMSs are. The European Society for Emergency Medicine [2021] defines EMS as the following:

> **Emergency Medical Service** A primary specialty established using the knowledge and skills required for the prevention, diagnosis, and management of urgent and emergency aspects of illness and injury, affecting patients of all age groups with a full spectrum of undifferentiated physical and behavioral disorders.

According to World Health Organization et al. [2008], A short definition is that EMS *"typically refers to the delivery of medical care at the site of the adverse medical event"* at a certain amount of time.

Historically, Dominique Jean Larrey is considered the father of emergency medicine, as pointed out by Skandalakis et al. [2006]. He distinguished himself during the French wars. After seeing the speed of the flying french artillery due to carriages, he implemented this idea in health care, inventing the *"ambulances volantes"* or flying ambulances in 1792. The flying ambulances aimed for the rapid transport of injured soldiers from the battlefield to a more safe place. However, the word ambulance's first appearance was in the 15th century when the Queen Isabella of Spain introduced the term. She used this word to refer more to the military field hospitals, which were located in tents, than to carriages or wagons that allowed the transport of the wounded to a safe place, according to Ortiz [1998].

Wilford [2008] points out that civilians had to wait until 1832 for the implementation of early ambulances in a city area after the outbreaks of cholera in London.

The United States of America was the pioneer in terms of a hospital-based ambulance service. In 1865, the first service in Cincinnati began to operate, followed by another one in New York (Barkley [1990]).

Through the years has been several improvements and new techniques, such as the development of the first training program for ambulance attendants by the American College of Surgeons on mid 1950s and the demonstration of the efficacy of mouth-to-mouth ventilation by the Dr. Peter Safar, according to Sanders et al. [2012], among others. Nevertheless, in the 1960s started a modernization and professionalization of processes and professionals, especially in the United States, as described by Kouwenhoven et al. [1960]: (i) Cardiopulmonary reanimation (CPR) and defibrillation became as the standard manner to care cardiac arrest outside of hospitals; (ii) the role of the paramedic was created; (iii) it started the specialization of physicians orientating their formation specifically to emergency accidents started; (iv) the German emergency doctor was introduced in the German EMS, among others.

Each country has a specific organization and structure according to its needs in this context. The differences could be present in the response time targets, the payment structure, whether the service is entirely public, private, or a combination of both approaches, among others.

Some authors, such as Al-Shaqsi [2010], recognize two main philosophies in organization and paradigms: The Anglo-American model and the Franco-German model. The Anglo-American philosophy is "Scoop and Run". Meanwhile, the Franco-German philosophy follows the idea of "Stay and Stabilize". As described by Dick [2003], the patient is transported to the treatment in the Anglo-American model, whereas in the Franco-German model, the treatment is carried to the patient. However, most EMS systems in the world are derivations or combinations of these models.

2.2 The Anglo-American model

The Anglo-American model, as pointed out, follows the idea of transporting the patient to the treatment as soon as possible. This shapes the entire process. Commonly, an EMS based on the Anglo-American model is composed of the following resources related to the emergency scene, according to Sanders et al. [2012]:

Dispatchers: Their are telecommunicators with the aim to stablish the primary contact with the public, and to direct the proper agencies to the scene. These agencies are the typical resources presented in a EMS such as ground and air ambulances, fire departments, utility services, among others.

Emergency Care Assistants (ECA): Their main task is to support the emergency Medical Technicians or the Paramedics. They are also capable of driving and giving first aid.

Emergency medical Technicians (EMT): They can perform monitoring procedures such as blood pressure monitoring or intravenous and intraosseous access. Could work without supervision in some low-risk emergencies.

Paramedics: They are the most competent and skilled professional on the emergency scene. Hence, they can perform advanced life support procedures such as intubation, defibrillation, and drug administration under the direction of a physician.

In the Anglo-American Model, Physicians or emergency doctors are present as consultant figures, giving some feedback and supervision to the paramedics. Usually, they are located in a hospital. However, they can also be on-site since they are also part of the air ambulance team. When an air ambulance is called, critical care medicine is required on the scene, which a doctor provides.

There are two types of ambulances: Basic Life Support (BLS) and Advanced Life Support (ALS) ambulances. BLS ambulances can transport patients to and between hospitals. The crew is composed of a least two persons: an ECA and an EMT. Sometimes they also attend low-risk emergencies. ALS ambulances are prepared for life-threatening emergencies. The crew is composed of at least two persons but, in this case, are a Paramedic and EMT. Usually, an ECA is also part of the team. Sometimes, they also perform transport duties when the patient is at risk.

2.3 The Franco-German model

The Franco-German Model is well-known for stabilizing and treating the patient in place. Some countries following this model use the acronym SAMU for their EMS, which means "Mobile Emergency and Resuscitation Service". With this aim, an EMS based on the Franco-German Model is composed of the following resources on the scene, according to Organization et al. [2008], and Bos et al. [2015]:

Dispatchers: They address similar functions as their colleagues in the Anglo-American model, such as to receive and process calls for EMS assistance, to dispatch and coordinate EMS resources, to relay medical information and the coordination with public safety agencies as required by the emergency.

Basic Emergency Medical Technicians: They are equivalent to the ECAs and perform almost the same tasks as their colleagues in the Anglo-American model.

Advanced Emergency Medical Technicians: Also known as Paramedics in some Latin-American countries, but equivalent to the EMT. Some countries have something between the EMT and Paramedics from the Anglo-American Model.

Physicians: They are specially qualified for applying treatment at the emergency scene. The entire ambulance Team is under the direction of an emergency doctor. They usually arrive at the scene in the same vehicle as the ambulance team. Sometimes they arrive in a separate vehicle driven by an Advanced EMT.

We also recognize BLS and ALS ambulances in this model with differences in capacity and capability when attending an emergency. By definition, ALS should also be able to attend a birth in some Latin-American countries. The crew is equivalent since there are composed by least two persons who have different skills depending on whether they are a team of a BLS or an ALS ambulance.

2.3.1 The German EMS system

In Germany, each of the 16 federal states is organized in several EMS regions, composed of either rural areas, urban areas, or a mix. Each EMS region is controlled and managed by a coordination center attending emergency calls and managing the resources. In some EMS regions, the coordination center is an integrated one, which means that it is in charge of both medical emergency and firefighter resources. This approach yields a better response in emergency rescue cases by coordinating resources, such as traffic accidents and landslides.

The main emergency resources are call-takers, ambulances, emergency doctors, and helicopters. They can be briefly described as follows:

Call-takers: They are located at the coordination center. Their tasks include answering emergency calls, gathering the relevant information about the emergency event, entering corresponding information in the IT system, booking, dispatching, and alarming the emergency team. The last three tasks are assigned to a dedicated dispatcher in some coordination centers.

Ambulances: They are located in hospitals and ambulance stations. Their teams are always composed of two persons: two emergency medical technicians (EMT) when the ambulance is assigned to transport duties or at least one paramedic if the ambulance is required for emergency rescue. These ambulances have different equipment according to what task they perform, whether transport or emergencies. In some regions, the same kind of ambulance performs transport and emergency duties.

Emergency doctors: In Germany, by law, an emergency doctor is required on-site for particular types of treatments such as defibrillation or life-saving drug administration. Doctors are located in hospitals or their private practices. An emergency doctor generally drives to the scene in a separate vehicle and meets the patient and the ambulance team at the emergency location. Emergency doctors can be dispatched either at the same time as the ambulances or afterward an evaluation by the ambulance team.

Helicopters: Commonly, helicopters are used when immediate treatment at a hospital is necessary. They are used for emergency rescues or intensive care transports. Sometimes, accessibility is a decisive factor in dispatching a helicopter.

Finally, one ambulance can serve one patient at a time. Then, in the case of a multiple accident, several ambulances should be dispatched, depending on the number of people involved in the medical emergency.

Figure 2.1 presents a scheme with the main stages in an emergency medical service process, according to Reuter-Oppermann et al. [2017].

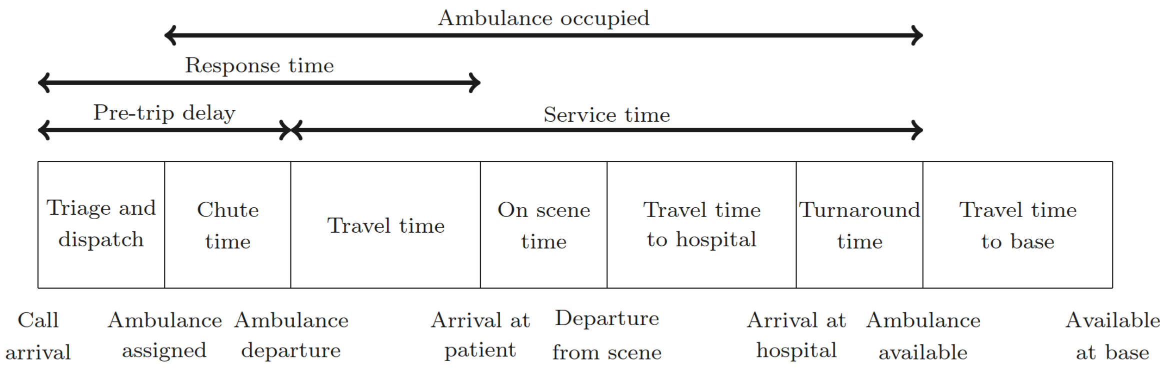


Figure 2.1. Schematic process of a emergency medical service.

Despite the differences, both philosophies pursuit the same goal in life-threatening cases, which is a fast response to saving lives.

Chapter 3

Methodological background

This chapter presents the methodological background used to develop this work. Since the methodology is based on simulation and optimization models, we focus this chapter on these model techniques. We organize this chapter as follows: first, we present some simulation approaches, with a focus on those we use in this thesis; second, we present simulation validation techniques; finally, we present the elements used in this work related to online optimization and the reasons about why is this type of optimization one of the best approaches for the problems addressed by this thesis.

3.1 Simulation

Simulation is one of the high-level tools in system analytics for aiding decision-makers since they allow us to know and analyze a complex process deeply. These models include each stage of the process and the interrelation between them. Each simulation approach has advantages and disadvantages, depending on which kind of decisions we desire to study. Baseline literature in this field is Law et al. [2000] and Banks et al. [2013]. Most of the following descriptions and definitions are based on these books, as well as key literature such as Carter and Price [2017], Ingalls [2011], Loper [2015], Nelson et al. [2013] and White and Ingalls [2018].

In order to understand what simulation is and what it is capable of, let us define the concepts of model and simulation, according to Loper [2015]:

Model is a physical, mathematical or otherwise logical representation of a system, entity, phenomenon or process.

Simulation is a method for implementing a model over time.

Modeling is the process of creating a simplified representation of a real system. This is done through simplification, or removing unimportant details, and abstraction, which is representing the essential features of a system in a different manner. The resulting model should demonstrate the qualities and behaviors of the real-world system that are relevant to the questions that the modeler is trying to answer.

Simulation is the process of running a model over time. This can be done on a computer, which allows for the study of how the modeled system works. By changing variables in the simulation, predictions can be made about the behavior of the system. There are many different types of computer-based simulations. Some of the most common approaches include discrete event, continuous system, agent-based, and system dynamics. A common feature they share is generating or predicting an artificial time history of the system, allowing the observer or experimenter to draw inferences concerning the operating characteristics of the real system that is represented.

While a model is essential to a simulation, other elements, such as data to stimulate model inputs and the target computer architecture, directly affect performance and accuracy and must be included in the definition of a simulation. Considering these various aspects of a simulation leads to the following conceptual definition, according to Tavernini [1996]:

$$Simulation = Model + Data + Method + Implementation + Realization \tag{D.1}$$

Where the components of D.1 are defined as the following:

Model could be a mathematical formulation, a flowchart a logic diagram, a state diagram, or a workflow, among others. It should be also formulated for a specific problem with defined border conditions, and also robust enough for address practical situation related to the phenomenon to be studied.

Data are the model inputs. They are commonly constrained to a number of representative situations that the system addresses. The combination of data and model results in a unique solution independent of the selected method, implementation and realization. When the parameters are modified between runs, then the treatment of them are similar to data. Nevertheless, in some simulations, a number of parameter values are defined as an integral part of the model. Moreover, random variables could be part of the data in order to define some probabilistic parameters of the model.

Method refers to different numerical procedures and algorithms for solve the model's equations. The selection of an appropriate method defines the mathematical accuracy and the computational complexity of the entire simulation. The better is the accuracy, the longer is the computation time.

Implementation corresponds to how the selected methods are implemented. For instance, a non-linear relationship between two continuous variables could be represented as a table function (the method) using discrete values and intervals or steps. Specifying how to adjust the step size and defining the values where the values are valid are part of the implementation.

Realization refers to the considerations and characteristics related to the computing hardware, variable size (i.e., integer vs. float vs. double), coding language and operating system, among others.

3.1.1 Simulation approaches

3.1.1.1 Discrete event simulation

Discrete Event Simulation (DES), as the name describes, models processes focusing on discrete events. First works related to this approach are those which define and introduce the terminology and key features of DES, such as simulation clock by Kiviat [1969], virtual clock and event ordering by Lamport [1978] and time-stamp order by Chandy and Misra [1979].

In general, DES models are composed of stages in the same manner as processes. Then, events trigger stage changes of specific agents, which represent entities in the simulation model. Hence, discrete events define when an agent enters the process, the stage duration, when an agent leave the process, etc. A discrete-time unit defines the minimum amount of time represented in the DES model. Some stages need resources. These are represented also as agents on the simulation model. For instance, in an emergency department, patient arrival is considered as an event, and stages represent each operation in the process on the simulation model. Nurses and Medical Doctors are resources of the system which are also modeled as agents. Every stage in the model takes a certain amount of time, which could be deterministic or stochastic. These characteristics present DES as one of the best approaches for processes where operational level decisions occur or short-term planning times.

Some examples of systems that might be evaluated using DES include any queueing system, such as a bank service counter, manufacturing systems and inventory and supply chain systems. We present an easy example of a DES representation in Figure 3.1, where circles represent entering or leaving the process. Blue boxes denote stages, and sky-blue boxes represent seizing or releasing a resource.

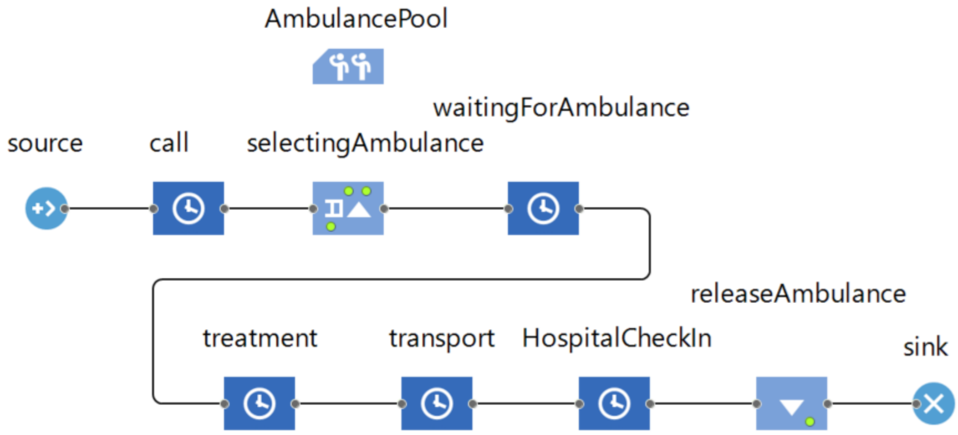


Figure 3.1. Example of a Discrete Event Simulation model for a simple pre-hospital emergency process.

3.1.1.2 Agent-based simulation

Agent-Based Simulation (ABS) focuses on the behavior of individual agents. This approach allows us to model a behavior protocol for each agent, addressing multiple situations and representing more complex behaviors. For instance, the interaction between agents could be addressed in detail. Sending information and instructions as well as synchronizing resources or implementing a shift schedule is also possible.

In the 1990s, some early agent-based models were built using SoftWare and Algorithms for Running on Multicore (SWARM) software, according to Minar et al. [1996], which was developed for artificial life and was focused on investigating emergent behaviors. In the 1980s and 1990s scientists in the field of ecology were using an approach called individual-based modeling, a paradigm very similar to agent-based modeling, according to Grimm and Railsback [2006].

ABS is especially suitable for highly complex processes where several agents interplay. These characteristics present ABS as one of the best approaches for tactical-level decisions or long-term planning times. We present a schematic version of an ABS model for an ambulance in Figure 3.2. There, we represent stages with yellow boxes and transitions with arrows. Specific parameters could define following one branch over the other, as we represent in the bifurcation between *DrivingToBase* and *Transport*.

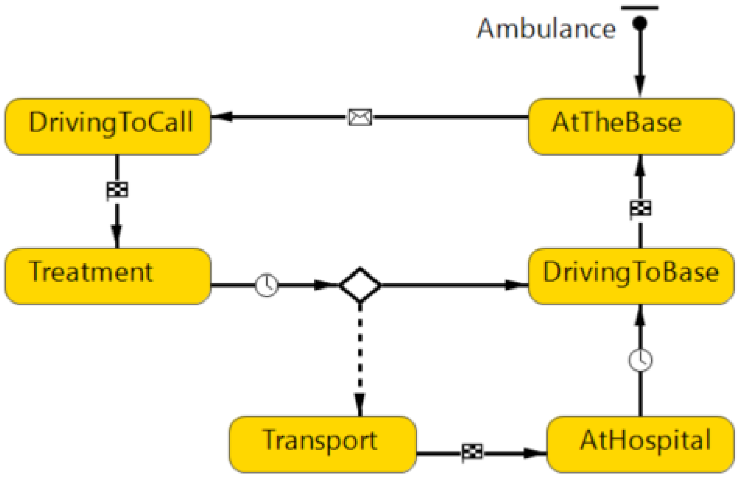


Figure 3.2. Example of a Agent-Based Simulation model for an Ambulance protocol.

3.1.1.3 Dynamic simulation

Dynamic Simulation (DS) or system dynamics (as presented also in literature) was developed by Jay Forrester at MIT Sloan School of Management in the 1950s to help solve management problems, as presented by Forrester [2007], and it is focused on flows. These flows could be information, products, patients, materials, among others. These elements call resources and are collected in Stocks. The incoming and outcoming flows from and to each Stock is determined by the flow rates. These flow rates could be modified by information links which affect differential equations.

Dynamic Simulation represents a top-down approach with a high level of abstraction. Hence, DS is one of the best approaches for long-therm decisions. Examples where DS could be implemented for what-if analysis are military, environmental designs, population growth analysis, engineering design, weather and crop production, among others. Figure 3.3 presents a simple version of a DS model applied in a emergency process context. In the example, *population* denotes a stock, *decreaseFactorCalls* illustrates a flow rate, and arrows represent information links.

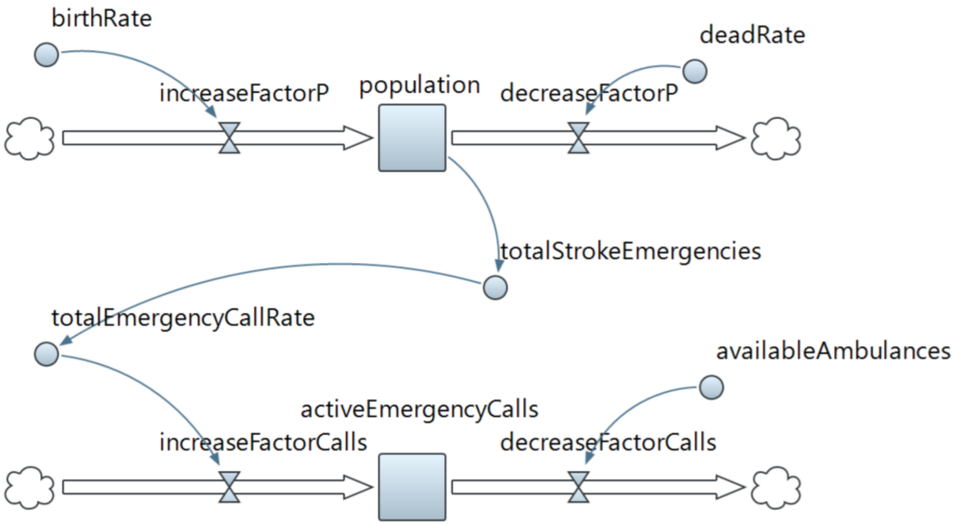


Figure 3.3. Example of a Dynamic Simulation model for a top-level emergency model.

3.1.2 Validation techniques

Validation and verification of simulation models are crucial since a simulation model aims to allow us to understand and experiment with situations and find solutions for real-world problems without the costs of doing it in the real system. Proper validation ensures that the model is a valid representation of reality. It also defines under which parameters and characteristics the model is representative and, therefore, under which scenarios the conclusions obtained from the model are implementable.

3.1.2.1 Without historical data

Animation: In animation, we can see the agents and event movements graphically in the layout of our simulation model. This possibility allows the observation and the comparison of the agents' behavior to find inconsistencies between the model and the real world. Some works pointing this out are Churilov et al. [2013], Jánošíková et al. [2019], Kleijnen [1995].

Extreme condition test: Testing the simulation model in extreme conditions allows us to see if the simulation model does what is expected in those test conditions. For example, if there are no emergencies in a particular region, none of the ambulances should respond. Hence, we should see zero travels in our simulation. Some works including this are Churilov et al. [2013] and Silva and Pinto [2010].

Face validity: When there is no access to real data, the opinion of an expert group could be the best approach. Experts have enough experience to identify some errors in the results, in the process or in the behavior of the agents. They also can be helpful to define critical parameters such as travel speed, treatment time, among others. Some works such as Aboueljinane et al. [2013], Churilov et al. [2013], Jánošíková et al. [2019] and McCormack and Coates [2015] pointed out this importance.

Trace: In complex simulation models, tracing an event or an agent in detail could be the right approach. This is based on saving the time in every step of the process where an agent is taking part for post-analysis. For instance, if a nurse has to vaccine a patient, we save the timestamps for the entire vaccination process. Then, we compare the timestamps with the expected ones collected from the actual process. Works addressing this technique are Granberg and Nguyen [2018], Kergosien et al. [2015] and Kleijnen [1995].

3.1.2.2 With historical data

Event validity: It compares the event time distribution from real data and the simulation model. This could be implemented for a specific step of the process or the entire system. Some references for this validation are Aboueljinane et al. [2013], and Churilov et al. [2013].

Graphic comparison: It takes into account the visual comparison of graphics. These graphics are statistical parameters or distributions from the simulation model and the historical data. Works describing this are Carson [2002], Granberg and Nguyen [2018], and Kleijnen [1995].

Inspection approach: It consists of the inspection and comparison of key performance indicators (KPIs), such as mean values or variances of specific output parameters. The comparison is performed between the KPIs from the simulation model and their real-world counterparts. Some works that point this out are are Aboueljinane et al. [2013], Jánošíková et al. [2019], Mendez-Giraldo et al. [2018], and Pinto et al. [2015].

Hypothesis test: It is an alternative to the inspection approach and consists of a statistical test, refusing the null hypothesis at a specified confidence level. In this case, the null hypothesis means no difference between the output parameters from both the real system and the simulation model. Some references describing this technique are Kleijnen [1995], and Churilov et al. [2013].

Confidence intervals: It is a very objective validation system based on statistical theory. Nevertheless, it takes several independent datasets from the existing system and simulation model for proper validation, which could be a drawback. Law et al. [2000] and Sargent [2013] mention this technique in the literature.

In this thesis, we apply ABS and DES to develop a Hybrid Simulation model. We use this approach since the problem faced in this thesis includes emergencies, which we model as events, and ambulances and medical doctors, which we model as agents (chapter 4 details why this approach is the best in our context). We use animation, face validity, trace, event validity, and graphic comparison as validation and verification techniques.

Several of these techniques are also presented in machine learning algorithms since they are used to model and predict the behavior of a particular variable, including sometimes an extensive set of independent parameters.

3.2 Machine learning

In this thesis, we also use machine learning algorithms. These algorithms are suitable for recognizing patterns, characteristics, and tendencies presented in a database to replicate or predict them. The concept of an algorithm with a learning capacity is not new. The first work speaking about this idea is Samuel [1959]. However, recently this topic has been getting more attention since the actual computer capacity can manage instances large enough as presented in real-world problems. Some literature related to this topic is Amat Rodrigo [2020], IBM Cloud Education [2020], James et al. [2013], Kubat [2017], Mitchell et al. [1990], and Zhou [2021]. In general, the literature identifies three main machine learning algorithm groups:

Supervised learning: It consists of a mathematical model that has access to both input and output data. The data is, in this case, a kind of "teacher" since it shows the correct answer for a particular input of parameters. A common approach is to split the dataset into two groups. The first group is designated for training purposes, changing the parameters of the mathematical model to meet the predicted values with the actual output data based on the input data. The second group is used for testing purposes with the aim of identifying the response of the machine learning algorithm under unknown data. Classes of algorithms included in this group are classification and regression algorithms. Classification algorithms are better for those situations where the output is restricted to a limited set of values. Regression algorithms are conceived for those cases where the output may have any numerical value within a specific interval. Examples of supervised learning algorithms are Support Vector Machine, Random Forest, and Decision Trees.

Unsupervised learning: It includes algorithms capable of recognizing patterns or similar characteristics by using only output data. Typical problems where this approach is applied are clustering problems, anomaly detection, and density function finding. In general, these algorithms need some entry parameters for proper performance besides the data. For example, some entry parameters in clustering applications are the maximum distance between two points, the minimum amount of points for being considered a cluster, and the desirable number of clusters. Some examples of unsupervised algorithms are k-nearest neighbors (k-NN), Local outlier factor (LOF), k-means clustering and Density-based spatial clustering of applications with noise (DBSCAN). The first two are examples of anomaly detection algorithms and the second two are algorithms for clustering problems.

Reinforcement learning: It follows the idea of try and error. The algorithm learns by following a system of rewards for the correct answers or predictions and penalties for the incorrect ones. The algorithm generates new knowledge under this approach, which the algorithm uses to explore unknown data and generate more knowledge. In this case, the data is called the environment, which is affected by actions coming from the algorithm. The results are evaluated and interpreted through rewards. The objective of these rewards is to indicate if the evaluated action is in the correct direction or not. Some examples are Q-learning, State-action-reward-state-action algorithm (SARSA), and Temporal difference learning.

Machine learning algorithms are handy tools for modeling and predicting datasets. However, we should consider the fact that they require large amounts of data to obtain accurate results.

In the context of this thesis, we use supervised and unsupervised machine learning algorithms. Specifically, we apply random forest for a regression problem and DBSCAN for a clustering problem.

3.2.1 Error and validation strategies

We use machine learning methods as a technique to recognize patterns and predict the values of a certain dependent variable, according to a number of independent variables. In order to achieve it, we use a dataset, which is split in a least two partitions: the first one is the training set, which it is used to train the model, and the second one is the testing set, which it is used to test the model performance.

In order to quantify the accuracy of our model, we have to define the amount of error at the outputs. More generally, the difference between the output predicted by the learner and the groundtruth output (presented in the dataset) is called error. In this case, the literature establish that the error calculated on the training set is called training error or empirical error and the error calculated on the new samples or testing set is called generalization error. We wish to have a learner with a small generalization error. However, since the details of the new samples are unknown during the training phase, we can only try to minimize the empirical error in practice.

In order to define and quantify how accurate is a machine learning-based model, there exist a number of techniques, which we proceed to describe:

Simple Validation This method consists of randomly dividing the available observations into two groups, one used to train the model and the other to evaluate it. However, this technique has two main considerations: **(i)** The error estimate is highly variable depending on which observations are included as the training set and which are included as the validation set (variance problem); and **(ii)** By excluding part of the available observations as testing data, there is less information available to train the model, resulting in an overestimation of the error compared to what would be obtained if all the observations were used for training (bias problem).

Leave One Out Cross-Validation It is an iterative method that starts using all available observations as a training set except one, which is excluded for use as validation. If a single observation is used to compute the error, the error varies greatly depending on which observation is selected. To avoid this, the process is repeated as many times as available observations, excluding a different observation in each iteration, adjusting the model with the rest and calculating the error with the excluded observation. Finally, the error estimated by Leave One Out Cross-Validation (LOOCV) is the average of all calculated errors. The LOOCV method makes it possible to reduce the variability that arises if the observations are randomly divided into only two groups. This is so because at the end of the LOOCV process, all available data is used for both training and validation. As there is no random separation of the data, the LOOCV results are fully reproducible.

The main disadvantage of this method is its computational cost. The process requires that the model be readjusted and validated as many times as available observations in the dataset. Exceptionally, in least squares regression and polynomial regression, due to their mathematical characteristics, only one adjustment is necessary, which greatly speeds up the process.

LOOCV is a very widespread validation method since it can be applied to evaluate any type of model. However, there exist the risk of falling into overfitting since the method uses all the observations as training, with the aim of minimizing the error.

k-Fold Cross-Validation This method is also an iterative process. It consists of dividing the data randomly into k groups of approximately the same size. $k-1$ groups are used to train the model and one of the groups is used as validation. This process is repeated k times using a different group as validation in each iteration. The process generates k estimates of the error whose average is used as the final estimate. In this case, the number of necessary iterations is determined by the selected value of k. In general, a k between 5 and 10 is recommended, which present a good computational tractability in comparison with LOOCV.

However, the main advantage of K-fold CV is that it achieves an accurate estimation of the test error thanks to a better balance between bias and variance. LOOCV uses n-1 observations to train the model, which is practically the entire available data set, thus maximizing the fit of the model to the available data and reducing bias. Nonetheless, for the final estimate of the error, the estimates of n models trained with practically the same data are averaged (there is only one data difference between each training set), so they are highly correlated. This translates into a higher risk of overfitting and therefore of variance. In the K-fold CV method, the k groups used as training are much less overlapping, avoiding the risk of overfitting.

Bootstrapping It is a technique based on obtaining a sample from the original one by means of replacing observations (resampling with replacement), and of the same size as the original sample. Resampling with replacement means that, after an observation is extracted, it is made available again for subsequent extractions. As a result of this type of sampling, some observations will appear multiple times in the bootstrap sample and others will not appear at all. Unselected observations are called out-of-bag (OOB). For each bootstrapping iteration, a new bootstrap sample is generated, the model is fitted with it, and it is evaluated with the out-of-bag observations.

In this thesis, we use the Simple Validation and the k-Fold Cross-Validation since the combination of both techniques ensure that we can avoid bias and variance problems with an acceptable computational resource consumption.

3.3 Online optimization

In this thesis, we use online optimization, which we could define by comparison with the standard and well-known offline optimization. The main difference between both is the amount of information accessible at the decision moment.

Some literature related to this topic are Borodin and El-Yaniv [2005], Bubeck [2011], Fiat and Woeginger [1998], Grötschel et al. [2013], Karp [1992] and Dunke et al. [2014], which are the baseline for this section.

According to Borodin and El-Yaniv [2005], in an online optimization problem, an algorithm decides how to act under incoming information items without any knowledge of the inputs in the future. In an offline optimization problem, the algorithm is aware of the incoming information items since it has complete access to information. In colloquial terms, we could say that online optimization is being a general on the battlefield, looking for the best decision while the battle is happening. Meanwhile, offline optimization is being a general after the battle, defining the best decisions, and knowing the events that happened on the battlefield.

In a similar manner, Dunke et al. [2014] point out that decisions have to be made on an ongoing basis, and algorithms for determining them have to deal with incomplete information available. This information is usually presented as an input sequence $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ where the individual elements σ_i, with $i = 1, \ldots, n$, are only gradually known to an online algorithm Alg for processing σ. Hence, the characteristics of σ in terms of the amount of available information for Alg, define how complete or partial the solution could be.

An excellent example in the EMS context is the ambulance dispatch problem. An offline optimization approach is to solve the problem using the historical data for a specific time interval, meeting a threshold response time. An online optimization approach is solving the problem according to how the emergency calls arrive at the system.

In order to know the quality of the performance of an online algorithm, competitiveness has been established as the main criterion [see e.g. Sleator and Tarjan, 1985, as early work related to this topic]. In general, the competitiveness is defined by comparison with the offline solution of the problem. For instance, in a cost minimization problem, an online algorithm Alg is c-competitive if, for all σ, $Alg(\sigma) \leq c\ Opt(\sigma) + a$ with a constant a, where $Opt(\sigma)$ is the optimal solution for σ of the offline algorithm.Then, the competitiveness of Alg is defined as the smallest value of $c \geq 1$ such that Alg is c-competitive.

However, a competitive comparison with an offline algorithm is not always the best approach since it means that Alg has to deal with a double worst-case orientation: Alg not only has to compete against the strongest possible offline algorithm Opt, but also has to keep the competitive factor on any given input sequence σ. In this context, comparing two online algorithms could give more information considering real-world situations, as proposed previously by Dunke and Nickel [2016].

In general, Online Optimization could be divided in two main groups: *classical* online optimization, and online optimization with look-ahead. The first group corresponds to the given definition previously, where the Alg is aware of σ_i at the decision moment. Hence, the Alg its not aware about the future in any case. In the second group, Alg is aware of not only σ_i, but also it is aware at least of σ_{i+1} at the decision moment. The more elements of the input sequence σ in advance Alg is aware, the bigger the look-ahead is and a better and more competitive decision is achieved.

In this dissertation we use the *classical* online optimization and we implement algorithm variations inspired on the concept of look-ahead to address the problem that inspires this work.

Chapter 4

Modeling a pre-hospital Emergency Medical Service using hybrid simulation and a machine learning approach

This chapter presents the problem related to the representation of a pre-hospital emergency medical process, which is addressed by developing and implementing a hybrid simulation approach. First, we present the literature related to this topic. Subsequently, we present a process model for the German rescue chain and the validation process. Then, we establish the characteristics of the agent modeling and the simulation modeling in general. Moreover, we propose a machine learning approach for a realistic simulation modeling process. Finally, we design an experiment with real-world curated data and discuss the results.

Contents of this chapter have been published in Olave-Rojas and Nickel [2021].

4.1 Related work

The use of simulation models to evaluate the performance of EMSs are not new and they have been used in the early stages of the professionalization and modernization of emergency medical process [see Swoveland et al., 1973, Wilmot, 1969, as early examples of the use of this technique in EMSs] As explained in the previous section, EMS is complex, with multiple interactions between resources. Since simulation is one of the best approaches in terms of performance under these complex conditions, this approach is also suitable for proposing and modeling solutions, besides other operations research techniques. Therefore, our review focuses not just on articles presenting simulation models as a central contribution but also on introducing simulation models combined with optimization or statistical models.

Commonly, in the second group of articles, simulation models are proposed as an evaluation or assessment tool to determine the solution quality or the impacts of the solution implementation. We adopt this approach since most of the literature related to healthcare emergency pathways, and simulation models are concentrated on Emergency Departments (ED) and less on EMS. Furthermore, we take into account five characteristics and their study in the literature.We present an overview of the literature analysis in Table 4.1. These characteristics encompass, which EMS resources are modeled, the EMS philosophy characterized by the simulation model, how travel distances are considered, which population area is modeled, and if emergency differentiation is taken into account.

The information displayed in Table 4.1 shows that most of the work in recent years does not address the interaction between two or more resources. For instance, Koch and Weigl [2003] model both ambulances and helicopters but do not explore their relation. However, both Aringhieri et al. [2007] and Aringhieri [2010] are exceptions since they explore the relation between ambulances and call centers. Note that, Aringhieri [2010] is the only work presented as an agent-based simulation (ABS) and a discrete-event simulation (DES) simultaneously. The authors call DES to model workflow, identifying bottlenecks, and ABS as a better approach for modeling the interactions among ambulances and operation centers.

As Al-Shaqsi [2010] presents, EMS's two main design and process philosophies are the Franco-German approach and the Anglo-American. The main differences are that the Anglo-American philosophy depends on paramedics and rapid patient transport to the hospital. Meanwhile, the Franco-German philosophy includes emergency doctors and focuses on patient stabilization at the accident place. Nevertheless, both approaches are focused on the same objective in life-threatening cases, which is a fast response to saving lives. According to Al-Shaqsi [2010], the process differences between both systems are presented in non-life-threatening situations and by patient transport.

Although both philosophies are well studied, none of the models in the literature under the Franco-German philosophy are tested using a German EMS.

We also identify works focused on call-centers, such as Ünlüyurt and Tunçer [2016], van Buuren et al. [2017], and Petitdemange et al. [2019]. In these works, the authors concentrate the efforts on call-center problems, but ambulance resources are not modeled as agents and are represented through events or timestamps.

The research has been concentrated in urban areas instead of rural areas in the last years. We assume this is a consequence of having more available data related to cities. However, the complexity of the interchange between urban and rural areas has been hardly addressed. Concentrating efforts on models based in more populated areas is expected since the more available data, the better the understanding of a particular region. This understanding could be critical in the early stages of modeling since the possibility of errors is less with a more painless validation process. This situation presents research gaps for applying simulation models to model emergency calls from both urban and rural areas simultaneously, sharing resources, and establishing an interplay between both areas.

Besides the characteristics mentioned above, we consider travel speed and distance as critical parameters since they are strongly connected to travel time and response time. As presented in Table 4.1, time estimation for distance and travel speed, either from data or statistical distribution, is used in most cases. Only Pinto et al. [2015] establish a relation between speed and some independent variables. By contrast, Jánošíková et al. [2019] use Geolocalization Information System (GIS) data to estimate a particular path's speed. This approach considers speed limits for city streets and highways but does not consider traffic or related historical data. These simplifications could lead to mistakes or a restricted vision when evaluating emergency policies in our simulation model, as described by Aringhieri [2010].

As well as the works described in this section, the summarized articles in Table 4.1 address, but not entirely, the interplay between not just agents, such as ambulances, call-takers, and emergency doctors, but also between crucial variables such as response time, distance, traffic, and travel speed. Nonetheless, none of them provides a simulation model consolidating all agents and variables by integrating them in a detailed manner. This work, which is also described in Table 4.1, proposes a simulation hybrid model integrating crucial agents in the EMS, modeled by ABS, and variables and parameters, mainly part of the DES, building a hybrid simulation model. The simulation is embedded in a GIS environment, allowing a detailed analysis of critical travel variables. Furthermore, we present a speed-based approach for modeling the dynamical traffic on streets and highways and its influence on response times.

Reference	Resources			Emergency	Process philosophy		Population	Travel		Country
	Amb	Cc	Others	Triage	Fra-Ger	UK-USA	Area	Distance	Speed	
Koch and Weigl [2003]	✓		✓		✓		urban + rural	euclidean	fixed	Austria
Aringhieri et al. [2007]	✓	✓		✓	✓		urban	euclidean	time estimation	Italy
Bayer et al. [2010]	✓					✓	urban	time	time estimation	UK
Aringhieri [2010]	✓	✓		✓	✓		urban	euclidean	fixed	Italy
Churilov et al. [2013]	✓					✓	urban	time	time estimation	Australia
Aboueljinane et al. [2014]	✓			✓	✓		urban + rural	euclidean	time estimation	France
Kergosien et al. [2015]	✓					✓	urban	time	time estimation	Canada
Pinto et al. [2015]	✓			✓		✓	urban	GIS	avg. by zones	Brazil
Ünlüyurt and Tunçer [2016]		✓			✓		urban	time	fixed	Turkey
van Buuren et al. [2017]		✓				✓	urban	time	time estimation	Netherlands
Aringhieri et al. [2018]	✓			✓			—	euclidean	avg. time	—
Lanzarone et al. [2018]	✓					✓	urban	time	fixed	Canada
Petitdemange et al. [2019]		✓		✓	✓		—	time	—	France
Jánošíková et al. [2019]	✓			✓	✓		urban + rural	GIS	GIS info	Slovakia
Karatas et al. [2020]			✓				sea	euclidean	time estimation	Turkey
Our work	✓	✓	✓	✓	✓		urban + rural	GIS	multi-variable function	Germany

Table 4.1. Summary of the literature related to EMS simulation models, incorporating resources (Amb = ambulances, Cc = call center), emergency triage and methodologies for modeling travel distance or time.

4.2 Process model for the German rescue chain

In the *4th Bad Boller Reanimationsgespräche workshop* organized by the German Society for Anesthesiology and Intensive Care [Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin e.V., 2017], a group of 70 experts from different disciplines and interest groups stated the need for an in-depth analysis of the survival chain in the German emergency medical service. This group aims to model the German EMS with a strong focus on the heart attack pathway. However, the resulting German EMS model from the analysis also includes a general process pathway for other diseases validated by the interdisciplinary group of experts.

This pathway includes steps from the emergency scene, rescue team, coordination center, and hospital team. Nevertheless, our simulation model shall include all of them, except for the hospital team, which we consider as a particular process.

Figure 4.1 presents a flowchart of the general EMS model based on the findings of the workshop. It shows the interaction between actors and which of them completes each procedure. Furthermore, it is possible to identify the patient pathway through the entire emergency process, starting when the emergency occurs (*Emergency event*). Thereafter, someone on the scene (i.e., it could be the patient or someone else as a qualified first responder or a bystander) calls to the emergency service, establishing a communication link (dashed line) with the coordination center. The coordination center receives the call (*Emergency call entry*) and activates the respective procedure (*Processing emergency call*). In case of the emergency involves a heart attack, the call-taker performs cardiopulmonary resuscitation assistance to the person on the scene. After collecting the information from the scene, the coordination center (more specifically, the dispatcher in) selects the resources (*Booking Resources*) and sends the alarm message, which is received by the corresponding rescue team.

The alarm message triggers the rescue team process, starting with the preparation according to the received information from the coordination center. When the rescue team arrives to the emergency place, they take the responsibility for the situation (*Taking over from a first responder*), and carry out an evaluation in order to define the best treatment in place. If the patient needs a complex treatment, the rescue team contacts the coordination center to demand more resources such as an emergency doctor. Once the patient has been stabilized, the rescue team proceeds to transport him to the hospital, in case of a transport is required (*Transport to Hospital*). Meanwhile, the coordination center selects the hospital and sends the information for patient registration.

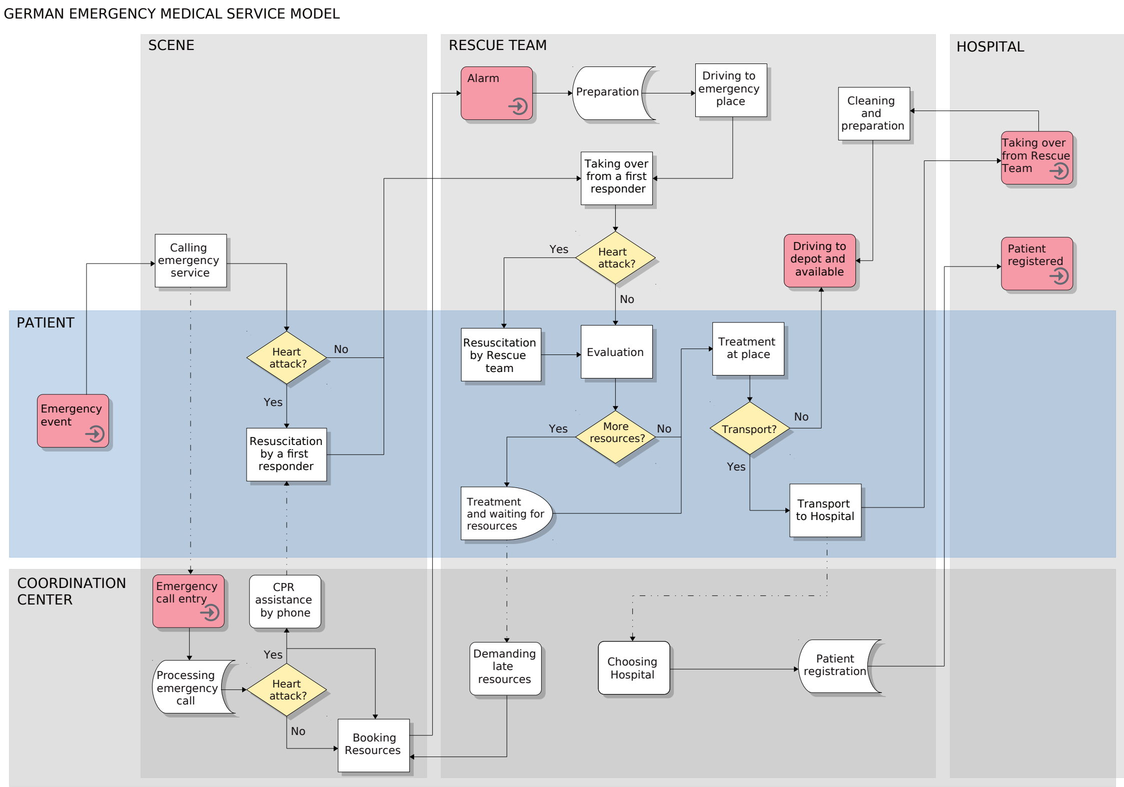


Figure 4.1. Flowchart of the general model of German emergency medical service. Dashed lines represent communication links, whereas lines represent transitions. Red boxes are milestones on the process

As explained in Section 2.3.1, model is generally performed by four main agents (call-takers, ambulances, emergency doctors, and helicopters). These agents accomplish complex tasks following a specific protocol for each situation. This leads us to the idea that each agent should be modeled as an independent entity in an agent-based simulation model. As in discrete event simulation models, emergencies enter the system as events, triggering a succession of operations. Hence, a hybrid simulation model arises to be the best approach to address the situation's complexity. It allows us to model details that profoundly affect the system's performance after emergency events. Furthermore, this approach permits us to model situations where operational decisions must be taken by coordinating a group of agents. In the following, we describe the characteristics of the three agents *Ambulance team*, the *Emergency doctor* and the *Call-taker*, as well as some necessary considerations for our simulation.

4.3 Agent modeling

4.3.1 Ambulance team

In the German EMS, ambulance teams perform different tasks such as patient transport and handling accidents. Due to the importance of this resource in the system, we model the ambulance team as an agent. In Figure 4.2a, we present the real-world agent protocol for ambulance teams. This protocol includes two combined states which are a novelty of this thesis: the *Idle* state and the *PreTreatment* state. The *Idle* state encompasses three sub-states where the ambulance team is available for dispatching and handling an incoming emergency call. These three sub-states are the following:

***Free* state:** In our simulation model, the connection between the patient and ambulance team is over when the patient does not need transport after a treatment at the emergency site or when the hospital takes over the patient. Then, after the *Preparation2* state, the ambulance team is available to attend an incoming emergency call.

***DrivingToBase* state:** The ambulance Team is in this state when the team returns to the base. It happens when the ambulance team has finished a certain commitment.

***AtTheBase* state:** This state represents the moment when the ambulance team is located at the base.

The ambulance team can be selected and dispatched in these three states since it is idle, as presented in the real world. As expressed in Section 4.1, this feature is not correctly addressed by the literature.

Since our simulation is aware of the emergency seriousness and the emergency call prioritization, we can re-dispatch ambulance teams that have been assigned to low priority calls. The *PreTreatment* state contains the activities between the ambulance dispatch and arriving at the emergency site. At this moment, the ambulance team has not arrived at the patient. Hence the ambulance team is available to re-dispatch it for a high-priority emergency call. The sub-states in the *PreTreatment* state are the following:

***Preparation* state:** In this state, the ambulance team receives information about the emergency call from the call-taker. Hence, this is the entry state of the process.

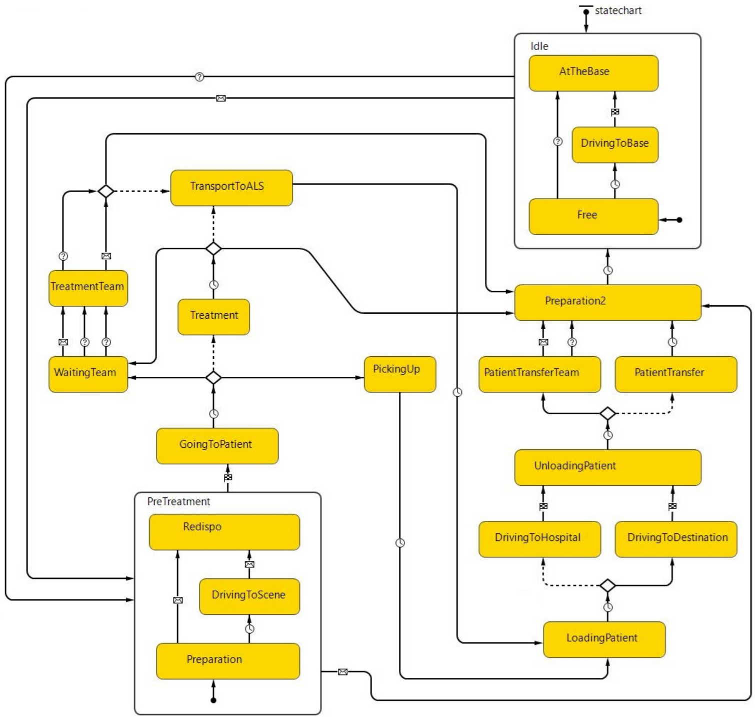
***DrivingToScene* state:** This state represents the moment when the ambulance team travels to the patient after receiving the information.

***ReDispo* state:** If the call-taker considers a dispatched ambulance team the best one for a high-priority level emergency, the ambulance team shall be re-dispatched. The *ReDispo* state state represents this moment..

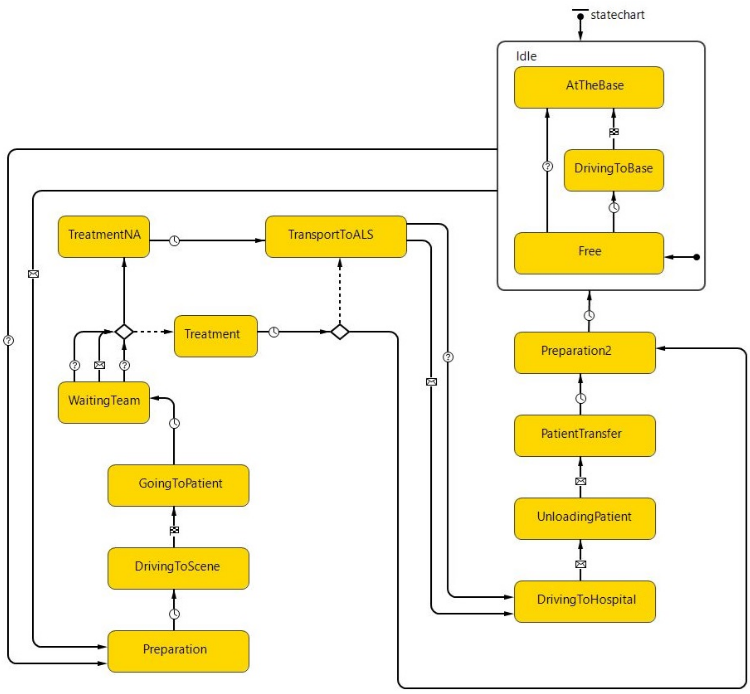
The idea related to the *PreTreatment* state exists in real life and is helpful for life-threatening emergencies, improving the response of the entire system. Another interesting state is the *Preparation2* state, which involves activities (ambulance cleaning, supplying, and ambulance disinfection) after serving a patient and before being ready for dispatch. These actions are presented in the process, but they are not adequately addressed in the literature. We consider these features critical for a realistic evaluation of the pre-hospital emergency process since the involved time in these actions could represent an essential amount of time in the entire process.

4.3.2 Emergency doctor

As described above, in the German EMS exist the figure of emergency doctors, who provide help in the most challenging situations, such as life-threatening accidents. Figure 4.2b depicts the agent protocol for an emergency doctor, which specifies the teamwork between the ambulance teams and the emergency doctors. The team communication between both agents is quite significant, considering that an emergency doctor



(a) ambulance team agent protocol



(b) emergency doctor agent protocol

Figure 4.2. Graphical representation of agent protocols implemented in AnyLogic™.

never attends an emergency without an ambulance team. We model this by creating a link between the agents and controlling state transitions, synchronizing both agents. For instance, in life-threatening accidents, the emergency doctor guides the treatments at the accident location, performing the treatment as a team with the ambulance team. The *TreatmentTeam* state represents this situation in Figure 4.2a and the *Treatment* state in Figure 4.2b.

Hence, the ambulance team's state changes are synchronized with the state changes from the emergency doctor, and there exists a communication link for instructions in both directions. For instance, in the *DrivingToHospital* state, the patient, the ambulance team, and the emergency doctor usually travel to the hospital in the ambulance. The ambulance team performs this task. Consequently, the emergency doctor has to wait for a message from the ambulance team to change its state to the *DrivingToHospital* state. To the best of our knowledge, no other work considers this interplay between both agents working as a team. Therefore, the analysis of this feature is also a novelty of this thesis.

4.3.3 Call-taker

This agent handles crucial information quickly since its main task is the correct assessment and the collection of the incoming information. The more precise and detailed the information, the better and faster the emergency team's preparation and treatment are. As a matter of simplicity, we combine both the call-taker and the dispatcher in the same agent. Although, in some coordination centers, they are different persons who accomplish these tasks. This approach unifying both agents is not uncommon in real-life systems. Finally, the call-taker agent supports cardiopulmonary resuscitation per phone until the emergency team arrives. During this time, the call-taker cannot answer other calls but is still in contact with the emergency team.

4.3.4 Further considerations

4.3.4.1 Geographic Information System

We implement our simulation model using Geographic Information System (GIS) data. This approach helps to explore routing options or dispatch strategies to simulate actual travel routes for different scenarios, considering the actual infrastructure of the studied area.

4.3.4.2 Average travel speed

We consider average travel speed (ATS) one of the critical parameters to simulate the emergency process accurately since most of the response time belongs to travel time. Its prediction depends on several parameters. For instance, road traffic affects travel times: the more traffic, the slower the travel speed, resulting in a longer response time. Additionally, road traffic is not constant throughout the day, causing differences in response times at different moments of the day for the same emergency type in the same place. Furthermore, travel distance also affects the speed: long distances are commonly covered using highways or main roads, promoting high speeds. Meanwhile, short distances are covered using primary and secondary roads, streets or passages.

In this context, we identify the following critical factors for the travel speed: emergency resource state (i.e., *at the base*, *travel to the base*, *at the hospital*), the emergency priority (the more significant the severeness, the higher the priority), the resource type (emergency doctor or ambulance team), whether the siren is on or off, and the moment of the day. Considering this, we propose developing a machine learning algorithm model to predict travel speed and, subsequently, to predict travel times for each resource involved in an emergency. In Section 4.5, we present the application of this approach more in deep.

4.4 Validation

As explained by Churilov et al. [2013] and Law et al. [2000], there exist different approaches of simulation validation. In this work, we use three of these approaches: **(i)**

The expert validation, represented by the interdisciplinary group of 70 experts reunited by the German Society for Anesthesiology and Intensive Care [Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin e.V., 2017] in different workshops; **(ii)** the visual validation, performed by means of the implementation of a control panel shown in Figure 4.3, allowing us to control the accuracy and the logic of the process; and **(iii)** data validation by comparing historical data with those obtained from the simulation. We explain more about the last validation in Section 4.5.3.

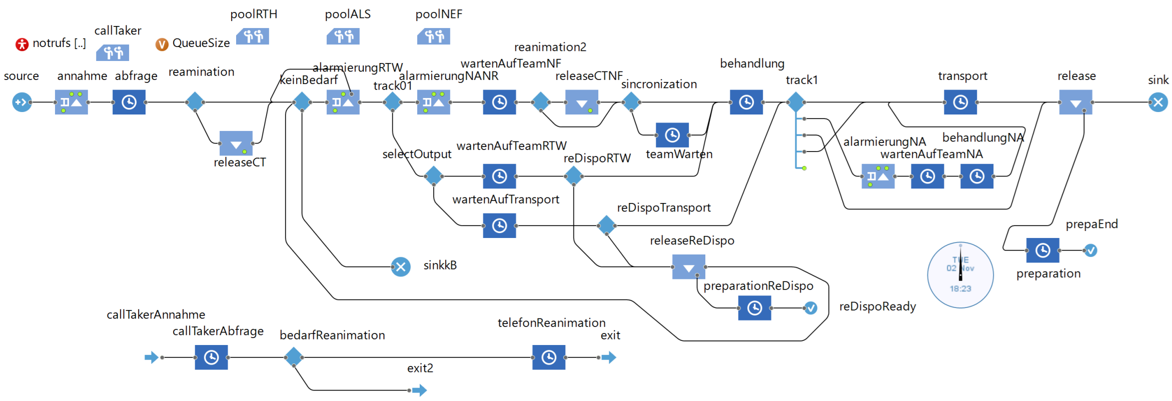


Figure 4.3. Detail of the control panel of the Simulation EMS model implemented in Anylogic™. This detail is part of the discrete event simulation model following the emergency call flow through the process.

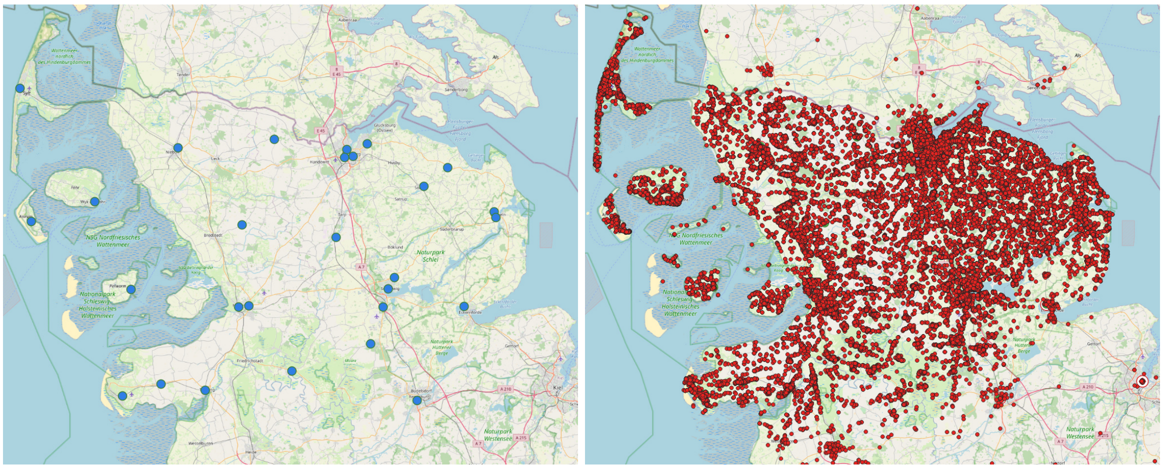
4.5 Experiment

4.5.1 Data

Since the EMS started a professionalization process in the early 60s, as described in 2.1, Germany has become a world benchmark in EMS, according to Lechleuthner [2019]. A majority of EMS regions have an integrated coordination center, i.e., the ambulance services and the fire department services are coordinated by the same coordination center, allowing better service to all types of accidents.

A good example is the Coordination Center North (*Leitstelle Nord* in German), which is located in Germany's northernmost region. This center is in charge of 58 ambulances, 14 emergency physicians, and one helicopter. Furthermore, it works together with six hospitals, covering a region of 4,211.29 $[Km^2]$ and 453,356 Inhabitants, according to Statistical Service for Hamburg and Schleswig-Holstein [Statistische Amt für Hamburg und Schleswig-Holstein, 2020].

Additionally, the region comprises a mixture of interesting characteristics such as urban areas, rural areas, and island and mainland regions. Moreover, the Coordination Center North is located next to the Danish border. This situation means that some accidents require resource coordination from Germany and Denmark simultaneously.



(a) Ambulance allocation across the Coordination Center North

(b) Emergency call localization

Figure 4.4. Ambulance bases and Emergency call locations in the Coordination Center North from 2017 and 2018.

In this context, the coordination center controls an area where it can find a great variety of emergency events. Figure 4.4b, presents around 270,000 emergency events from a curated database between 2017 and 2018 for the coordination center. As expected, there is an intense event concentration at the city locations. Nevertheless, a homogeneous distribution of emergency calls is present in rural areas. We also present in Figure 4.4a the existing ambulance base positions in the coordination center area.

4.5.2 Travel speed modeling

We consider travel speed the most crucial parameter for addressing a region's complexities in terms of orography, response time, traffic circulation, and resource availability.

Following this idea, we propose a machine learning model approach to predict travel speed. There are commonly two trips involved in an emergency event: the first one between the place where the ambulance is located at the moment when the emergency call comes in and the emergency place. The second one, if necessary, is between the emergency place and the hospital where the patient is finally allocated. Furthermore, an ambulance is dispatchable if it is *Idle*, being at the hospital, traveling back to the base or at the base.

Ambulances track their rides. They register when they receive a new emergency call and at which location. They track the localization of the patient and the time of arrival. Likewise, they enter into the system when they arrive at a hospital to drop off a patient. Generally, this happens by pressing a button in their IT-System, stating *trip finished*. This activity is automatically attributed with a timestamp and the ambulance's GPS location at that very moment.

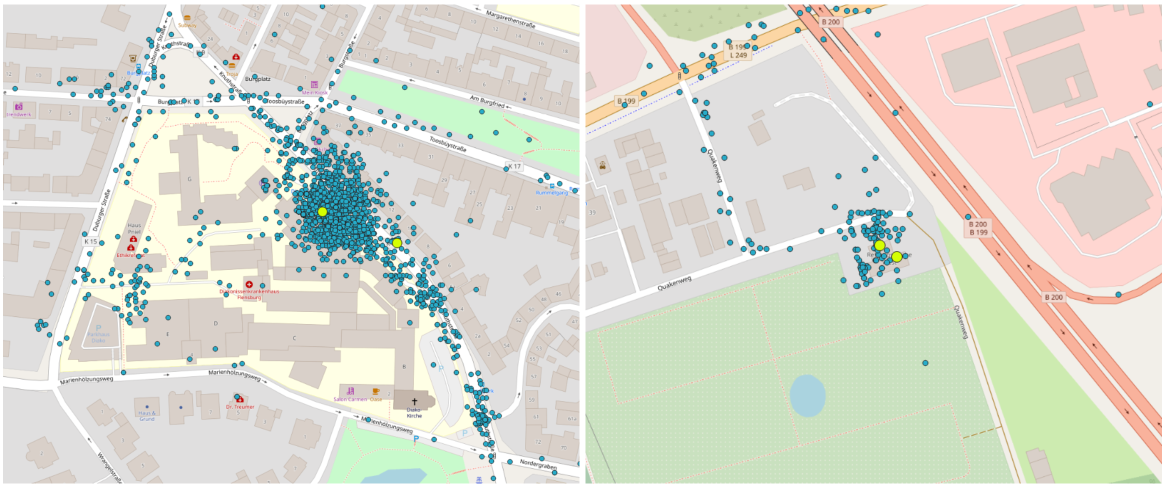
Finishing a trip, i.e., pressing the button to state the end of a trip, often happens not just at the hospital itself but often some meters or even some streets before. The same applies when being alarmed and leaving the hospital premises or an ambulance base.

In Figure 4.5, we show examples of ambulance positions at the alarming moment found in the data for both a hospital location and an ambulance base, represented by blue dots. These points represent the same place, i.e., the same hospital or ambulance place.

For the sake of tractability, we cluster the positions related to the same location, applying a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm Ester et al. [1996]. Figure 4.5 presents the situation related to a hospital (Figure 4.5a) and an ambulance base (Figure 4.5b). The blue dots are the alarm positions for ambulances; the yellow dots are the related clusters' centers. As parameters for the DBSCAN algorithm, the minimum cluster size is 30, and the maximum distance between two points of the same cluster is 5 meters.

Following the idea of calculating the travel speed, we need the distance between the start point and destination and the required time for covering the distance for travel speed calculation. The travel time is easy to get having timestamps corresponding to the two moments between the travel occurs.

Then, identifying the start point and destination coordinates related to the respective cluster center, we estimate the actual travel distance utilizing the *Openrouteservice* tool [openrouteservice, 2020], developed by Heidelberg University.



(a) Hospital **(b)** Ambulance Base

Figure 4.5. Examples of real alarm positions for ambulances (blue dots) and the respective cluster (green dots).

Using this information, we estimate the travel speed for each trip associated with a cluster, obtaining a total of 87,453 emergency events with travel speed data.

The travel speed prediction can be seen as a regression problem. In order to predict a suitable regression function, we apply the following machine learning algorithms: Generalized Linear Model (GLM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), Support Vector Machine (SVM), and Decision Tree (TREE).

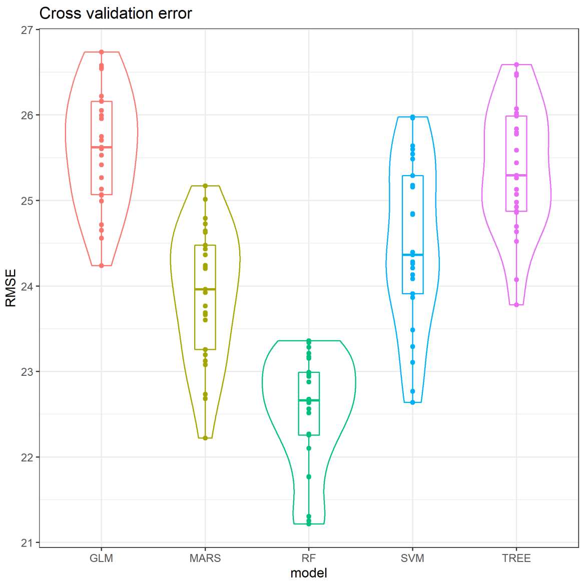
We select these algorithms since they are well-known and commonly used for regression problem applications.

After a brief data exploration, we identified as travel speed predictors the emergency priority level (the so-called triage level), the resource type (e.g., ambulance, emergency doctor), the resource state (i.e., at the base, going to the patient, among others), the siren state (on, off), the travel distance, and moment in the day (clustered by six groups of 4 hours each). For data training purposes, we use 15% of the data. The remaining

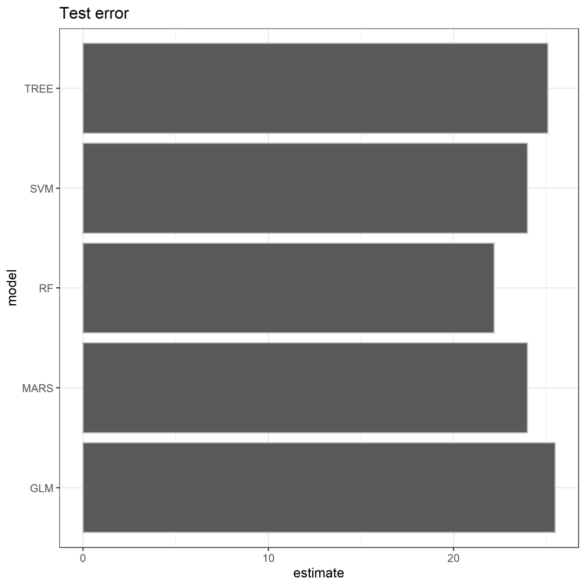
85%, we use for testing. We select this split due to the amount of data and the available computing capacity.

In Figure 4.6, we present the test results for the five machine learning algorithms. The tests are the 5-fold cross-validation error (Figure 4.6a) and the test error (Figure 4.6b). The metric is the root mean square error (RMSE), and the RF is the model with the best performance.

RF is the machine learning algorithm with the best performance for our problem structure, showing the smallest test error of 22.19, compared to 23.98 for TREE and SVM, 25.09 for TREE, and 25.48 for GLM. Also, considering the cross-validation error, the Random Forest performs best since it is the algorithm with the smallest average error and the smallest standard deviation compared to the other algorithms.



(a) Cross Validation Error

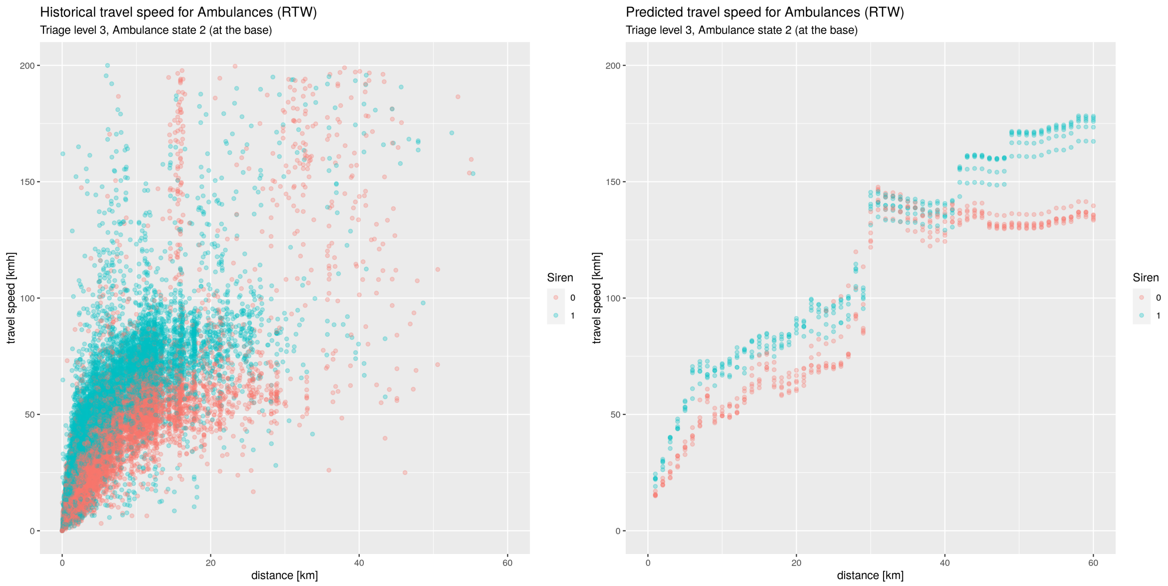


(b) Test Error

Figure 4.6. Test results of different machine learning models for travel speed prediction.

Furthermore, for illustrative purposes, we show in Figure 4.7 the historical and the predicted data for different distances, given an emergency priority level of 3, i.e., low priority, and an ambulance state being *AtTheBase* for both siren states. The RF model's predicted data consistently follows the data's tendency despite the strong variability presented for distances shorter than 30 $[km]$. The RF model also works for distances greater than 40 $[km]$, despite the few data compared with short distances.

Due to the complexity and variety of the predictors, we consider the machine learning approach as one of the best possibilities to predict travel speed applied to travel time modeling.



(a) Historical data for travel speed **(b)** predicted data for travel speed

Figure 4.7. Historical and predicted data for ambulance travel speed.

4.5.3 Simulation parameters and validation

As presented in chapter 3.1.2, we use animation, face validity, trace, event validity and graphic comparison as validation and verification techniques in order to define the correct parameters that are involved in the simulation model. In this context, we use the curated database to obtain empirical density curves for most of the parameters. We prefer this approach over statistical curve fitting since it is more accurate for representing what occurs in reality.

In Figure 4.8, we present four graphs comparing the time for four process stages: information request time used by call-takers to recollect the information from the event scene; preparation time, which is the required time by the emergency team for being ready (also called reaction time); travel time, which is the time for travel to the scene; and treatment time, which is the time required for the emergency time at the scene for patient's treatment.

The implemented curves in the simulation are step functions since they are empirical. Furthermore, in Figure 4.8, we point out that there is a difference between historical and simulation curves for travel time and preparation times. This difference occurs because travel times and preparation times depend on many factors and not only on the type of emergency to be attended to. This circumstance presents a scenario where the preparation time for the same type of accident and the same team is not equal. However, we consider this difference between the curves to be positive since it makes the modeled scenarios more pessimistic than those presented in the real world.

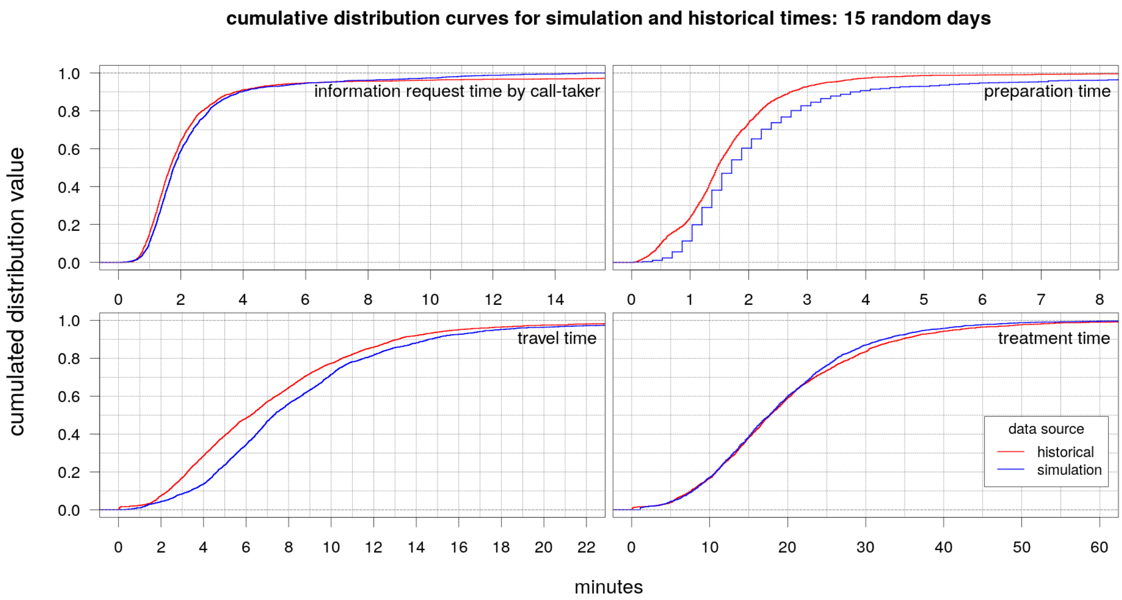


Figure 4.8. Cumulative distribution curves for simulation and historical data.

4.5.4 Experiment setup

To expose some of the use possibilities of our simulation model, we study the impact of the call-takers capacity change on the system's general performance by analyzing the response time. For this analysis, we use the concept of counting distribution presented in Dunke and Nickel [2016]. However, we apply this concept for scenario comparison instead of optimization function comparison in this work.

Call-takers are one of the critical resources of the entire emergency process. The call-taker crew is in charge of several tasks: attending the incoming emergency calls, assessing the emergency seriousness, selecting the emergency team and doctor according to the seriousness, and assisting with cardiopulmonary reanimation by phone. Call-takers' performance and the selection of the right capacity could lead to better results in terms of shorter response times. Furthermore, the required investment in a new call-taker could be less than the required for a new ambulance team, but the entire system's impact could be the same or more profound.

For this experiment, we propose five scenarios where the total amount of call-takers is different: The base scenario is the existing configuration in the control center, consisting of 4 call-takers during the day shift (07:00 to 19:00 Hours) and two workers on the night shift (19:00 to 07:00 Hrs). We call it "4+2 call-takers", and the other four scenarios are variations of this defined base scenario. The second and third scenarios are the base scenario plus one and minus one call-taker from the day shift (5+2 and 3+2, respectively). The last two are the half capacity and double base capacity (2+1 and 8+4 call-takers, respectively).

Additionally, we perform these scenarios employing our simulation model, running 40 instances. Each instance is composed of one day randomly selected from the curated emergency call database. Furthermore, this data corresponds to emergency calls between 2017 and 2018, as explained above.

The simulation time unit corresponds to one minute, and the model is implemented on Anylogic™8.2. Furthermore, we run our experiment in an Intel® Core™ i7-4600U 2.10GHz machine with 8GB RAM and Windows™10.

4.6 Results

We present in Figure 4.9 three graphs with the results for our experiment set up according to the five scenarios presented in Section 4.5.4. In Figure 4.9a, we display the counting distributions for the scenarios. In this context, the call-taker capacity variation does not statistically represent a real variation in response times. Furthermore, the utilization resource proportion is the ratio between the busy time of the resources over the total available time. By comparing the respective utilization proportions, we expose in Figure 4.9b that this indicator is stable for each scenario and less than 20%, except for the “2+1 call-takers” scenario. This result means that the actual “4+2 call-takers” scenario could be oversized.

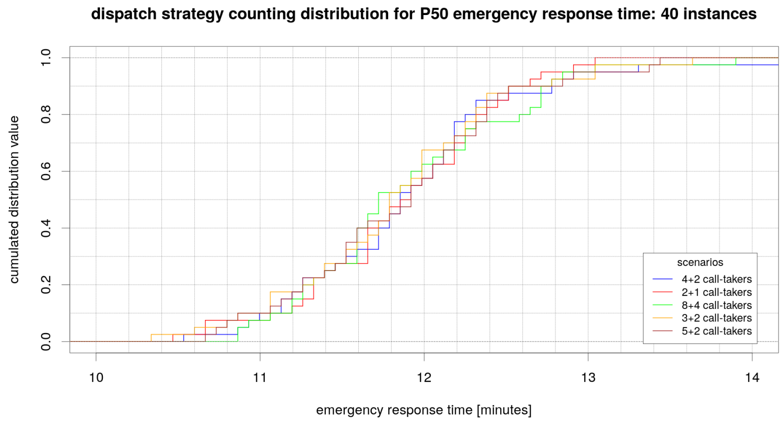
Nevertheless, we consider the current scenario the best option, considering a robust approach, since the exact future demand is stochastic, despite some profile tendency existing in demand. In the scenario “2+1 call-takers”, the call-taker utilization ratios are between 20% and 40%. Given the actual demand, we consider this result the capacity’s collapsing border. Any increment in the number of calls is not manageable in this scenario.

Response time and resource utilization are two indicators that help to understand the system’s performance or, in this case, the performance of a resource. However, the most important agents in this process are the beneficiaries, i.e., the patients. Furthermore, response times are not strictly linked to call-takers performance but the emergency team. Hence, we could argue that response time is not the most significant indicator for assessing call-takers performance.

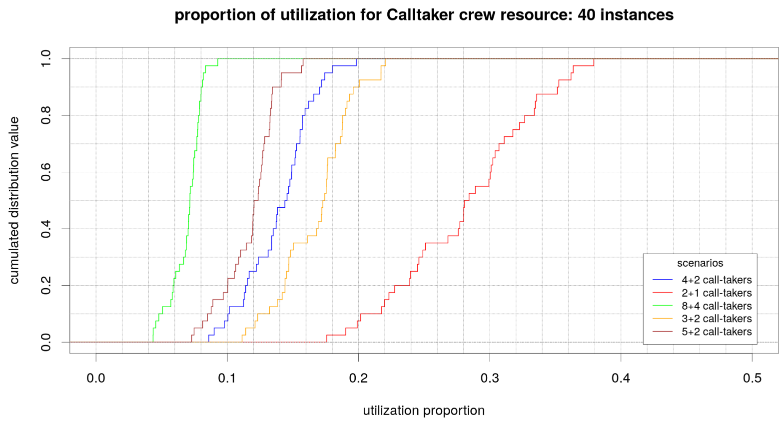
Taking this into account, we present in Figure 4.9c the results for the maximum waiting queue for entering the system, i.e., the queue for waiting for some idle call-taker. Again we notice the same tendency as in Figure 4.9b. The longest queue for each instance is between 2 and 6 emergency calls for all the scenarios except “2+1 call-takers”. This scenario presents a queue of between 2 and 8 emergency calls and consistently presents between 1 and 3 calls more than other scenarios for 90% of the instances. We again consider this last scenario as the border where the system collapses.

For the scenarios different than “2+1 call-takers”, we consider the results are the same since the performance is equal for 70% of the instances.

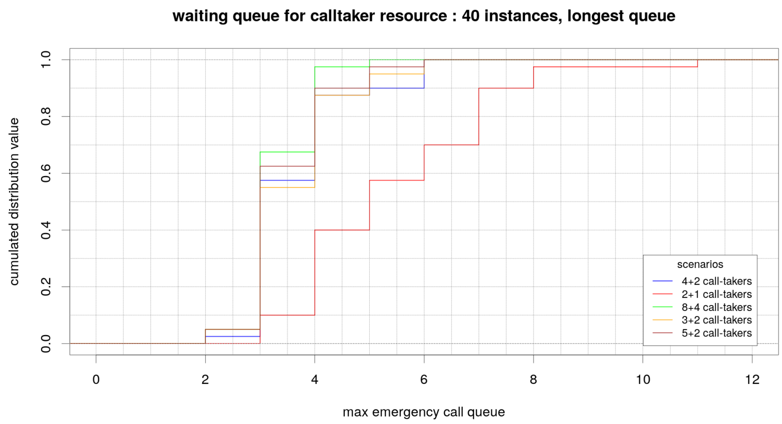
After this experiment, we consider the base scenario (“4+2 call-takers”) as the coordination center’s best configuration. This configuration offers almost the same performance as the scenarios with more capacity. Moreover, it presents a buffer in



(a) counting distribution for emergency response time by experiment scenarios



(b) cumulated distribution for call-taker utilization by experiment scenarios



(c) maximum waiting queue by experiment scenarios

Figure 4.9. experiment results for proposed call-taker capacity scenarios.

terms of utilization for unexpected disasters such as terrorism and natural disasters where demand escalates quickly.

4.7 Conclusion and outlook

Emergency Medical Services is one of the most important organizations to ensure life quality. Since there are many unpredictable situations in the real world, such as earthquakes, tsunamis, terrorism attacks, fires, and traffic accidents, an in-depth analysis of these services is required. However, this analysis and understanding generate the need for tools that include the multiple variables and existing parameters, presenting a closer representation of reality.

In this paper, we presented a hybrid simulation model of a pre-hospital emergency medical service. This model is based on a resulting process from a workshop. In this workshop, experts from several disciplines developed and analyzed the emergency rescue chain and established the general German system-based emergency process.

The simulation model considered the interplay between the resources, which affects the entire process's performance for the incoming call and future calls. Furthermore, there is a particular emphasis on the connection between call-takers, ambulance teams, and emergency doctors for life-threatening emergencies, such as a cardiac arrest. This is possible through a hybrid simulation approach, which is unexplored in this field. Additionally, we also presented a machine learning approach for travel speed modeling. This technique admits several parameters which influence the travel speed of ambulances and emergency doctors to the accident scene. Since much of the response time is determined by travel time, we considered the travel speed critical for the entire system performance. This approach is also new and unexplored, as presented in Section 4.1. Hence, this work pursuits addressing non-linearities presented in the system such as traffic, weather and maximum travel speed by means of the implementation of the techniques described above in emergency medical services.

Using a real case study located in a northern region in Germany and characterized by curated data, we showed improvements compared to existing tools. These improvements lie in a more realistic representation of the entire pre-hospital emergency service since it considers the abovementioned approaches..

Since there are just some works in the context of the study of call centers (see Table 4.1), including urban and rural areas, we tested our simulation model's capability by performing a crew capacity experiment based on the operational decisions of the entire process. This experiment is built using real-world data from the coordination center north, located in the North of Germany.

Nevertheless, the capabilities of our model are broader, including dispatch strategy analysis, ambulance facility location analysis, and system stress analysis, among others. These possibilities are based on considering the operational decisions presented in the process, which enhance the decision analysis's accuracy on the tactical and strategical decision level.

Finally, we consider our model to represent Anglo-American-based EMS since the main difference between Anglo-American and Franco-German systems is the emergency doctor. Nevertheless, our simulation model needs to be tested with real-world data from one system based on the Anglo-American philosophy.

Chapter 5

A methodological framework for ambulance dispatch strategy evaluation in a hybrid simulation model context

In this chapter, we present the dispatch algorithms included in this thesis. Then we present the Simulation-Online Optimization Framework to evaluate online ambulance dispatch algorithms. After that, we show an experiment that aims to understand how some spatial and resources variation could affect the performance of dispatch algorithms. Finally, we discuss the results obtained from the experiment.

Contents of this chapter have been published in Olave-Rojas and Nickel [2021] and submitted in Olave-Rojas and Nickel [2023].

5.1 Dispatch strategies for an ambulance coordination center

Dispatch problems applied to emergency systems are well-known and have been well studied at least, in the last 50 years [see Daskin, 1983, Fitzsimmons, 1973, as examples in the literature]. Basically, these are the decision problems to determine which vehicle to assign to an emergency call. Is a matter of fact, that dispatching decisions about where to send emergency vehicles impact the system's future ability to meet demands adequately. If a vehicle is sent to a call in a particular region, it can be expected that the coverage in that region will degrade. However, relocating vehicles can help reduce such degradation. Therefore, dispatch decisions and relocation strategies are closely related, as presented in the literature [see Aringhieri et al., 2017a, Bélanger et al., 2019, Lu and Wang, 2019, Zaffar et al., 2016, as reviews where this relationship is exposed].

Nevertheless, we concentrate our effort on the effect and performance of dispatch strategies without any relocation or allocation improvement. In this context, we identify the following main ambulance dispatch strategies in the literature:

5.1.1 Nearest ambulance dispatch strategies

These dispatch strategies are the most common in coordination centers since the implementation is straightforward and inexpensive. Furthermore, several studies include it in order to evaluate the performance and to compare with other strategies, as presented by Lim et al. [2011], McLay and Mayorga [2013a], Schmid [2012]. In this case, we consider as representative the *nearest ambulance by euclidean distance* (*NearestByDistance*) and the *nearest ambulance by route* (*NearestByRoute*).

We consider these versions different since the response after the selection criteria could not be the same due to geographical features such as a river or a hill that shall be surrounded.

Considering γ as the emergency call and W as the set of ambulances in the system, these dispatch strategies are represented by the following algorithm:

Algorithm: Nearest Vehicle Dispatch Algorithm
input : γ, W.
output: selected ambulance for dispatch.
$flag$ = **false**;
$Vehicle_0$ = GetNearestVehicle(γ,W);
while $W \neq \varnothing$ *and* $flag \neq$ **true** **do**
    $Vehicle$ = GetNearestVehicle(γ,W) ; `// searching idle closest vehicle`
    if *Vehicle.inState(Idle)* **then**
        $flag$ = **true**;
    else
        Remove(W,$Vehicle$);
if $W = \varnothing$ **then**
    $Vehicle = Vehicle_0$;
return $Vehicle$;

5.1.2 Covering dispatch strategies

The aim of covering strategies is to improve the average response of the system in the long term using covering criteria. We consider in our work the following:

5.1.2.1 Maximal covering

It is a classical formulation implemented using mixed-integer programming. A so-called covering matrix represents the covering criteria. The area is represented by quarters assigned to a respective ambulance base. Let us define W as the set of ambulances in the system and Z as the set of quarters where the incoming emergency calls are located. Then, we define y_{ij} as the binary variable indicating if a call from quarter i is assigned to ambulance j; x_i as the binary variable indicating if quarter i is covered for some idle ambulance ($i \in Z$); and u_i as the binary variable indicating if ambulance i is idle or not ($i \in W$).

As parameters, let us define d_i, which is the demand from quarter i, and c_{ij}, which indicates if the ambulance i covers quarter j. After this definition, we implement the following formulation:

$$MAX_Y = \sum_{i \in Z} x_i \tag{FO.1}$$

s.t.:

$$\sum_{i \in W} y_{ij} = d_j, \qquad \forall j \in Z \tag{C.1}$$

$$y_{ij} \leq c_{ij}, \qquad \forall i \in W,\ \forall j \in Z \tag{C.2}$$

$$\sum_{i \in Z} c_{ij} y_{ij} \leq 1, \qquad \forall j \in W \tag{C.3}$$

$$u_j \leq (1 - y_{ij}), \qquad \forall i \in \delta\{j\},\ \forall j \in W \tag{C.4}$$

$$x_i \leq \sum_{j \in W} c_{ij} u_j, \qquad \forall i \in Z,\ \forall j \in W \tag{C.5}$$

$$y_{ij}, x_j \in \{0, 1\}, \qquad \forall i \in W,\ \forall j \in Z \tag{C.6}$$

Constraint C.1 ensures that the demand from each quarter is covered. Constraint C.2 assigns an ambulance j to an emergency call from a quarter i only if the quarter i is covered by the ambulance. Constraint C.3 ensures that only one ambulance can serve one emergency call. Constraints C.4 and C.5 ensure that a quarter is covered by idle ambulances. The nature of the involved variables is expressed in C.6.

Once the formulation is solved, we analyzed the solution to know which ambulance should be dispatched for a specific call. In this context, we look for the y_{ij} where i corresponds to the quarter where the emergency call is located. We call this strategy the *MaxCoveringByZones*.

5.1.2.2 Maximum expected covering location problem

The first version of this formulation is proposed by Daskin [1983]. After some modifications, Jagtenberg et al. [2017] proposed a dispatch formulation based on the Maximum Expected Covering Location Problem (MEXCLP). The principal idea is that ambulance resources have a certain probability of being busy based on the time when the ambulances are attending some emergency call. Then, the expected covering of an ambulance through time depends directly on this probability. Furthermore, the expected coverage of a specific point also depends on the among of ambulances reaching this point under a threshold time.

This concept is represented under the following formula for the dispatching problem:

$$\sum_{i \in D} d_i (1-p) p^{k-1} \tag{M.1}$$

Where D is the set of demand points, d_i is the number of emergency calls related to each point i, p is the ambulance busy fraction, and k is the number of ambulances covering point i.

Therefore, the dispatched ambulance for the emergency call will be the one with the smallest sum of marginal coverings over the emergency calls. This idea is summarizes by Jagtenberg et al. [2017]:

$$arg\ min_{x \in W_{idle}} \sum_{i \in D} d_i (1-p) p^{k-1} \cdot \mathbf{1}_{\tau_{Loc(x),i} \leq T} \tag{M.2}$$

Where W_{idle} is the set of idle ambulances, T is the threshold time for arriving at the patient, $Loc(x)$ is the location of the ambulance x, $\tau_{Loc(x),i}$ is the driving time between $Loc(x)$ and emergency i at the emergency categorization of i.

5.1.2.3 Mixed dispatch strategies

We also consider a mixed dispatch strategy. This approach is new since we found insufficient literature related to [see Bandara et al., 2014, as an example]. In our case, this strategy consists of using the *nearestByRoute* strategy for the worst and life-threatening cases; and the *MEXCLM* strategy for the other emergencies. With this approach, we expect a faster response in those cases where time is critical. Note that we suppose a relation between time response and survival probability, particularly for life-threatening cases. We call this strategy as *MEXCLM+NearestByRoute*.

5.2 Simulation-optimization framework for ambulance dispatch algorithms

Chapter 4, presents a realistic simulation model for the pre-hospital EMS based on exchanges with practitioners and decision-makers from the Health Care area. In Section 5.1, we present the dispatch approaches that we consider as principals in the literature. In order to evaluate the impact of dispatch algorithms in EMS, we present a bi-level approach for the evaluation framework.

Roughly speaking, the simulation model replicates the reality by means of the running of different instances. These instances are composed of a stream of emergency calls as input in the EMS process. Every call follows the EMS process, going through the flowchart presented in Figure 4.1. After the call-taker has collected the patient information, a dispatch decision has to be made. At the moment of the dispatch decision, the Simulation Model is in a specific configuration that represents the state of the EMS system at that particular moment.

We call it Process Snapshot $\boldsymbol{\lambda}$, and it is defined by the set of parameters that describes this moment, and it is the input for the online optimization model. Then, after running the dispatch algorithm, we obtain an ambulance $\boldsymbol{\varphi}$ which is the choice for dispatching.

Note that the online optimization model is required in real-time since the dispatch algorithm shall deliver the solution within a very tight time-bound. Then, the ambulance $\boldsymbol{\varphi}$ is an input in the dispatch process stage, allowing the simulation model to run according to the online solution. Finally, the emergency call goes through the following stages until it is officially out of the EMS process.

In Figure 5.1, we present the diagram of the Simulation-Online optimization framework that summarizes the interaction between both models.

With the aim of comparing the performance of dispatch algorithms, some indicators should be collected. The most common indicators are the response time and a covering indicator based on an area or a certain number of calls. In this chapter, we use both the average response time (since it is the most common indicator) and a covering indicator based on the area.

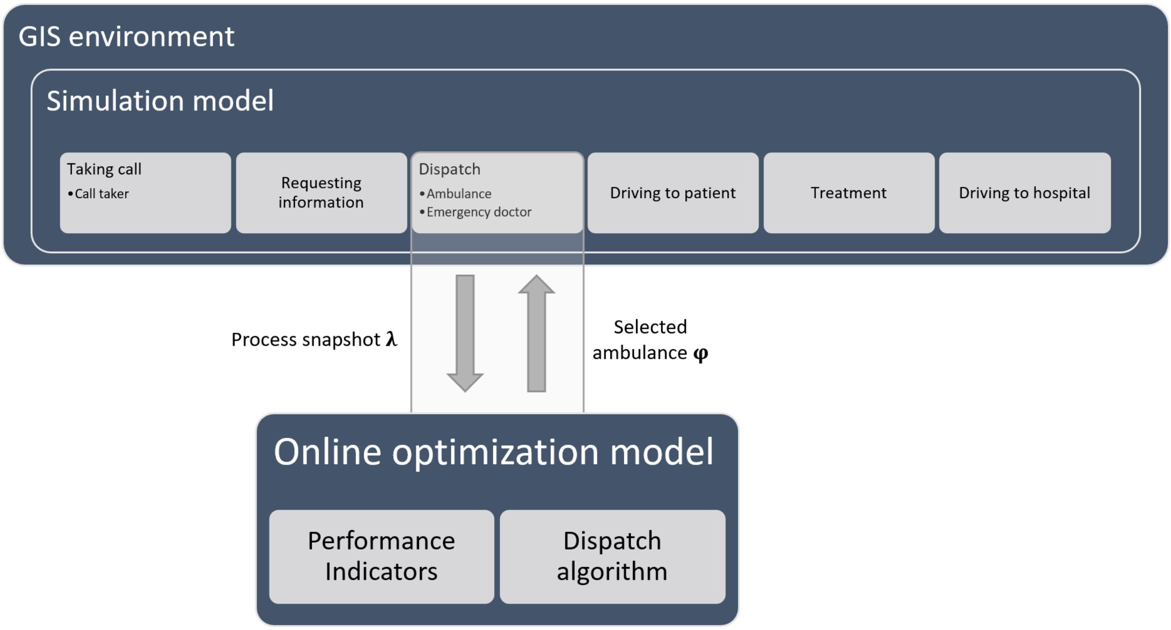


Figure 5.1. Simulation-Online Optimization framework for ambulance dispatch problem.

5.3 Experiment setup

We are aware that several events are taking place simultaneously in the EMS process. The nature of these events is stochastic. Even when they are not directly correlated, the configuration and how some resources are interrelated in the EMS process result in a no-linear environment. This non-linearity could be challenging to understand what is happening and which dispatch strategy is the best. To understand more deeply how some spatial variation could affect the performance of the dispatch algorithms, we build a theoretical experiment, proposing deterministic scenarios, which are easier to understand.

Inspired by the example in Jagtenberg et al. [2017], we create several spatial configurations presented in Figure 5.2, where $Z_i \in Z = \{1, \ldots, 9\}$ are the emergency call places and $A_j \in A = \{1, 2, 3\}$ are the ambulance depots. Please note that Z_2 in Figures 5.2a, and Z_5 in Figure 5.2b and 5.2c are in cyan since they are places where a hospital is located. The time between the knots is the response time for an ambulance in this environment, which means that the travel speed is fixed and is the same for every emergency.

Based on the spatial configurations of Figure 5.2, we define five scenarios. We present these scenarios in detail in Table 5.1.

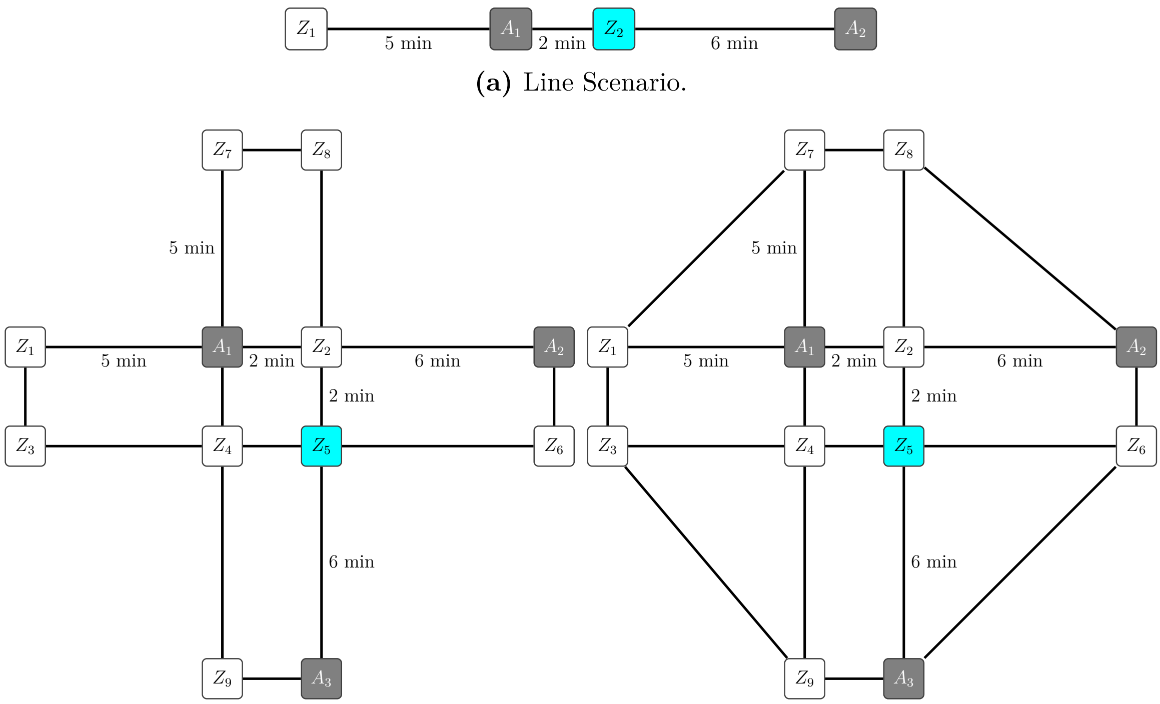
Scenario	Description
Line Scenario	Instance configuration shown in Figure 5.2a
Cross Scenario	Instance configuration shown in Figure 5.2b
Rhombus Scenario	Instance configuration shown in Figure 5.2c
Big Cross Scenario	Same configuration as Figure 5.2b but with double travel times
Big Rhombus Scenario	Same configuration as Figure 5.2c but with double travel times

Table 5.1. Description of the scenarios based on the configuration presented in Figure 5.2.

In order to obtain a deterministic version of the emergency process, we use the average time of each step in the emergency process based on two datasets coming from two emergency coordination centers: The coordination center North [In German, Leitstelle Nord, 2021] and the coordination center Karlsruhe Department of Quality Assurance of Emergency Medical Services in Baden Wuerttemberg [In German, Stelle zur trägerübergreifenden Qualitätssicherung im Rettungsdienst Baden-Württemberg, 2021]. The first one is located in Schleswig-Holstein, the northernmost state of Germany, and the second one is located in Baden-Württemberg, which forms the southern part of Germany's western border with France.

For the emergency events, we implement the following idea: every 18 minutes, a new emergency call entries to the system, randomly located in Z. Considering this, we build 30 instances of two days long. Finally, our experiment consists of the ambulance and emergency doctor capacity variation, composing several configurations. Table 5.2 presents the configurations included in the experiments.

Since some dispatch strategies presented in Section 5.1 are variations of a previous one, we consider these strategies a family and implement just one of them in our experiment. For instance, *NearestByRoute* and *NearestByDistance* are in the same dispatch strategy family since the dispatch philosophy is the same. Considering this, we include one strategy of each family: *NearestByDistance*, *MaxCoveringByZones*, *MEXCLM*, and *MEXCLM+NearestByDistance*, respectively. Note that *MEXCLM+NearestByDistance*



(a) Line Scenario.

(b) Cross Scenario.

(c) Rhombus Scenario.

Figure 5.2. Spatial configurations for analytical experiments. Knots A_i are ambulance depots. The Z_i (Cyan) is where the hospital is located. Finally, emergency calls come from Zones Z_i.

Version Name	Description
Version 1+2	resources with 1 ambulance and 2 emergency doctors.
Version 2+2	resources with 2 ambulances and 2 emergency doctors.
Version 3+2	resources with 3 ambulances and 2 emergency doctors.
Version 4+2	resources with 4 ambulances and 2 emergency doctors.
Version 5+2	resources with 5 ambulances and 2 emergency doctors.
Version 5+3	resources with 5 ambulances and 3 emergency doctors.

Table 5.2. Description of the resource configurations for experimentation.

induces the same results as *MEXCLM+NearestByRoute* since the distance and the route are the same in almost all cases for our instances.

5.4 Results

In Figure 5.3, we present the average emergency response time result for the "Line Scenario", which corresponds exactly to the representation in figure 5.2a. The first situation we recognize is that *MaxCoveringByZones* is clear the strategy with the worst performance. This strategy presents a strong dependence on the system's capacity: the bigger the capacity, the better the performance. This performance is entirely understandable since we consider it as a static strategy (based on a static and defined covering matrix) for a dynamic situation. For *NearestByDistance* we notice a slightly better performance for "Version 3+2" and "Version 4+2" compared to the other strategies.

Nevertheless, we cannot conclude that this strategy seems to be the best by the analysis of Figure 5.3. The same result appears for the Cross Scenario and the Rhombus Scenario. We include these graphs in Appendix A.

For a better analysis we have to examine Figure 5.4, which is a larger scenario than the "Line Scenario" (Figure 5.3), and Figure 5.5, which includes longer travel times than "Rhombus Scenario" (Figure 5.4). In this case, we also recognize a strong dependence not only for *MaxCoveringByZones*, but also for *NearestByDistance*. Both strategies present almost the same behavior under ambulance quantity changes, with a better performance in terms of response time for *NearestByDistance* by around 2 minutes in the last tested version, which are the "Version 4+2" for the "Rhombus Scenario" and the "Version 5+3" for the "Big Cross Scenario". Nevertheless, when the resources are not enough for the demand, *NearestByDistance* presents the worst performance as presented for "Version 3+2" in both Figure 5.4 and Figure 5.5.

The performance for *MEXCLM* and *MEXCLM+NearestByDistance* is almost the same no matter how the resource capacity is. Since we present aggregated results, there exists almost no difference between these strategies. In "Version 4+2" we present the cross point, where the tendency changes about which strategy is the best. Furthermore, these results are also representative of the "Cross Scenario" and "Big Rhombus Scenario", which we include in Appendix A.

From the covering point of view, we identify a better performance of the *MaxCoveringByZones* strategy when the resources are limited, as presented in Figure 5.6 for the Line Scenario. However, the tendency changes when more resources are available since the "Version 3+2" *NearestByDistance* outperforms the other strategies.

To better understand the strategy behavior, we present the results for the average

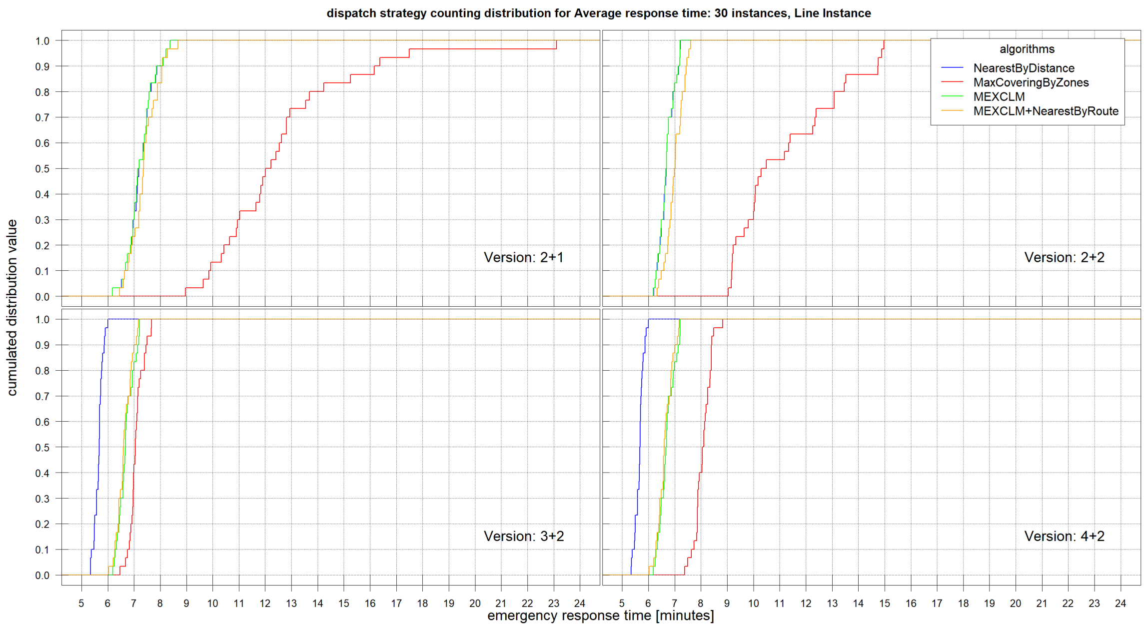


Figure 5.3. Dispatch strategy counting distribution for average emergency response time: Line scenario.

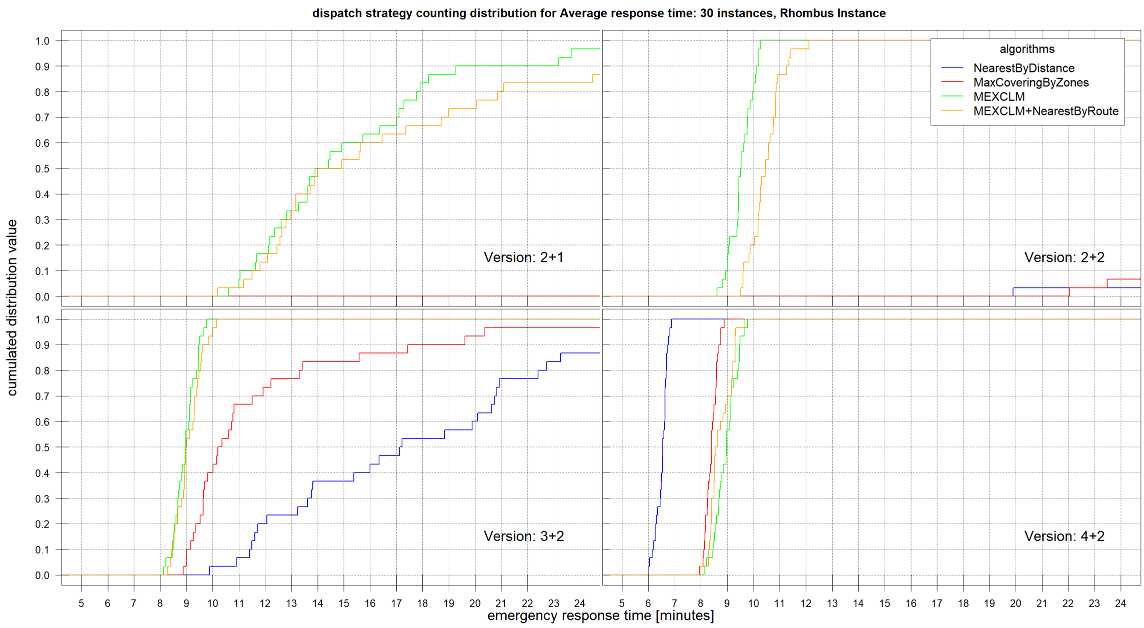


Figure 5.4. Dispatch strategy counting distribution for average emergency response time: Rhombus scenario.

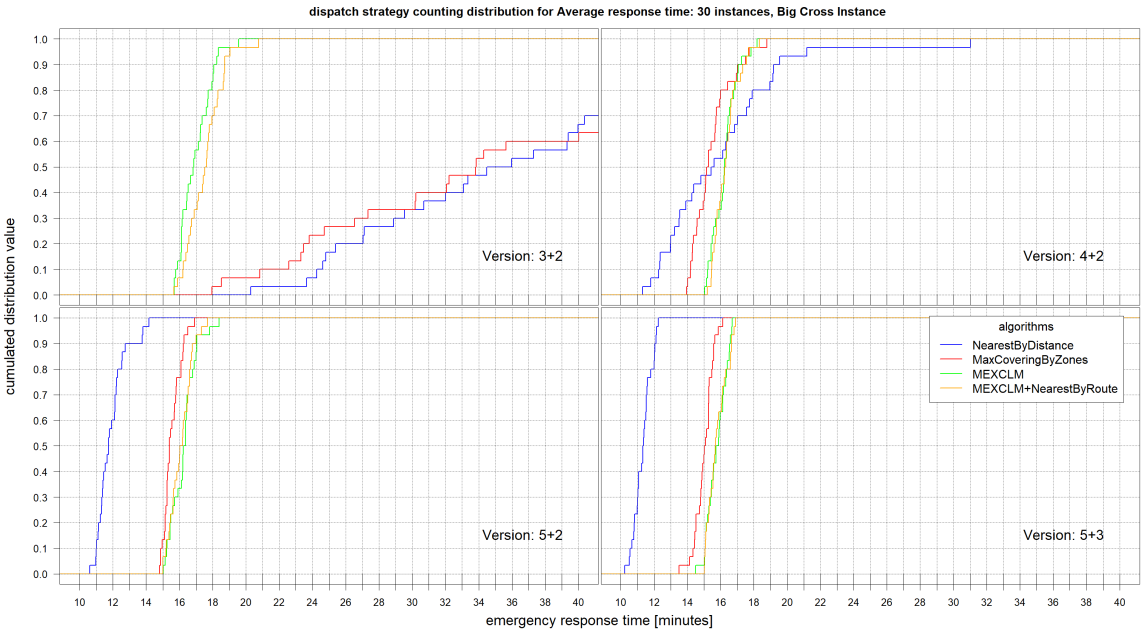


Figure 5.5. Dispatch strategy counting distribution for average emergency response time: Big Cross scenario.

emergency call covering related to the "Cross Scenario" in Figure 5.7 and the "Big Rhombus Scenario" in Figure 5.8. In these scenarios, we identify a clear outperformance of *NearestByDistance* in the first place and *MaxCoveringByZones* in second place over the *MEXCLM*-based strategies. We understand this as a result of the faster response of *NearestByDistance* than the *MEXCLM*-based strategies. In this context, it is a matter of fact that a faster response is based on shorter travel times, which promotes shorter process times for emergency calls. Finally, shorter process times mean that the emergency resources are free sooner, and they have more time to be ready and in position.

However, again we find some interesting characteristics. For instance, it seems that the *MEXCLM*-based strategies are distributed in a smaller difference between the minimum and the maximum value than the non-*MEXCLM*-based strategies. This difference is evident in Figure 5.7, especially in limited resources scenarios, such as "Version 2+1" and "Version 2+2".

These results are almost identical to the scenarios we do not include in detail here but they are included in Appendix A.

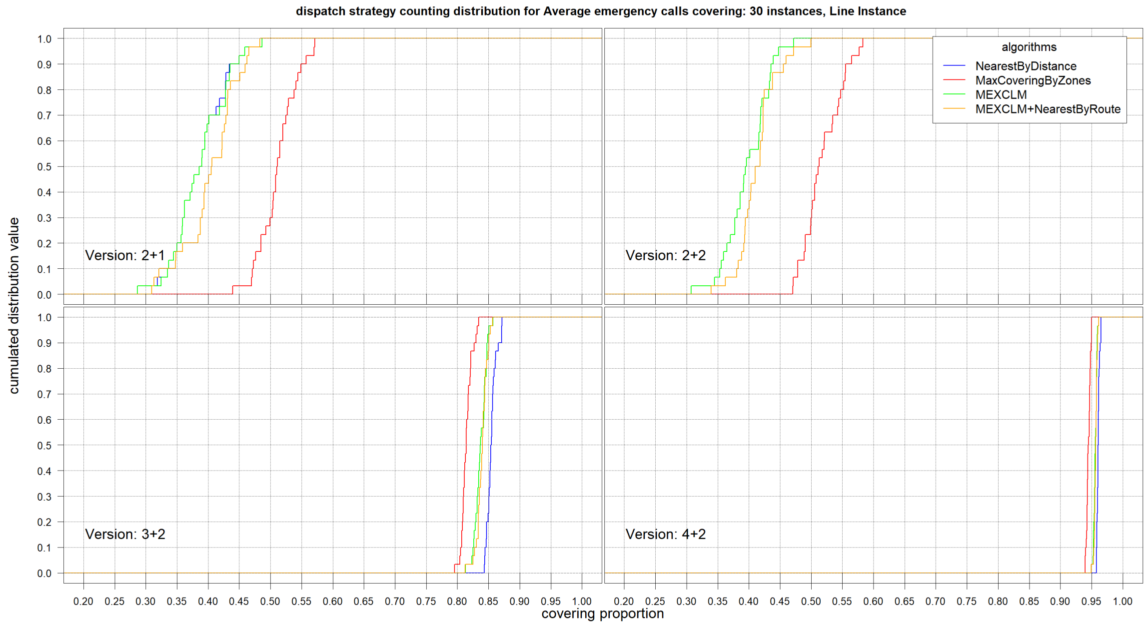


Figure 5.6. Dispatch strategy counting distribution for average emergency calls covering: Line scenario.

After analyzing the results, we can declare that the answer to which dispatch strategy is the best is strongly related to the ambulance capacity and the stress presented in the system. There are clear advantages for the non-*MEXCLM*-based strategies when the system is well planned or oversized, and a quick response is needed. However, these strategies are susceptible to changes in the demand or the capacity of the system. From our point of view, the *MEXCLM*-based strategies are more predictable in terms of performance which means they are more robust than the rest of the strategies.

5.5 Conclusions

In this chapter, we present a Simulation-Optimization Framework for ambulance dispatch algorithms, including a theoretical analysis of some of the most important dispatch strategies.

In this context, we tested *NearestByDistance*, *MaxCoveringByZones*, *MEXCLM*,

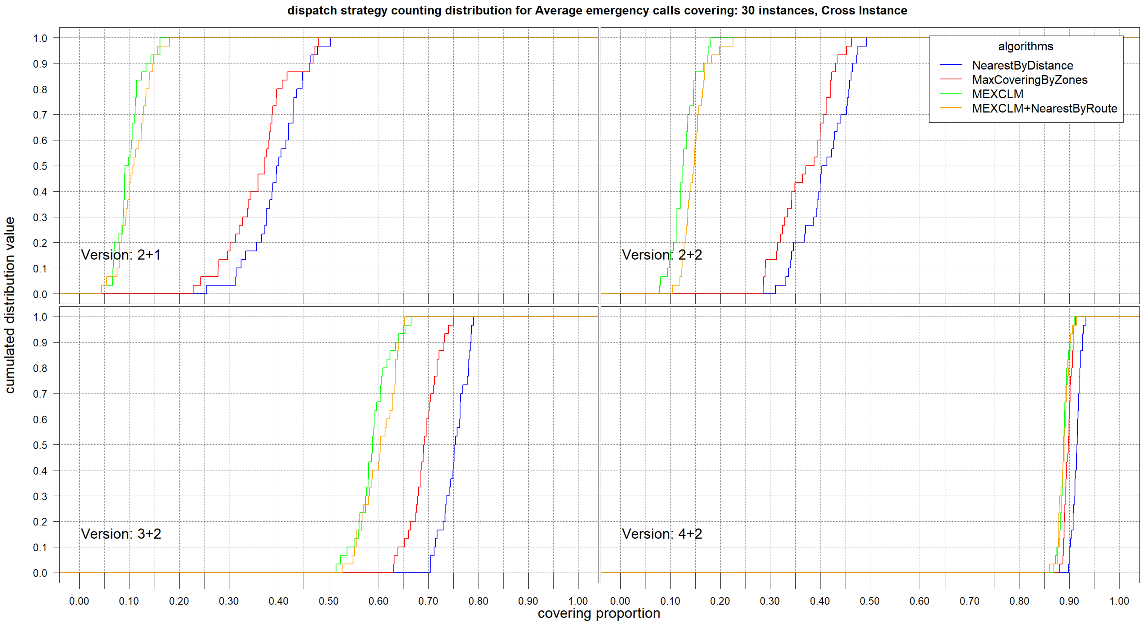


Figure 5.7. Dispatch strategy counting distribution for average emergency calls covering: Cross scenario.

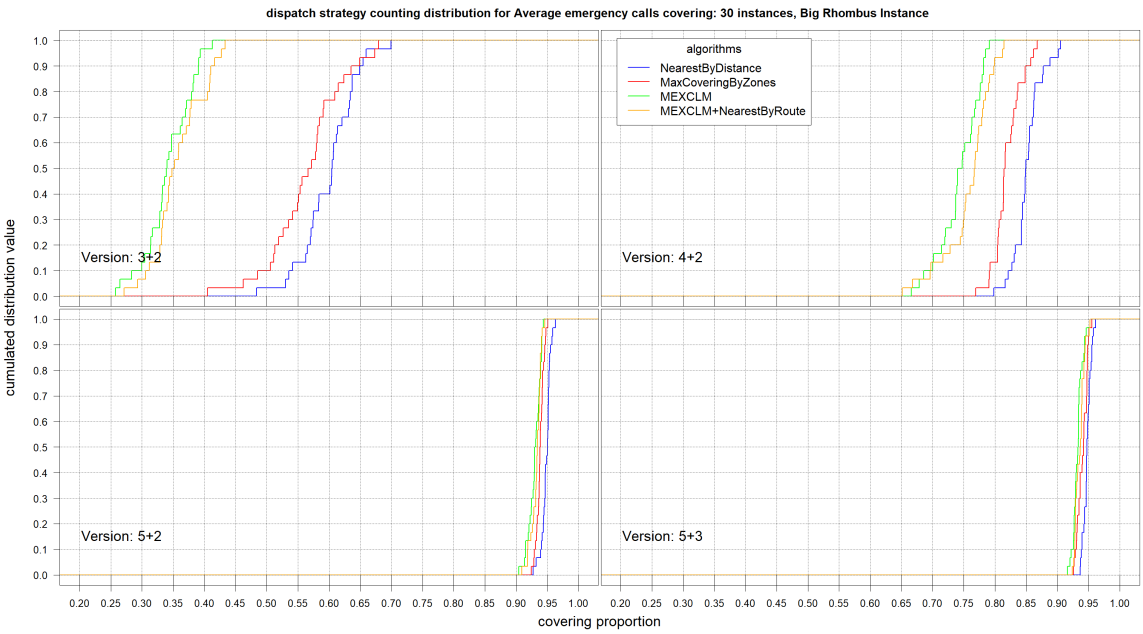


Figure 5.8. Dispatch strategy counting distribution for average emergency calls covering: Big Rhombus scenario.

and *MEXCLM+NearestByDistance*.

The results allow us to establish a relation between stress and dispatch performance of the system since we isolate the performance behavior related to the system's capacity from the other variables, such as traffic, and difficulties on the emergency site, among others. In this context, we find that the relationship between resource capacity vs. demand affects the performance of the dispatch strategies and is critical for concluding which strategy is the best performer. *MEXCLM*-based strategies are more reliable and predictable, maintaining their performance when the system is under stress. We also identify that these strategies are not the best performer when the emergency system is in a relaxing configuration or when the system is not under pressure. Furthermore, we identify configurations where the classical *NearestByDistance* or its "brother", *ClosestAmbulance*, are better in response time. Considering this, it seems to be necessary the detection of the inflection point, which means finding the capacity or the configuration when the non-textitMEXCLM-based dispatch strategy performs better than the *MEXCLM*-based strategies.

However, we consider the *MEXCLM*-based strategies the best all-around performers since they deliver better responses when unpredictable scenarios are taking place.

Finally, we consider that these findings should be confirmed under a realistic scenario with real-world data. Some assumptions and simplifications in this chapter could cause non-representative results.

Chapter 6

A study for dispatch strategies: A simulation-optimization approach for a real world scenario

This chapter addresses the ambulance dispatch problem under a realistic simulation model. Using two curated datasets from emergency regions in Germany, we present two case studies to find which ambulance dispatch strategy is the best. Both regions are very different in terms of resources, orography, and challenges, presenting various scenarios and options in real-world situations.

Part of the content of this chapter will be published in Olave-Rojas and Nickel [2023].

6.1 Literature review

As explained previously in this thesis, life-threatening emergencies require a fast response to have good survival possibilities. These emergencies are taking place in the real world with complex EMS trying to coordinate several resources and attending to multiple emergencies simultaneously. Several techniques for addressing the ambulance dispatch problem exist in the literature, such as offline optimization Andersson and Värbrand [2016], Markov process McLay and Mayorga [2013a,b], among others.

However, our review focuses on articles presenting simulation and optimization models as a significant contribution to facing the ambulance dispatch problem.

In this context, Zarkeshzadeh et al. [2016] propose a hybrid dispatch model considering relocation with generated data. This method includes the first-in-first-out (FIFO) approach, the maximum centrality approach, and the nearest ambulance approach. Although the simulation result shows an improvement in response times, the authors did not include call priorities, road traffic, variable ambulance speed, and different types of ambulances, as pointed out in the work.

Bélanger et al. [2020] propose a recursive simulation-optimization framework to face the ambulance location and dispatching problem. The simulation model performs a dispatching list based on the busy fraction of the ambulances. Using data from Montreal, Canada, the busy fraction is updated after every simulation run through two different methods based on the Bernoulli distribution and the queueing theory. The busy fraction is used in an optimization formulation to determine the relocation position and the dispatching list for the next run. However, the simulation model is simple and does not address the system's complexity since the main simulation model objective seems to be the actualization of the entire region regarding which ambulance is occupied and which is not.

Jagtenberg et al. [2017] ask if the closest-idle policy is always optimal. To answer this, the authors propose first a Markov Decision Process (MDP) and second a heuristic for dynamic ambulance dispatching based on the Maximum Expected Covering Location Problem (MEXCLP) proposed by Daskin [1983]. The authors pointed out that the second approach is tractable for real-world instances. For validation purposes, the authors use a discrete event simulation model representing the region of Utrecht, Netherlands. Nevertheless, the simulation model includes only ambulances with two states: idle and busy, neither re-dispatching nor dispatching during the travel to the base.

Lim et al. [2011] review dynamic ambulance relocation models from the perspective of dispatch policies. The authors declare that deterministic and probabilistic ambulance location models cannot handle the fluctuating demand over time. A dynamic model including dispatch policies overcomes this problem. Furthermore, the authors use a simulation model for testing purposes by implementing a hypothetical region represented by a grid. They also established a fixed ambulance travel speed of 60 km/h and a fixed treatment time of 10 min.

Bandara et al. [2014] study dispatch strategies, including the call's severity, using call priorities to increase patient survivability. The authors propose a dispatch heuristic algorithm to send the closest available ambulance to priority one calls (the most severe) and the less busy ambulance for the priority two calls. This algorithm is tested through a simulation model characterized using data from Hanover County, Virginia, USA. One crucial finding in this work is that priority dispatching policies could improve the performance for urgent calls, meaning a cost in performance for non-urgent calls. Nevertheless, the authors do not consider the realistic possibility of ambulance dispatching when the ambulance is returning to the base since they consider the implementation not easy in systems with limited resources.

McLay and Mayorga [2013a] propose a model for optimal ambulance dispatch considering classification errors in patient priorities. The authors compare an MDP with myopic policies, considering a zero-length queue assumption since the dispatch strategy could depend on the length of the queue, and the problem could be untractable with this technique. The myopic policies presented in this paper are the closest approaches to online optimization models. Although this work does not implement a simulation model, it is the only work considering classification errors to the best of our knowledge.

Sudtachat et al. [2014] study the multiple-unit dispatch to multiple call priorities problem through simulation, optimization, and heuristics. This work aims to maximize the most severe patients' overall expected survival probability. After analyzing the performance of different dispatch policies in small-size instances, the authors propose a heuristic algorithm for real-size problems. One of this work's findings is that considering emergency severity could improve the overall expected survival probability. Implementing a simulation-optimization model, this work is the only one that includes multiple resource types since the model differentiates between advanced life support (ALS) units and basic life support (BLS) units, which should work together in life-threatening accidents. Nevertheless, the authors recognize the complexity of the emergency system and declare as future research the formulation of a more exact model.

Van Buuren et al. [2012] present a Testing Interface For Ambulance Research (TIFAR) simulation tool. The authors claim that one application of the tool is testing different dispatch strategies. This tool includes call priority differentiation, generates calls by postal codes, and allows the dispatching of ambulances when they are returning to the base. However, the authors recognize that the generation call approach does not locate calls in forests, water, or highways. Furthermore, this work includes a redeployment strategy but does not allow the re-dispatch of ambulances.

Aringhieri et al. [2018] present a simulation and online optimization approach to managing ambulances in real-time. The work address three decisions for defining a policy. These decisions are dispatching, routing, and redeployment of ambulances. The authors test several strategies for generating instances based on planar graphs. We also point out that the authors include different average speeds for each arc based on three-speed levels: low, medium, and high. However, the study focuses on urban areas, not on the countryside or the entire emergency department.

Zaffar et al. [2016] present a simulation model based on a seven steps ambulance dispatch process for a comparative study to solve the ambulance location problem. The objective functions are maximizing coverage, survivability and minimizing average response time. The authors' main finding suggests that the better results come with the survivability objective. However, the authors also declare that survivability statistics could be challenging to collect and explain to the policymakers and the general public. The authors also recognize that the limitations could play an essential role in the findings since all demand is assumed to be aggregated to knots in a grid, the ambulances could not attend a call when are returning to the base, and the calls are assumed with an equal priority.

Lanzarone et al. [2018] propose a recursive simulation-optimization approach for the Ambulance Location and Dispatching Problem (ALDP). The simulation model is considered for validation purposes and for improving the optimization model. This work is one of the first attempts to apply this methodology to the ALDP. Nevertheless, the simulation model is simple regarding ambulance process complexity, including fixed travel speed.

Aringhieri et al. [2017b] discuss the possibilities of quantitative analysis based on Health Care Big Data to evaluate dispatching policies for a region to fairly distribute workload after clustering by emergency departments (ED). Using data from Piedmont in Italy, the authors developed a discrete event simulation model, including call prioritization. Furthermore, this is the only work developed under the Franco-German philosophy. Although the authors recognize the need for an analysis of the entire emergency department network, they also declare the need to improve the current model by adding a more detailed representation of the transportation network.

Amorim et al. [2018] analyze two dispatching rules with the closest idle ambulance dispatch policy: the random and the intelligent survival dispatch rule. In order to test the performance, the authors also propose a simulation model based on an algorithm. This algorithm includes two states for ambulances, busy and idle, which cannot represent dispatching when the ambulance returns to the base. A grid represents the spatial area of San Francisco, USA, with nodes for the emergency call positions and hospitals. Since travel times are modeled using data from Google and its Direction API, the authors declare that real-time traffic information is essential in every simulation.

Following the works described in this section, they address the ambulance dispatching problem using a combination of operation research methods. However, there are some concerns in implementing the findings in the real world since they are tested in simulation models that do not represent a real emergency system properly. For instance, none of the works address the interplay between call-takers and ambulances, as the different travel speed profiles presented during a day neither. We consider this characteristic extremely important since speed defines time response and covering under a threshold time. Since most of the response time is used for the ambulance travel, we point out that none of the simulation models include a Geographical Information System (GIS) environment to replicate the travel distances. According to Aringhieri [2010], this kind of simplification could lead to errors and a restricted vision at the moment to evaluate emergency policies.

This chapter proposes a comparison between realistic dispatch policies performed in a simulation model, which is a detailed representation of the complex emergency system. Employing real-world data, the simulation model can interact with the dispatch strategies facing situations such as classification errors, dispatching direct after the patient transport, and re-dispatching if one dispatched ambulance is better located for a life-threatening emergency. A detailed explanation of the methodology for the simulation model is presented in Olave-Rojas and Nickel [2021]. A resume is presented in Chapter 4.

6.2 Dispatch strategies for an ambulance coordination center

Dispatch strategy problems have been well-known and well studied for years. In ambulance dispatching, we concentrate our effort on those that are tractable in terms of computer implementation and less costly in terms of implementation in the real world. In this context, we consider the following ambulance dispatch strategies:

***NearestByDistance*:** This dispatch strategy is the same as presented in Section 5.1.1, i.e., the emergency call will be attended by the nearest ambulance by euclidean distance.

***NearestByRoute*:** This strategy is similar to the previous one. The only difference is the way to define the distance, which is the existing route to the patient.

***MaxCoveringByZones*:** This strategy is explained in detail in Section 5.1.2 and is based on a covering matrix. This matrix is not updated during the experiments.

***MEXCLM:aC*:** This strategy is based on the concepts explained in Section 5.1.2 for the Maximum Expected Covering Location Problem. In this case, we consider for covering purposes not only the unattended calls but also the attended calls that are still in the system. We consider that areas with a high population could lead to a high density of accidents, and consequently, emergency calls should come from these areas more frequently. The idea of including the attended calls is an effort to address this phenom dynamically.

***MEXCLM+NearestByRoute*:** This is one of the two strategies considering a mixed approach, as explained in Section 5.1.2.3. In this version, we consider only the unattended calls for covering purposes.

***MEXCLM+NearestByRoute:aC*:** This strategy is similar to the previous one, but it considers all the existed calls in the system for covering purposes.

6.2.1 Further considerations

There are two characteristics present in all strategies: The idea of re-dispatching and dispatching when the ambulance is returning to the base. Both are present in the real world, according to Hackstein [2019]. Not including these features could lead to errors in analyzing the results since they improve the response of the entire system.

6.3 Cases

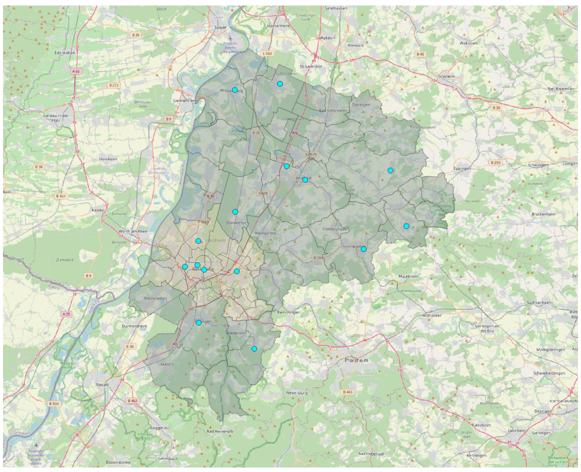
6.3.1 Coordination Center North

As explained in Section 4.5.1, the coordination center North is the Northernmost region in Germany, including a variety of essential characteristics since this region includes urban and rural areas and mainland and islands. Furthermore, this area is located beside the German border with Denmark. We use a curated dataset for this region, including around 270,000 emergency events from 2017 and 2018.

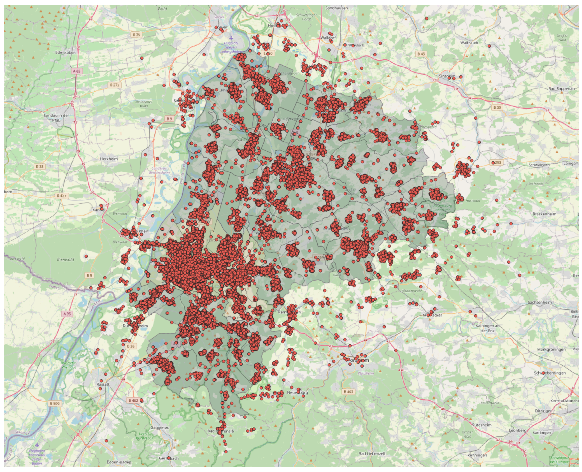
In Figure 6.1, we present the ambulance allocation bases and emergency call localization included in the respective dataset.

6.3.2 Coordination Center Karlsruhe

Karlsruhe's coordination center controls an area, including lands, towns, and cities. The name of this coordination center comes from the second biggest city in the state of Baden-Würtemberg. This area is located in the southwest of Germany beside the Rhine river, one of the mayors of Europe. The coordination center (CC) area is composed of two administrative regions: the city of Karlsruhe and the county of Karlsruhe. These regions include 25 ambulances, seven emergency doctors, and one helicopter, working together with 19 hospitals and covering an area of 1,261.42 [Km^2] and 755,288 Inhabitants Statistisches Landesamt Baden-Württemberg [2021b]. Some characteristics to emphasize are: **(i)** the region concentrates almost 3/4 of the population on the city of Karlsruhe; **(ii)** it is limited to the west by the Rhine river. Hence, the river is a natural barrier to the cooperation between the coordination center and the region on the other side, located in the Rhineland-Palatinate state. In this case, the emergency call dataset is composed of 65,000 emergency events from 2019.



(a) Ambulance allocation across the Coordination Center Karlsruhe



(b) Emergency call localization

Figure 6.1. Ambulance bases and emergency call locations in the Coordination Center Karlsruhe from 2019.

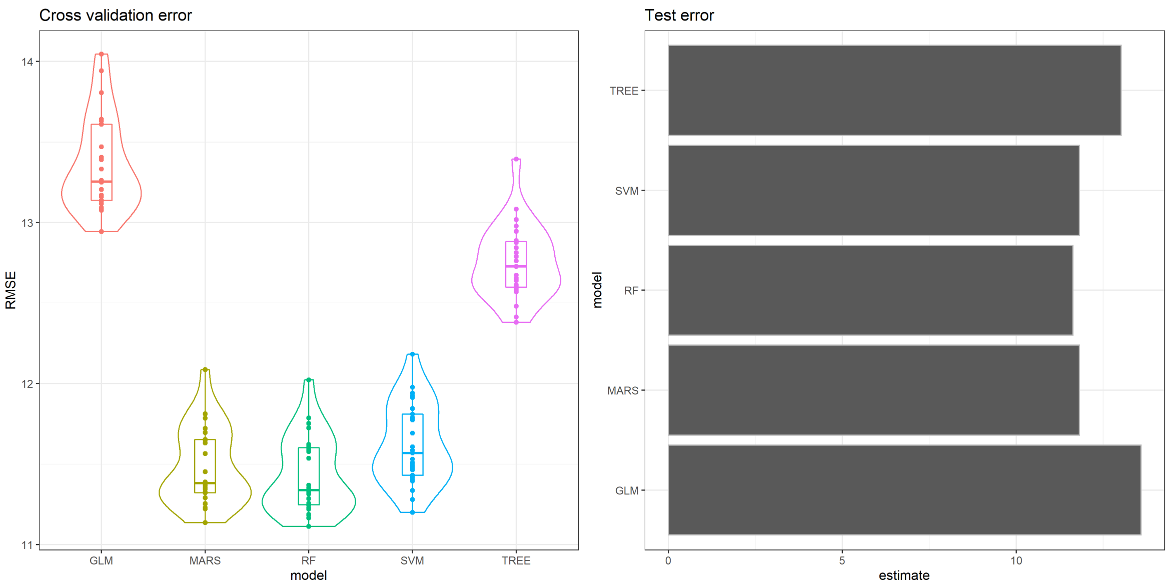
We consider both regions necessary for our study since one region is located in the north and the other is the south of Germany. This geographical difference means the type of emergencies are not the same because weather affects what kind of accidents are presented and how the system can respond to these emergencies. Furthermore, the resource distribution is different in both regions: In the CC North, there is one ambulance for every 7816 Inhabitants; In the CC Karlsruhe, there is one ambulance for every 30211 Inhabitants, which is almost four times more Inhabitants for each ambulance.

Moreover, the area of the CC North is almost four times bigger than the area of the CC Karlsruhe. Finally, in the CC North, there are 72.61 [Km^2] per ambulance compared to CC Karlsruhe, where there are 50.46 Km^2 per ambulance. These differences establish two scenarios that help us specify each strategy's performance more profoundly and detailed.

6.3.3 Simulation parameters

To achieve a proper representation of the real world, we use the methodology proposed by Olave-Rojas and Nickel [2021]. Based on this methodology, the authors developed a simulation model using the data from CC North, which we use for our experiments in this work. Using the methodology of Olave-Rojas and Nickel [2021], we developed the simulation model for CC Karlsruhe. For instance, we use a machine learning approach for the average ambulance speed. Specifically, we use a Random Forest algorithm for travel speed prediction since it has the best test results among five different options, as presented in 6.2 for the CC Karlsruhe.

The parameters are based on curated data to obtain empirical density curves. These data from CC North and CC Karlsruhe are results of the cooperation with the Coordination Center North [Leitstelle Nord, 2021] and the Department of Quality Assurance of Emergency Medical Services in Baden Wuerttemberg [Stelle zur trägerübergreifenden Qualitätssicherung im Rettungsdienst Baden-Württemberg, 2021], respectively. As Validation methodology, we use the same approach as Olave-Rojas and Nickel [2021]: Expert validation, Visual validation, and Data validation.



(a) Cross Validation Error **(b)** Test Error

Figure 6.2. Test results of different machine learning models for travel speed prediction in the Coordination Center Karlsruhe region.

6.3.4 Experiment setup

In order to compare the performance of the dispatch strategies, we use the idea of the counting distribution function defined and developed by Dunke and Nickel [2016]. The authors present this function with the aim of evaluating online algorithms:

$$F_{ALG}(v) = \frac{\sum_{i \in I} \mathbf{1}_{[-\infty, v]}(v_{ALG}(i))}{|I|}, \;\; F_{ALG} : \mathbb{R} \to [0,1] \tag{F.1}$$

Where $v_{ALG}(i)$ is the resulting value of the algorithm ALG on instance i, I is a set of instances, and $\mathbf{1}_A(x)$ is the indicator function (it is 1 if $x \in A$ and 0 otherwise).

We also use the concept of performance ratio proposed by Dunke and Nickel [2016], which establishes a relation between two algorithms:

$$r_{ALG_1, ALG_2}(i) = \frac{ALG_1(i)}{ALG_2(i)} \tag{F.2}$$

The value of $r_{ALG_1,ALG_2}(i)$ in F.2 is greater than 1 when ALG_1 outperforms ALG_2 for the instance i, and it is smaller than 1 if ALG_2 outperforms ALG_1 for the instance i. In the case of both algorithms having the same value, F.2 is equal to 1.

Furthermore, we perform our experiment by running 40 instances. Each instance is composed of one day randomly selected from the curated emergency database for both Coordination Centers.

The simulation time unit is equal to one minute, and the model is implemented on Anylogic™8.2. Finally, we run our experiments in an Intel® Core™ i7-4600U 2.10GHz machine with 8GB RAM and Windows™10.

6.4 Results

6.4.1 Response time

In figure 6.3, we present the counting distributions of each dispatch strategy for average emergency response time. Each column of graphs corresponds to the CC North and the CC Karlsruhe results, respectively. The first row corresponds to non-life-threatening emergency calls, and the second one shows the results for the life-threatening emergency calls. For the calls without worst cases, *NearestByRoute* and *NearestByDistance* are the two dispatch strategies with the best performance, with an advantage for *NearestByRoute*. Furthermore, there is a clear dominance of the strategies based on distance over those based on covering for the average emergency response time since the difference in performance is about 3 minutes. However, for life-threatening emergencies, *MEXCLM+NearestByRoute* is the best strategy in terms of average emergency response time. There are also results in terms of compactness. A slight difference between the best and the worst instance means a more predictable response. Meanwhile, a more significant difference means that the response could not be stable for different scenarios.

For the worst cases in CC North, *NearestByRoute* has a difference of 8.7 minutes between the best and the worst instance. The *MEXCLM+NearestByRoute* presents a difference of 5.9 minutes. The CC North, *MEXCLM+NearestByRoute:aC* shows a difference of 3 minutes. For the same cases, in the CC Karlsruhe, the differences are

3.1, 4.5, and 3.6 minutes, respectively.

MaxCoveringByZones is the strategy with the worst performance, especially in CC North, which involves a more extensive area and a more complicated orography, including islands.

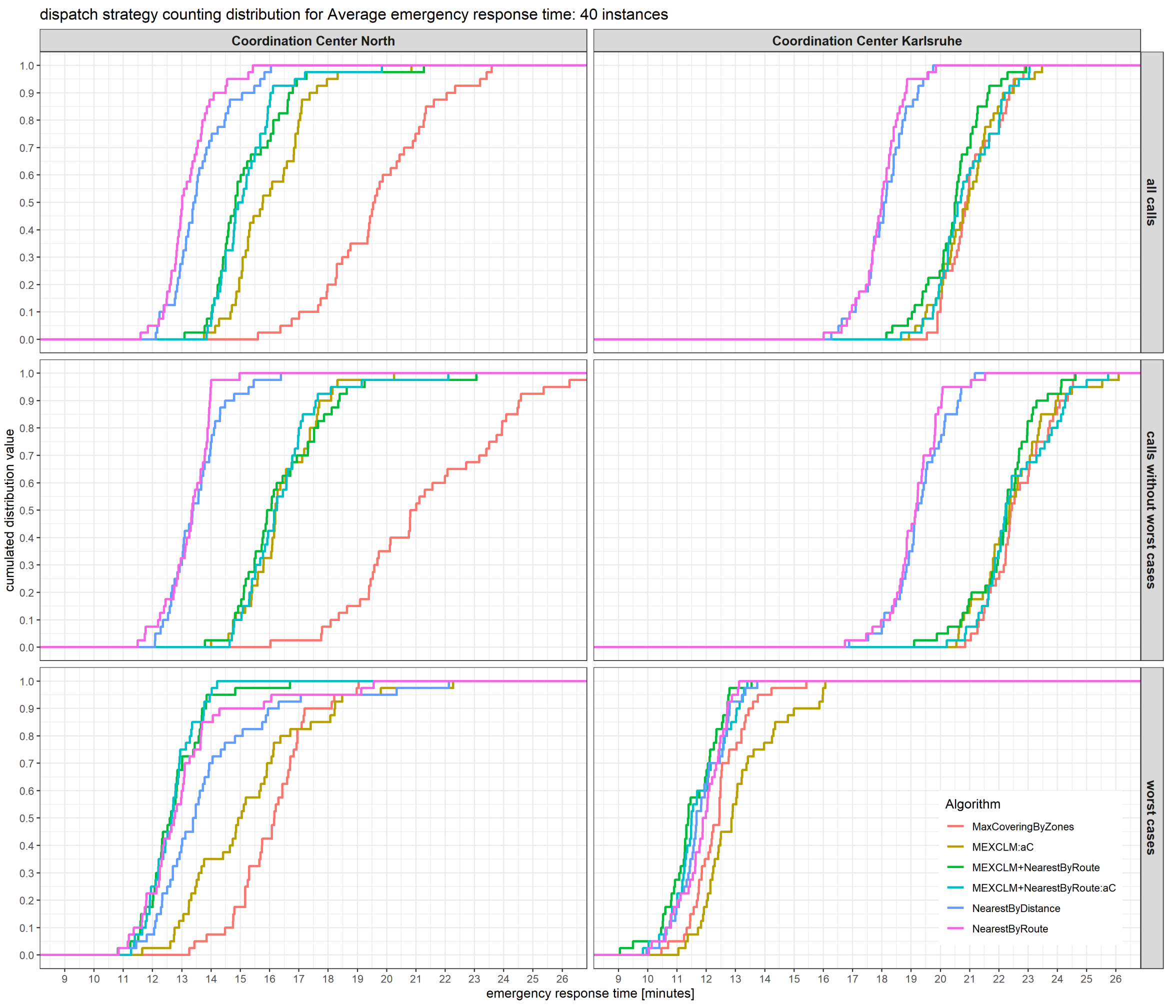


Figure 6.3. Dispatch strategy counting distribution for average emergency response time.

With the aim of a more detailed analysis, we present in Figure 6.4 the performance ratio of all the dispatch strategies or algorithms compared to *NearestByRoute* for both CCs. In this case, we point out that a performance ratio less than 1 means that the selected algorithm for the comparison outperforms the *NearestByRoute* algorithm. In the same way, a performance ratio bigger than 1 means that *NearestByRoute* outperforms the selected algorithm for the one-to-one comparison.

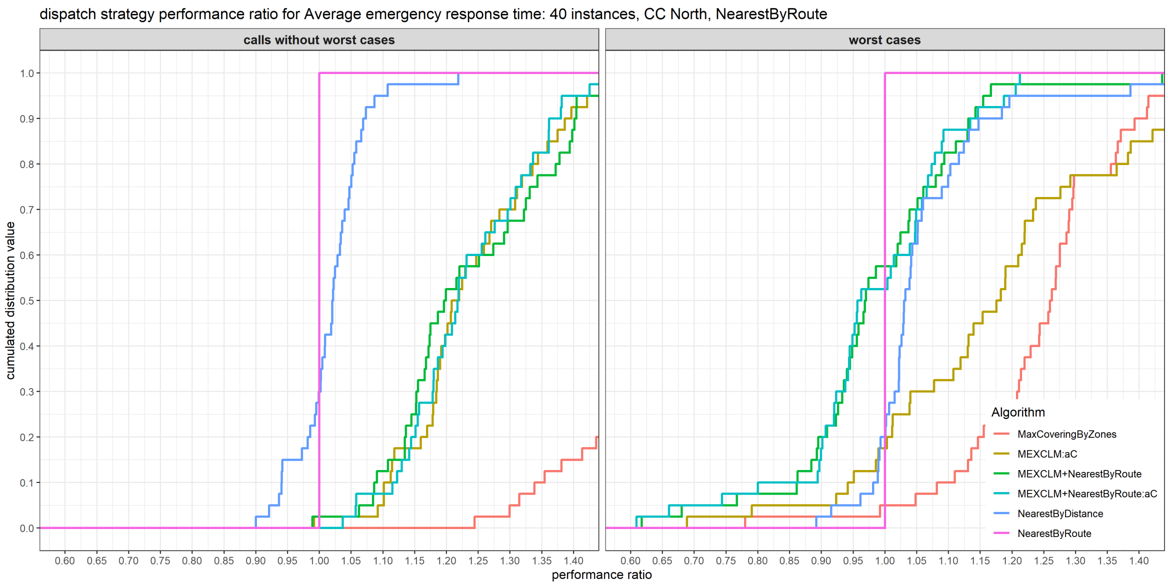
For CC North, *NearestByRoute* is the strategy with the best performance in 70% of the cases (see Figure 6.4a) for non-life-threatening emergencies. For CC Karlsruhe, the result is similar: the best performer is *NearestByRoute* in 60% of the instances for calls without including the worst cases (Figure 6.4b).

However, *MEXCLM+NearestByRoute* outperforms *NearestByRoute* in 57% of the cases for CC North and 70% of the instances for CC Karlsruhe, for the life-threatening cases.

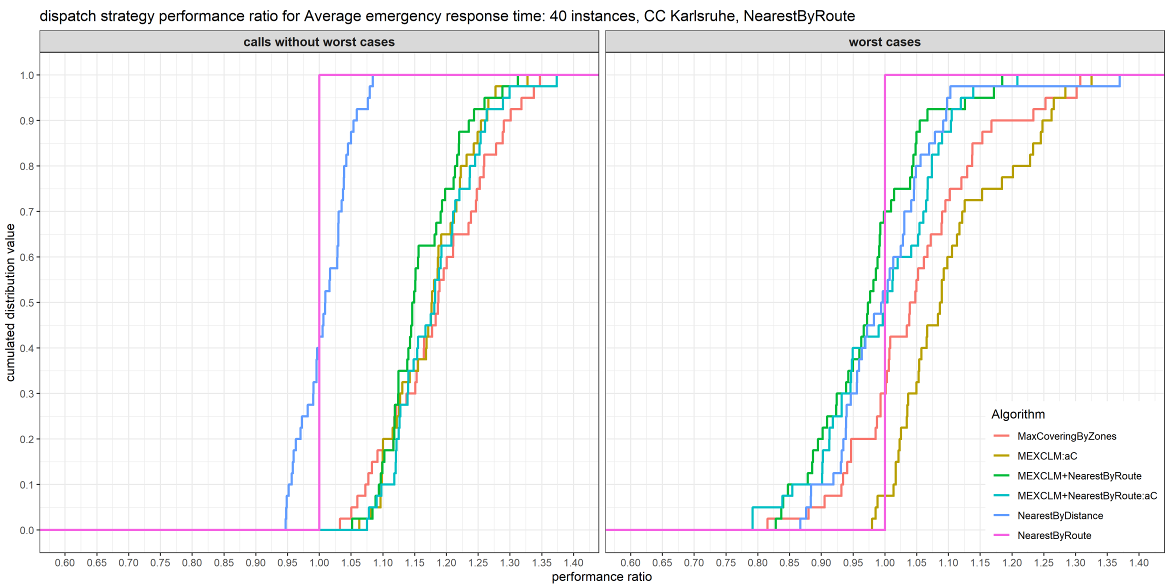
This result presents the idea that *MEXCLM+NearestByRoute* is the best competitor for the classical *NearestByRoute* in performance, where response time is critical. Furthermore, there is a difference of a least 40% between the *MEXCLM* and *MEXCLM+NearestByRoute* in comparison of *NearestByRoute* for life-threatening emergencies. For the emergencies without the worst cases, the difference in performance compared to *NearestByRoute* is similar.

Call Group	Algorithm	Response Time [min]					
		Mean	Min	P25	P50	P75	Max
All	NearestByDistance	18.10	3.55	11.18	15.42	22.80	55.11
	NearestByRoute	**17.97**	3.48	11.18	**15.36**	22.58	**53.10**
	MaxCoveringByZones	**21.04**	4.09	12.71	**18.45**	27.29	59.39
	MEXCLM:aC	20.94	3.94	13.75	17.88	24.68	87.90
	MEXCLM+NearestByRoute	20.47	3.61	12.99	17.60	24.30	85.59
	MEXCLM+NearestByRoute:aC	20.89	3.97	13.33	17.90	24.60	**91.39**
worst cases	NearestByDistance	11.81	4.66	8.78	11.22	13.99	26.43
	NearestByRoute	11.84	5.07	8.98	11.38	13.92	25.01
	MaxCoveringByZones	12.40	5.12	9.32	11.82	14.66	26.21
	MEXCLM:aC	**13.10**	5.23	9.76	**12.26**	15.01	**32.47**
	MEXCLM+NearestByRoute	**11.51**	4.95	8.71	**11.09**	13.51	**23.52**
	MEXCLM+NearestByRoute:aC	11.74	5.09	8.83	11.12	13.62	25.65
not worst cases	NearestByDistance	19.30	3.82	12.11	16.77	24.48	55.11
	NearestByRoute	**19.14**	3.85	12.00	**16.68**	24.02	**53.10**
	MaxCoveringByZones	**22.67**	4.67	14.33	**20.47**	29.22	59.15
	MEXCLM:aC	22.48	5.07	15.27	19.18	26.45	87.43
	MEXCLM+NearestByRoute	22.20	4.06	14.76	19.04	26.21	85.59
	MEXCLM+NearestByRoute:aC	22.61	4.55	15.06	19.20	26.58	**91.39**

Table 6.1. Average values of statistics for the response time over 40 instances of **Coordination Center Karlsruhe**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.



(a) Coordination Center North



(b) Coordination Center Karlsruhe

Figure 6.4. Dispatch strategy performance ratio for average emergency response time: 40 Instances, *NearestByRoute*

Call Group	Algorithm	Response Time [min]					
		Mean	Min	P25	P50	P75	Max
All	NearestByDistance	13.61	1.51	8.80	11.88	16.02	72.77
	NearestByRoute	**13.21**	1.59	8.85	**11.78**	15.88	**52.37**
	MaxCoveringByZones	**19.70**	2.02	11.43	**17.54**	22.88	**234.15**
	MEXCLM:aC	16.04	2.92	10.02	13.62	18.35	93.18
	MEXCLM+NearestByRoute	15.24	1.95	9.47	13.05	17.78	95.12
	MEXCLM+NearestByRoute:aC	15.19	1.89	9.46	12.99	17.67	87.43
worst cases	NearestByDistance	13.88	3.82	8.72	11.56	15.48	60.41
	NearestByRoute	13.08	4.01	8.78	11.38	15.06	44.76
	MaxCoveringByZones	**16.09**	4.58	10.34	**14.74**	20.17	47.88
	MEXCLM:aC	15.25	3.63	9.42	12.58	17.05	**68.98**
	MEXCLM+NearestByRoute	12.72	3.87	8.65	11.45	15.16	35.23
	MEXCLM+NearestByRoute:aC	**12.63**	4.00	8.64	**11.23**	15.06	**33.76**
not worst cases	NearestByDistance	13.50	1.54	8.91	12.14	16.32	53.83
	NearestByRoute	**13.26**	1.59	8.91	**12.01**	16.18	**40.78**
	MaxCoveringByZones	**21.40**	2.08	12.37	**19.06**	24.02	**234.15**
	MEXCLM:aC	16.42	3.38	10.41	14.16	18.94	83.66
	MEXCLM+NearestByRoute	16.43	1.97	10.13	13.98	19.04	95.12
	MEXCLM+NearestByRoute:aC	16.39	1.96	10.12	14.01	19.03	87.43

Table 6.2. Average values of statistics for the response time over 40 instances of **Coordination Center North**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.

In Tables 6.1 and 6.3, we present numerical results for both Coordination Centers. These results are the average value for each indicator presented in the Tables. Furthermore, we highlight the best and the worst values for *Mean*, *P50*, and *Max* indicators. Since the results in these Tables are statistical values for the response time, we understand the smallest value as the best of all strategies for a particular indicator. Subsequently, the worst value is the greatest of them.

In this context, *NearestByRoute* is the best performer when we do not consider the severity of calls. However, mixed strategies (*MEXCLM+NearestByRoute* for CC Karlsruhe and *MEXCLM+NearestByRoute:aC* for CC North) are the best performers when the life-threatening calls are prioritized. Furthermore, we find the best average maximum response time in these strategies (23.52 [min] for CC Karlsruhe and 33.76 [min] for CC North), which confirms that mixed strategies ensure a better system response in those cases when a rapid response is critical. From our point of view, the cost of implementing a mixed strategy is minimal since the worsening of performance for the *not worst cases* is about 3 minutes or less in both regions for at least 75% of the cases.

Sometimes average values could not be representative enough for describing phenomena. In Appendix B, we include Tables B.1 and B.2, which contain the median values (P50) of statistics for the response time. These results confirm the analysis presented in this section for the average values.

6.4.2 Covering

In Figure 6.5, we present the results for the average emergency calls covering. Covering is measured in this work using the following formula based on Daskin [1983]:

$$Cov_i = (1-p)^k \cdot \mathbf{1}_{\tau_{Loc(x),i} \leq T} \tag{Cov.1}$$

Where p is the busy fraction, k is the number of idle ambulances covering the emergency call i. Cov.1 assets the covering of the emergency call from the available ambulances under a threshold time T. In Germany, each state defines by law the time T. and is different according to each state. In CC North the threshold response time is 12 minutes, as presented by the Landesregierung Schlewig-Holstein [2021] and FASP Finck Sigl und Partner [2021a]. This time is measured from the moment the rescue team is alarmed until the rescue team arrives at the emergency place. In CC Karlsruhe, this time should be no more than 10, at most 15 minutes, according to the Statistisches Landesamt Baden-Württemberg [2021a] and FASP Finck Sigl und Partner [2021b]. However, the measure definition is different: from the moment when the call enters into the system until the rescue team is at the emergency place. Although the time T is generally defined according to the patient's severity, in both regions is defined the same T for all cases. Then, these parameters are a benchmark in terms of quality assessment of pre-hospital services for both CCs. This work considers 12 minutes as the threshold time for both regions to compare results. The time is measured between the entry into the system until the rescue team achieves the patient.

For the CC North, *MEXCLM+NearestByRoute:aC* is the strategy with the best performance in terms of covering since this strategy considers the unattended calls and

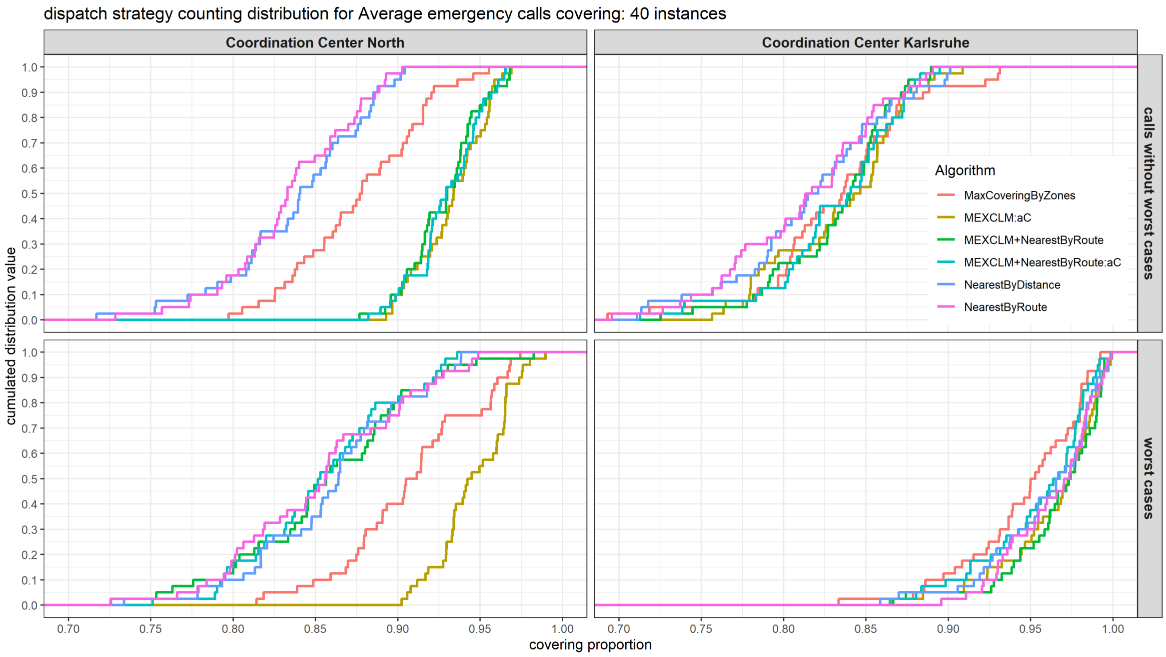


Figure 6.5. Dispatch strategy counting distribution for average emergency calls covering.

the emergencies that were attended to and are still in the system. This dominance is more significant for life-threatening calls. We consider this approach an advantage in terms of covering since we understand that the more inhabitants in a specific area, the more possibility for new emergencies coming from this area. Figures 4.4b and 6.1b support this idea through concentrations of emergency calls related to cities or towns. Hence, considering the attended calls in the system improve the emergency covering, which we consider more related to this context than the area covering.

For CC Karlsruhe, there is no clear differentiation in the strategy performance as in CC North. Nevertheless, we identify a slight dominance of the *MEXCLM*-based strategies.

For a more in-depth analysis, we shall necessarily analyze the results under the findings of Chapter 5 since we are aware that the capacity of the system and the amount of emergency calls plays an essential role in emergency dispatch strategy performance.

Considering this, we understand that CC North and CC Karlsruhe are system' situations that are somewhere between "Version 4+2" and "Version 5+2", both presented in Section 5.4. This conclusion is based on the situation presented for CC Karlsruhe in Figure 6.3: The performance for worst cases is similar between strategies, and they are also compact, as presented in Figure 5.5 for "Version 4+2". This scenario means that the strategy families are in the same performance region, mixed, with slight performance differences but slightly dominance by the *MEXCLM+NearestByRoute* strategy, which combines both families. For calls without worst cases, the situation is similar to "Version 5+2", with a clear dominance of the Distance/Route-based strategies over the Covering-based strategies. For CC North, the situation is similar to CC Karlsruhe. There is a substantial similarity to "Version 5+2" in Figure 5.5 since there is a clear dominance from the Distance/Route-based strategies over the *MEXCLM*-based strategies, for all and calls and the non-worst case calls. We believe the main reason for this is the greater capacity of the system in CC North in comparison with CC Karlsruhe, as we tested in Chapter 5. The performance is significantly similar in CC North and stays almost in the same region for all calls without the worst cases and life-threatening calls. Meanwhile, in CC Karlsruhe, there is a notorious performance difference in each group of calls.

From the covering point of view, it seems to be that the region extension plays an important role in the results. That explains the difference between both Regions for worst cases since the performance of the strategies seems to be similar to "Version 4+2" or "Version 5+2" of Figure 5.8 for CC North, and "Version 5+2" for CC Karlsruhe. Following this idea, we could conclude that CC Karlsruhe has a better capacity to attend the region since the strategy performance is similar to "Version 5+3" in Figure 5.8. Nevertheless, this interpretation does not consider that the simulation model includes re-dispatching, which is clearly present in CC Karlsruhe since there is a difference in performance between calls without worst cases and life-threatening calls. After our analysis, *MEXCLM+NearestByRoute* seems to be the best dispatch strategy for several reasons: **(i)** despite not having the best performance for non-life-threatening emergencies, it is the best performer for the worst cases in terms of response time. We understand that the benefits for life-threatening cases could justify the cost in terms of time for the other calls. **(ii)** the dispatch algorithm performance seems to be quite robust and predictable. There are almost no performance differences in Figure 5.5

and Figure 6.3 for worst cases. We consider this a strong argument because we are in an emergency context, and significant disruptions can occur in the system, which we cannot accurately predict. Reliable and predictable performances are desirable.

Call group	Algorithm	CDF [%] (15 min)	Covering[%]					
			Mean	Min	P25	P50	P75	Max
All	NearestByDistance	69.77	84.53	0.00	88.05	98.94	99.79	100.00
	NearestByRoute	**70.99**	**84.06**	0.00	88.24	**98.92**	99.79	100.00
	MaxCoveringByZones	**39.97**	88.67	0.00	92.16	99.45	99.80	100.00
	MEXCLM:aC	58.76	**93.72**	3.35	96.23	**99.59**	99.80	100.00
	MEXCLM+NearestByRoute	62.09	90.51	1.68	94.35	99.45	99.80	100.00
	MEXCLM+NearestByRoute:aC	62.43	90.55	0.00	94.47	99.45	99.80	100.00
worst cases	NearestByDistance	73.54	86.03	0.00	89.78	99.18	99.88	100.00
	NearestByRoute	74.76	**85.31**	0.00	88.31	98.99	99.88	100.00
	MaxCoveringByZones	**51.21**	90.75	10.05	94.11	99.44	99.92	100.00
	MEXCLM:aC	65.51	**94.62**	22.33	96.96	**99.65**	99.92	100.00
	MEXCLM+NearestByRoute	74.19	85.67	3.90	89.27	**98.96**	99.88	100.00
	MEXCLM+NearestByRoute:aC	**75.16**	85.44	0.00	90.80	99.10	99.88	100.00
not worst cases	NearestByDistance	68.02	83.79	0.00	87.26	99.03	99.91	100.00
	NearestByRoute	**69.23**	**83.44**	0.00	86.52	**98.96**	99.91	100.00
	MaxCoveringByZones	**34.74**	87.68	1.68	90.19	99.32	99.93	100.00
	MEXCLM:aC	55.43	**93.25**	6.70	95.83	**99.62**	99.93	100.00
	MEXCLM+NearestByRoute	56.24	92.82	5.03	95.70	99.59	99.93	100.00
	MEXCLM+NearestByRoute:aC	56.37	93.00	3.35	95.69	99.58	99.93	100.00

Table 6.3. Average values of statistics for covering over 40 instances of **Coordination Center North**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.

In Tables 6.3 and 6.4, we present the average values of statistic indicators for both Coordination Centers. Furthermore, we present the average for the cumulative distribution function call covering under 15 minutes of response time. The results show that dynamic covering outperforms distance-based dispatch strategies for CC North and CC Karlsruhe. However, the dynamic covering does not ensure better system response, but it helps to achieve it for the *worst cases* by penalizing the system response for the *not worst cases*. This effect is indeed presented in Table 6.4 since the best strategy for *not worst cases* under the *CDF* indicator is the worst in covering.

In appendix B, Tables B.3 and B.4 resume the results for the median values to confirm the finding related to average values.

Call Group	Algorithm	CDF [%] (15min)	Covering [%]					
			Mean	Min	P25	P50	P75	Max
All	NearestByDistance	47.47	83.93	0.00	87.94	99.48	99.80	100.00
	NearestByRoute	**48.01**	**83.65**	0.00	87.41	**99.55**	99.80	100.00
	MaxCoveringByZones	35.40	85.05	0.00	89.89	99.46	99.80	100.00
	MEXCLM:aC	**31.86**	**85.55**	0.00	91.34	99.53	99.80	100.00
	MEXCLM+NearestByRoute	35.42	85.38	0.00	89.87	**99.44**	99.80	100.00
	MEXCLM+NearestByRoute:aC	33.46	85.11	0.00	90.06	99.48	99.80	100.00
worst cases	NearestByDistance	82.09	95.83	48.58	98.87	99.86	99.93	100.00
	NearestByRoute	82.78	96.31	48.37	98.55	99.88	99.93	100.00
	MaxCoveringByZones	**77.61**	**94.66**	34.04	98.34	**99.87**	99.93	100.00
	MEXCLM:aC	75.15	96.34	50.79	99.03	99.89	99.93	100.00
	MEXCLM+NearestByRoute	**84.74**	**96.60**	52.12	99.01	**99.90**	99.93	100.00
	MEXCLM+NearestByRoute:aC	83.36	95.46	39.86	98.94	99.88	99.93	100.00
not worst cases	NearestByDistance	40.86	81.62	0.00	82.55	**99.34**	99.92	100.00
	NearestByRoute	**41.41**	**81.24**	0.00	80.24	99.41	99.93	100.00
	MaxCoveringByZones	27.45	83.24	0.00	86.64	**99.51**	99.93	100.00
	MEXCLM:aC	**23.33**	**83.42**	0.00	85.59	99.48	99.93	100.00
	MEXCLM+NearestByRoute	25.84	83.21	0.00	84.99	99.40	99.92	100.00
	MEXCLM+NearestByRoute:aC	24.16	83.16	0.00	87.37	99.38	99.92	100.00

Table 6.4. Average values of statistics for covering over 40 instances of **Coordination Center Karlsruhe**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.

6.4.3 Resource utilization

We define Resource Utilization in this work as the ratio between the amount of time when the resource is busy over the time when the resource is available. This indicator is critical to defining a resource's available probability for serving an emergency. As expressed by Aboueljinane et al. [2013], A low utilization rate is essential to ensure that the rescue team will be available to attend a particular emergency call. Furthermore, around 30-35% is the average utilization rate worldwide, according to Erkut et al. [2008] and Jánošíková et al. [2021]. Above this value, the system could be considered under stress and below this mark, there is free capacity to attend more emergencies than in the covered area.

In Table 6.5, we present the average value of the utilization ratio for the ambulances in the emergency system: In this context, the *NearestByRoute* dispatch strategy is the one that promotes the smallest ambulance utilization, and the *MaxCoveringByZones* strategy promotes the most considerable ambulance utilization.

Distance-based strategies are clearly less stressful for the system since the travel distances are shorter than other strategies. In the same manner, static-covering strate-

Coordination Center	Algorithm	Utilization [%]
North	NearestByDistance	15.98
	NearestByRoute	**15.87**
	MaxCoveringByZones	**17.26**
	MEXCLM:aC	16.08
	MEXCLM+NearestByRoute	15.95
	MEXCLM+NearestByRoute:aC	16.15
Karlsruhe	NearestByDistance	31.44
	NearestByRoute	**31.30**
	MaxCoveringByZones	**32.99**
	MEXCLM:aC	32.31
	MEXCLM+NearestByRoute	32.11
	MEXCLM+NearestByRoute:aC	32.06

Table 6.5. Summary the average utilization for ambulances after 40 instances in each coordination center.

gies (e.g., *MaxCoveringByZones*) do not take into account the distance and changes in the position of ambulances *"in real-time"*, which means that the covered area by an ambulance at the moment of selecting it could be incorrect

6.5 Conclusions and future work

This chapter, we presents a comparison of realistic ambulance dispatch strategies in a real-world context. We implement this through a simulation-online optimization approach. The simulation model is validated and is a real-world representation based on the German emergency medical system. It is essential to point out that the dispatch strategies are implemented under an online optimization approach, which we consider the most realistic approach under real-world conditions.

Furthermore, we applied these strategies to two real case studies (located in the North and the South of Germany, respectively), both characterized by curated data.

We summarize our findings into the following ideas:

- The relation between resource capacity vs. demand affects the performance of the strategies and is critical for concluding which strategy is the best performer.
- Mixed dispatch strategies seem to be better than the "pure" strategies. Our

work shows that the nearest idle ambulance for the life-threatening cases combined with a MEXCLM-based strategy for the non-life-threatening cases performs better than any other strategy. Hence, this strategy performs the rapid response required for the first ones but is supported by the second ones, improving the performance in those cases where time responses are critical.

- Dynamic covering-based dispatch strategies are more stable and robust to changes in demand or capacity compared to static covering-based strategies or the classic nearest idle ambulance. This robustness becomes more critical when disruptions such as terrorist attacks or natural disasters, among others, take place.

However, our findings are analyzed under the comparison with the classical idea of the faster, the better, which is one of the most important quality standards for EMS according to the Department of Quality Assurance of Emergency Medical Services in Baden Wuerttemberg [Stelle zur trägerübergreifenden Qualitätssicherung im Rettungsdienst Baden-Württemberg, 2021]. We consider that performance indicators could be improved to encompass more particularities, for instance, those related to patient wellness, EMS available resources, and orography. We are aware that both Regions have a configuration with one high population density city and countryside. Hence, we consider that a study including a big urban area could present interesting results since there are other parameters such as rush hours.

Furthermore, a comparison between EMS under the two main philosophies, such as the Franco-German and the Anglo-American, could also present results to improve the patients' service quality and survivability.

Chapter 7

Conclusions and future research

Emergency medical systems are complex in such a way that it is hard to enhance the performance of these systems using only optimization models. This context presents challenges related to integrating a number of methods, such as analytics, simulation and optimization research methods, with the aim of achieving a more profound understanding and enhancing these systems.

In this dissertation, we address the emergency ambulance dispatching problem and its complexity from the operational point of view. It includes the following characteristics and elements:

- the establishment of a general emergency medical system (EMS) process;
- the proposition of a hybrid EMS simulation model associated with the EMS process;
- the proposition of a machine learning approach for ambulance travel speed prediction;
- the study and establishment of the interrelation between emergency resources;
- the proposition of an evaluation methodology for ambulance dispatch strategies; and,
- the inclusion of essential characteristics such as orography, re-dispatch (re-routing), patient categorization, and land and urban areas simultaneously.

The items presented in the above list show the complexity of addressing the ambulance dispatch problem and to model it in a proper way, with the aim of obtaining answers applicable to the real world.

By addressing the described list, we determine that the number of ambulances in the system directly affects the performance of a particular dispatch strategy compared to others. It means, that the change in the ambulance availability or the variation of ambulance resource in comparison to the demand, does not affect in the same way the dispatch performance associated with each dispatch strategy. In this context, the covering-based strategies show a more stable and predictable performance in opposition to the classical *"nearest ambulance"* strategy. Nevertheless, we establish the conditions, cases and scenarios where the nearest ambulance strategy is the best performer under the well-known response time indicator.

We also found that the best *all-rounder* approach is a combination of strategies based on which type of emergency call the system should attend. In this context, we believe that emergency categorization is fundamental to correctly evaluate the EMS performance. Is a matter of fact that each patient needs different levels of care, which depends on the emergency severity. Since the modern EMS is prepare for each emergency type by categorizing patients and presenting a *"tailor-made"* response for each of them, we believe that the same approach should be considered for the performance evaluation of these systems.

Furthermore, we verify that several publications exist in the EMS context, but almost none of them are based in the German EMS. We consider this system as one of the most significant worldwide. Hence, we believe that more research based on the German EMS should be done since it is one of the foundations of the Franco-German philosophy. Following this idea, we consider as future research the possibility of testing the results of this work under accurate data coming from one Anglo-American philosophy-based system. Such comparison has not been performed so far in a realistic context, which presents the possibility of the following results and possible research: **(i)** an unified dispatch strategy approach, which works for both philosophies; **(ii)** a real comparison where is possible to define which situations one philosophy is better than the other in; and **(iii)** the definition of an optimal integration level of both philosophies since several EMSs worldwide are a mix between them.

Our hypothesis is that such comparison could lead to confirm the results presented in this work, but it is clear that there exist some challenges in order to perform this comparison property. We recognize at least the following challenges: **(i)** to define a common comparison indicator, in such a way that this should be fair enough for both philosophies; **(ii)** to determine which characteristics should be included for addressing the particularities of both systems, in order to perform the comparison property; **(iii)** to define which emergencies are presented in both systems since they make the comparison plausible; and, **(iv)** to determine the countries where lifestyle and traditions are similar with the aim of do not bias the results.

We also visualize the need for a general emergency demand prediction model. We are aware that exists some research in this field [see Grekousis and Liu, 2019, Martin et al., 2021, as examples of studies addressing this topic]. However, we consider that the applicability of the actual artificial intelligence (AI) techniques, such as neuronal networks, could lead to including characteristics that are not being enclosed so far in realistic approaches, such as dispatch strategies with lookahead, with the aim of establishing a general model for different scenarios. Our best effort in this work is the inclusion of the emergency calls, which are already in the system for the *MEXCLM*-based strategies, as places where the possibility for an emergency call in the future is higher because of population density.

We are also aware that there are challenges in implementing AI techniques since they require a considerable amount of good quality data. The access to data remains as one of the most critical challenges in this area. This point was an important issue to develop and refine the methodology presented in this thesis. In this context the partnership with the Coordination Center North [Leitstelle Nord, 2021] and the SQR-BW [Stelle zur trägerübergreifenden Qualitätssicherung im Rettungsdienst Baden-Württemberg, 2021] was invaluable for this work.

The access to reliable data was also an issue in order to study profoundly the call-taker-dispatcher interaction or the process presented in the coordination center, since in most of the cases, these data does not exist. Coordination centers deal with emergencies, the coordination of services, support activities for the emergency team and hospitals, and if necessary, the coordination with firefighters, police, rescue teams, etc. This presents the need of a deep study and analysis of the process since efficiency is critical in this unit. We believe that the existing methods today could improve the findings presented in Liston et al. [2017], Van Buuren et al. [2015], if there exist timestamps for the entire internal process in the coordination center.

Another important topic is the fact that mean emergency response time is still the most important indicator for evaluating the service quality of an EMS. This occurs not only in the literature but also in the praxis. Since the complexity of EMS is a matter of fact, more key performance indicators are needed, in such a way that they can grasp important characteristics such as survivability, patient satisfaction, fair access to the service, patient and illness categories, among others. Again, one of the main challenges in this topic is reliable data. Some studies modeling survivability for heart attacks exist in the literature, but they are not actualized nor generalized. In this case, these studies should include more diseases and more characteristics as described above in order to define the predictive parameters for these characteristics.

The integration and connection of different methods is an interesting future research topic. This means the development of tool and models based on AI-Simulation-Optimization approaches since the computing capacity allow us to address real size problems. EMS systems' complexity could be addressed by these approaches, promoting more exact and representative solutions. We consider this work as an effort in this line after including a machine learning model for predicting the travel speed with the simulation model and the online optimization models, but we are aware that this is not enough. Following this idea, we believe that the inclusion of neuronal networks presents big possibilities for improving the performance of online optimization approaches. Since neuronal networks are able to take as input several parameters and to establish a relation with the outputs if there these relation exists. A neuronal network could be trained by means of the offline solution, using complete information. Then, the trained neuronal network could solve the online problem, by the pattern recognition in the historical data. Our hypothesis is that this approach could at least outperforms the online approaches presented in the literature, but is complete unclear by how much and how similar these performance to the offline approach is.

An interesting topic is also the system response under stress in the long term. This kind of study presents two advantages: **(i)** they allow us to know the system's limits and which conditions or resources are critical for those variables that are structural and affect the system in the long term, such as climate change; and **(ii)** the policymakers can develop protocols and define the best practices and decisions for accidents, natural disasters, terrorism, among others. We understand that this kind of studies could be difficult to generalize and scale the results to others areas since every region has particularities. Nevertheless, we believe this could be possible after our experience dealing with data from two very different regions.

Finally, we understand the importance of cooperation between academia and practitioners. The bigger the cooperation is, the better the results and integration of operations research techniques into Health Care Systems. This cooperation should be based on understanding processes in EMS and Operation Research methods for both parts.

Bibliography

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.

L. Aboueljinane, E. Sahin, Z. Jemai, and J. Marty. A simulation study to improve the performance of an emergency medical service: application to the french val-de-marne department. *Simulation modelling practice and theory*, 47:46–59, 2014.

S. Al-Shaqsi. Models of international emergency medical service (ems) systems. *Oman medical journal*, 25(4):320, 2010.

J. Amat Rodrigo. Ciencia de Datos, Estadística, Machine Learning y Programación. https://www.cienciadedatos.net/, 2020. Online; visited 2021-06-29.

M. Amorim, S. Ferreira, and A. Couto. Emergency medical service response: analyzing vehicle dispatching rules. *Transportation research record*, 2672(32):10–21, 2018.

T. Andersson and P. Värbrand. Decision support tools for ambulance dispatch and relocation. In *Operational Research for Emergency Planning in Healthcare: Volume 1*, pages 36–51. Springer, 2016.

R. Aringhieri. An integrated DE and AB simulation model for ems management. In *Health Care Management (WHCM), 2010 IEEE Workshop on*, pages 1–6. IEEE, 2010.

R. Aringhieri, G. Carello, and D. Morale. Ambulance location through optimization and simulation: the case of milano urban area. *The 38th annual conference of the Italian operations research society optimization and decision sciences.*, 2007.

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

R. Aringhieri, D. Dell'Anna, D. Duma, and M. Sonnessa. Evaluating the dispatching policies for a regional network of emergency departments exploiting health care big data. In *International Workshop on Machine Learning, Optimization, and Big Data*, pages 549–561. Springer, 2017b.

R. Aringhieri, S. Bocca, L. Casciaro, and D. Duma. A simulation and online optimization approach for the real-time management of ambulances. In *2018 Winter Simulation Conference (WSC)*, pages 2554–2565. IEEE, 2018.

D. Bandara, M. E. Mayorga, and L. A. McLay. Priority dispatching strategies for ems systems. *Journal of the Operational Research Society*, 65(4):572–587, 2014.

J. Banks, J. S. Carson, B. L. Nelson, and D. M. Nicol. *Discrete-event system simulation: Pearson new international edition.* Pearson Higher Ed, 2013.

K. T. Barkley. *The ambulance: the story of emergency transportation of sick and wounded through the centuries.* Load N Go Press, 1990. ISBN 0962635723.

S. Bayer, C. Petsoulas, B. Cox, A. Honeyman, and J. Barlow. Facilitating stroke care planning through simulation modelling. *Health informatics journal*, 16(2):129–143, 2010.

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.

V. Bélanger, E. Lanzarone, V. Nicoletta, A. Ruiz, and P. Soriano. A recursive simulation-optimization framework for the ambulance location and dispatching problem. *European Journal of Operational Research*, 286(2):713–725, 2020.

A. Borodin and R. El-Yaniv. *Online computation and competitive analysis*. cambridge university press, 2005.

N. Bos, M. Krol, C. Veenvliet, and A. Plass. Ambulance care in europe: organization and practices of ambulance services in 14 european countries. 2015.

S. Bubeck. Introduction to online optimization. *Lecture Notes*, 2, 2011.

J. S. Carson. Model verification and validation. In *Proceedings of the winter simulation conference*, volume 1, pages 52–58. IEEE, 2002.

M. W. Carter and C. C. Price. *Operations research: a practical introduction*. Crc Press, 2017.

K. M. Chandy and J. Misra. Distributed simulation: A case study in design and verification of distributed programs. *IEEE Transactions on software engineering*, (5):440–452, 1979.

L. Churilov, A. Fridriksdottir, M. Keshtkaran, I. Mosley, A. Flitman, and H. M. Dewey. Decision support in pre-hospital stroke care operations: a case of using simulation to improve eligibility of acute stroke patients for thrombolysis treatment. *Computers & Operations Research*, 40(9):2208–2218, 2013.

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

Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin e.V. Bad Boller Reanimationsgespräche 2017 – TECHNIK SOLL HELFEN, MEHR LEBEN ZU RETTEN — Deutsches Reanimationsregister. https://www.reanimationsregister.de/aktuelles/1594-veranstaltungen.html, 2017. Online; visited 2020-02-24.

W. F. Dick. Anglo-american vs. franco-german emergency medical services system. *Prehospital and disaster medicine*, 18(1):29–37, 2003.

F. Dunke and S. Nickel. A general modeling approach to online optimization with lookahead. *Omega*, 63:134–153, 2016.

F. Dunke, J. Necil, and S. Nickel. Online-optimierung und simulation in der logistik. In *Zukunftsperspektiven des Operations Research*, pages 33–47. Springer, 2014.

E. Erkut, A. Ingolfsson, and G. Erdoğan. Ambulance location for maximum survival. *Naval Research Logistics (NRL)*, 55(1):42–58, 2008.

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231. AAAI Press, 1996.

European Society for Emergency Medicine. Eusem - What is EM? https://www.eusem.org/about-us/emergency-medicine/what-is-em, 2021. Online; visited 2021-09-15.

FASP Finck Sigl und Partner. Rettungsdienstrecht Schleswig-Holstein, rescuenomics. https://rettungsdienstgesetz.de/schleswig-holstein/, 2021a. Online; visited 2021-09-24.

FASP Finck Sigl und Partner. Rettungsdienstgesetz Baden-Württemberg, rescuenomics. https://rettungsdienstgesetz.de/baden-wuerttemberg/, 2021b. Online; visited 2021-09-24.

A. Fiat and G. J. Woeginger. *Online algorithms: The state of the art*, volume 1442. Springer, 1998.

J. A. Fitzsimmons. A methodology for emergency ambulance deployment. *Management Science*, 19(6):627–636, 1973.

J. W. Forrester. System dynamics—a personal view of the first fifty years. *System Dynamics Review: The Journal of the System Dynamics Society*, 23(2-3):345–358, 2007.

T. A. Granberg and H. T. Nguyen. Simulation based prediction of the near-future emergency medical services system state. In *2018 Winter Simulation Conference (WSC)*, pages 2542–2553. IEEE, 2018.

G. Grekousis and Y. Liu. Where will the next emergency event occur? predicting ambulance demand in emergency medical services using artificial intelligence. *Computers, Environment and Urban Systems*, 76:110–122, 2019.

V. Grimm and S. F. Railsback. Agent-based models in ecology: patterns and alternative theories of adaptive behaviour. In *Agent-based computational modelling*, pages 139–152. Springer, 2006.

M. Grötschel, S. O. Krumke, and J. Rambau. *Online optimization of large scale systems*. Springer Science & Business Media, 2013.

A. Hackstein. Technical information. personal communication, 2019. [112 — Leitstelle NORD].

IBM Cloud Education. What is Machine Learning?—IBM. https://www.ibm.com/cloud/learn/machine-learning, 2020. Online; visited 2021-06-29.

R. G. Ingalls. Introduction to simulation. In *Proceedings of the 2011 winter simulation conference (WSC)*, pages 1374–1388. IEEE, 2011.

C. J. Jagtenberg, S. Bhulai, and R. D. van der Mei. Dynamic ambulance dispatching: is the closest-idle policy always optimal? *Health care management science*, 20(4): 517–531, 2017.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

L. Jánošíková, M. Kvet, P. Jankovič, and L. Gábrišová. An optimization and simulation approach to emergency stations relocation. *Central European Journal of Operations Research*, 27(3):737–758, 2019.

L. Jánošíková, P. Jankovič, M. Kvet, and F. Zajacová. Coverage versus response time objectives in ambulance location. *International Journal of Health Geographics*, 20 (1):1–16, 2021.

M. Karatas, N. Razi, and H. Tozan. Assessing the performance of a sar boat location-allocation plan via simulation. In *Improving the Safety and Efficiency of Emergency Services: Emerging Tools and Technologies for First Responders*, pages 142–178. IGI Global, 2020.

R. M. Karp. On-line algorithms versus off-line algorithms: How much is it worth to know the future? In *IFIP congress (1)*, volume 12, pages 416–429, 1992.

Y. Kergosien, V. Bélanger, P. Soriano, M. Gendreau, and A. Ruiz. A generic and flexible simulation-based analysis tool for EMS management. *International Journal of Production Research*, 53(24):7299–7316, 2015.

P. J. Kiviat. Digital computer simulation: Computer programming languages. Technical report, RAND CORP SANTA MONICA CA, 1969.

J. P. Kleijnen. Verification and validation of simulation models. *European journal of operational research*, 82(1):145–162, 1995.

O. Koch and H. Weigl. Modeling ambulance service of the austrian red cross. In *Winter Simulation Conference*, volume 2, pages 1701–1706, 2003.

W. B. Kouwenhoven, J. R. Jude, and G. G. Knickerbocker. Closed-chest cardiac massage. *Jama*, 173(10):1064–1067, 1960.

M. Kubat. *An introduction to machine learning*. Springer, 2017.

L. Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7):558–565, 1978.

Landesregierung Schlewig-Holstein. Gesetze-Rechtsprechung Schleswig-Holstein SHRDG — Schleswig-Holsteinisches Rettungsdienstgesetz. http://www.gesetze-rechtsprechung.sh.juris.de/jportal/?quelle=jlink&query=RettDG+SH&psml=bsshoprod.psml&max=true, 2021. Online; visited 2021-09-24.

E. Lanzarone, E. Galluccio, V. Bélanger, V. Nicoletta, and A. Ruiz. A recursive optimization-simulation approach for the ambulance location and dispatching problem. In *2018 Winter Simulation Conference (WSC)*, pages 2530–2541. IEEE, 2018.

A. M. Law, W. D. Kelton, and W. D. Kelton. *Simulation modeling and analysis*, volume 3. McGraw-Hill New York, 2000.

A. Lechleuthner. Architecture of emergency medical services in germany. *Notfall+ Rettungsmedizin*, pages 1–8, 2019.

Leitstelle Nord. Service and Kontakt. https://www.leitstelle-nord.de/index.php/ct-menu-item-44, 2021. Online; visited 2021-09-02.

C. S. Lim, R. Mamat, and T. Braunl. Impact of ambulance dispatch policies on performance of emergency medical services. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):624–632, 2011.

P. Liston, J. Byrne, O. Keogh, and P. J. Byrne. Beyond calls: Modeling the connection center. In *2017 Winter Simulation Conference (WSC)*, pages 3804–3815. IEEE, 2017.

M. L. Loper. *Modeling and simulation in the systems engineering life cycle: core concepts and accompanying lectures*. Springer, 2015.

L. Lu and S. Wang. Literature review of analytical models on emergency vehicle service: location, dispatching, routing and preemption control. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 3031–3036. IEEE, 2019.

R. J. Martin, R. Mousavi, and C. Saydam. Predicting emergency medical service call demand: A modern spatiotemporal machine learning approach. *Operations Research for Health Care*, 28:100285, 2021.

R. McCormack and G. Coates. A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. *European Journal of Operational Research*, 247(1):294–309, 2015.

L. A. McLay and M. E. Mayorga. A model for optimally dispatching ambulances to emergency calls with classification errors in patient priorities. *IIE Transactions*, 45 (1):1–24, 2013a.

L. A. McLay and M. E. Mayorga. A dispatching model for server-to-customer systems that balances efficiency and equity. *Manufacturing & Service Operations Management*, 15(2):205–220, 2013b.

G. Mendez-Giraldo, E. López-Santana, and C. Suarez-Roldan. A simulation model for the attention to users in emergency situations in the city of bogotá. In *Workshop on Engineering Applications*, pages 233–245. Springer, 2018.

N. Minar, R. Burkhart, C. Langton, M. Askenazi, et al. The swarm simulation system: A toolkit for building multi-agent simulations. 1996.

T. Mitchell, B. Buchanan, G. DeJong, T. Dietterich, P. Rosenbloom, and A. Waibel. Machine learning. *Annual review of computer science*, 4(1):417–433, 1990.

B. Nelson et al. Foundations and methods of stochastic simulation. *A first course. International series in operations research & management science*, 187, 2013.

D. Olave-Rojas and S. Nickel. Modeling a pre-hospital emergency medical service using hybrid simulation and a machine learning approach. *Simulation Modeling, Practice and Theory*, page 102302, 2021. doi: 10.1016/j.simpat.2021.102302.

D. Olave-Rojas and S. Nickel. The challenge of dispatching the right ambulance: A simulation-online optimization approach for the real world. *Health Care Management Science*, 2023. Submitted.

openrouteservice. openrouteservice. https://openrouteservice.org/, 2020. Online; visited 2020-06-24.

W. H. Organization et al. Emergency medical services systems in the european union: report of an assessment project co-ordinated by the world health organization-data book. Technical report, Copenhagen: WHO Regional Office for Europe, 2008.

J. M. Ortiz. The revolutionary flying ambulance of napoleon's surgeon. *US Army Medical Department Journal*, 4:17–25, 1998.

E. Petitdemange, F. Fontanili, E. Lamine, M. Lauras, and U. Okongwu. A tool-based framework to assess and challenge the responsiveness of emergency call centers. *IEEE Transactions on Engineering Management*, 2019.

L. R. Pinto, P. M. S. Silva, and T. P. Young. A generic method to develop simulation models for ambulance systems. *Simulation Modelling Practice and Theory*, 51:170–183, 2015.

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.

A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229, 1959.

M. J. Sanders, L. M. Lewis, and G. Quick. *Mosby's paramedic textbook*. Jones & Bartlett Publishers, 2012.

R. G. Sargent. Verification and validation of simulation models. *Journal of simulation*, 7(1):12–24, 2013.

V. Schmid. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European journal of operational research*, 219 (3):611–621, 2012.

P. M. S. Silva and L. R. Pinto. Emergency medical systems analysis by simulation and optimization. In *Proceedings of the 2010 winter simulation conference*, pages 2422–2432. IEEE, 2010.

M. Singer and P. Donoso. Assessing an ambulance service with queuing theory. *Computers & operations research*, 35(8):2549–2560, 2008.

P. N. Skandalakis, P. Lainas, O. Zoras, J. E. Skandalakis, and P. Mirilas. "to afford the wounded speedy assistance": Dominique jean larrey and napoleon. *World journal of surgery*, 30(8):1392–1399, 2006.

D. D. Sleator and R. E. Tarjan. Self-adjusting binary search trees. *Journal of the ACM (JACM)*, 32(3):652–686, 1985.

Statistische Amt für Hamburg und Schleswig-Holstein. Zahlen + Fakten - Statistikamt Nord. https://www.statistik-nord.de/zahlen-fakten, 2020. Online; visited 2020-06-24.

Statistisches Landesamt Baden-Württemberg. Gesetz über den Rettungsdienst (Rettungsdienstgesetz - RDG). https://im.baden-wuerttemberg.de/fileadmin/redaktion/m-im/intern/dateien/pdf/Rettungsdienstgesetz_Stand_30122015.pdf, 2021a. Online; visited 2021-09-24.

Statistisches Landesamt Baden-Württemberg. Regionaldaten - Statistisches Landesamt Baden-Württemberg. https://www.statistik-bw.de/SRDB/?R=KR215, 2021b. Online; visited 2021-06-24.

Stelle zur trägerübergreifenden Qualitätssicherung im Rettungsdienst Baden-Württemberg. SQR-BW — Kontakt. https://www.sqrbw.de/kontakt/, 2021. Online; visited 2021-09-02.

K. Sudtachat, M. E. Mayorga, and L. A. McLay. Recommendations for dispatching emergency vehicles under multitiered response via simulation. *International Transactions in Operational Research*, 21(4):581–617, 2014.

C. Swoveland, D. Uyeno, I. Vertinsky, and R. Vickson. A simulation-based methodology for optimization of ambulance service policies. *Socio-Economic Planning Sciences*, 7(6):697–703, 1973.

L. Tavernini. *Continuous-time modeling and simulation: using Turbo Pascal and CTM-S/TP*. Gordon and Breach Science Publishers, Inc., 1996.

T. Ünlüyurt and Y. Tunçer. Estimating the performance of emergency medical service location models via discrete event simulation. *Computers & Industrial Engineering*, 102:467–475, 2016.

M. Van Buuren, R. van der Mei, K. Aardal, and H. Post. Evaluating dynamic dispatch strategies for emergency medical services: Tifar simulation tool. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages 1–12. IEEE, 2012.

M. Van Buuren, G. J. Kommer, R. van der Mei, and S. Bhulai. A simulation model for emergency medical services call centers. In *2015 winter simulation conference (WSC)*, pages 844–855. IEEE, 2015.

M. van Buuren, G. J. Kommer, R. van der Mei, and S. Bhulai. Ems call center models with and without function differentiation: A comparison. *Operations Research for Health Care*, 12:16–28, 2017.

S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, F. N. Delling, et al. Heart disease and stroke statistics-2020 update: A report from the american heart association. *Circulation*, 141(9):e139, 2020.

K. P. White and R. G. Ingalls. The basics of simulation. In *2018 Winter Simulation Conference (WSC)*, pages 147–161. IEEE, 2018.

J. N. Wilford. Cholera Epidemic in New York City in 1832 - The New York Times. https://www.nytimes.com/2008/04/15/science/15chol.html, 2008. Online; visited 2021-09-15.

R. D. Wilmot. *Simulation and evaluation of ambulance systems*. PhD thesis, Georgia Institute of Technology, 1969.

World Health Organization. Cardiovascular deseases. https://www.who.int/health-topics/cardiovascular-diseases, 2020. Online; visited 2020-03-24.

World Health Organization et al. Emergency medical services systems in the european union: report of an assessment project co-ordinated by the world health organization-data book. Technical report, Copenhagen: WHO Regional Office for Europe, 2008.

World Health Organization regional office for Europe. WHO/EUROPE — Cardiovascular deseases - Data and statistics. http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cardiovascular-diseases/data-and-statistics, 2020. Online; visited 2020-03-24.

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.

M. Zarkeshzadeh, H. Zare, Z. Heshmati, and M. Teimouri. A novel hybrid method for improving ambulance dispatching response time through a simulation study. *Simulation Modelling Practice and Theory*, 60:170–184, 2016.

Z.-H. Zhou. *Machine learning*. Springer Nature, 2021.

Appendix A

Graphs corresponding to the results of Chapter 5

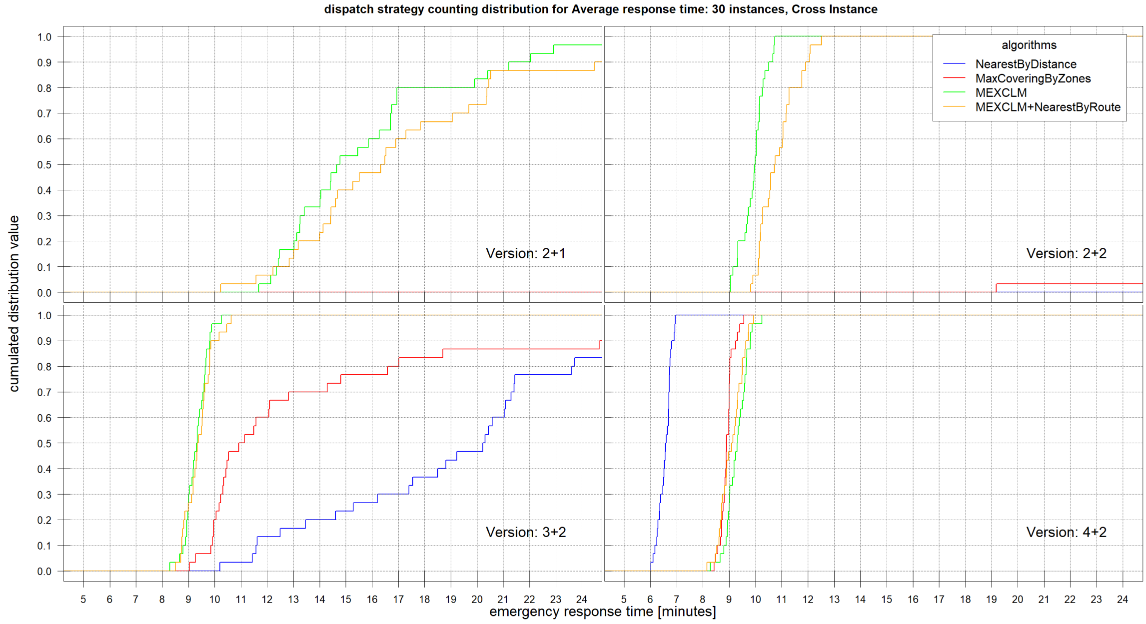


Figure A.1. Dispatch strategy counting distribution for average emergency response time: Cross scenario.

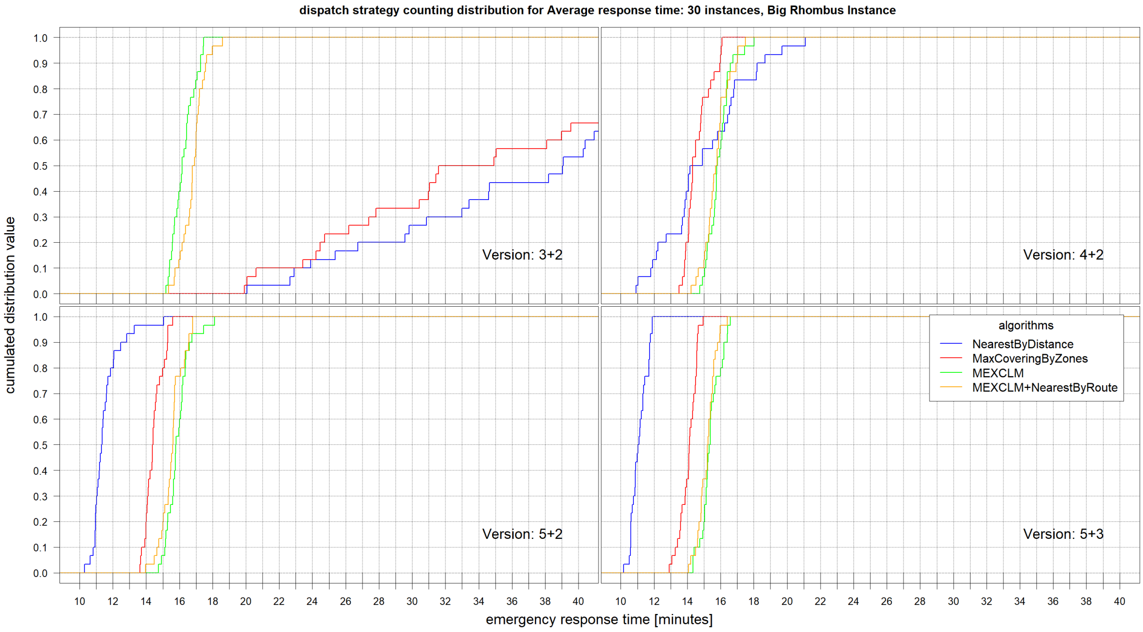


Figure A.2. Dispatch strategy counting distribution for average emergency response time: Big Rhombus scenario.

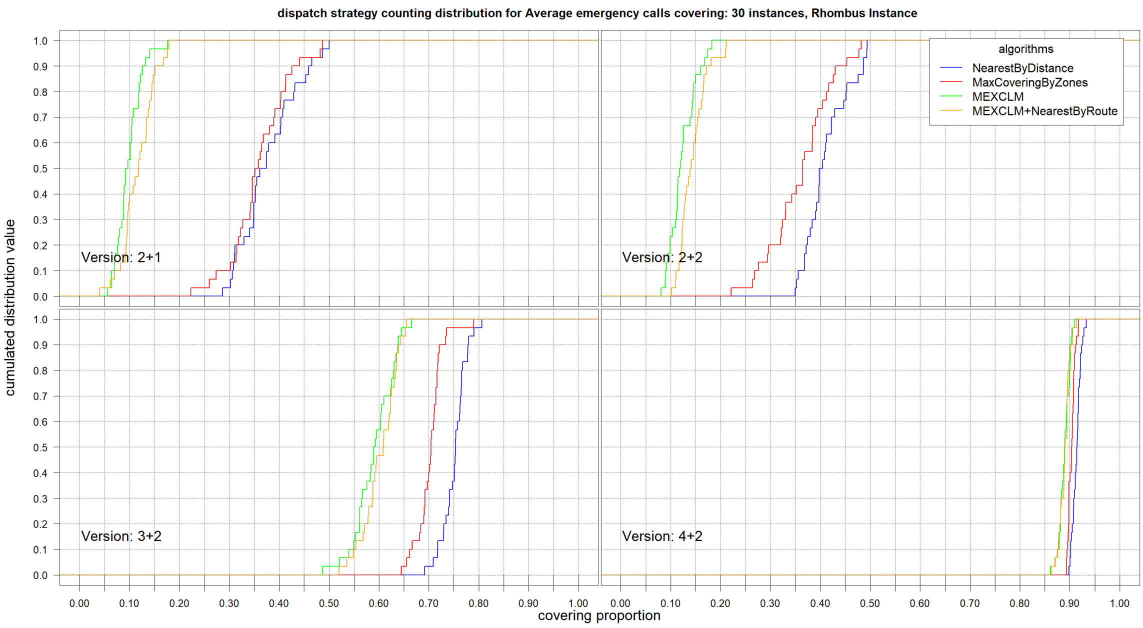


Figure A.3. Dispatch strategy counting distribution for average emergency calls covering: Rhombus scenario.

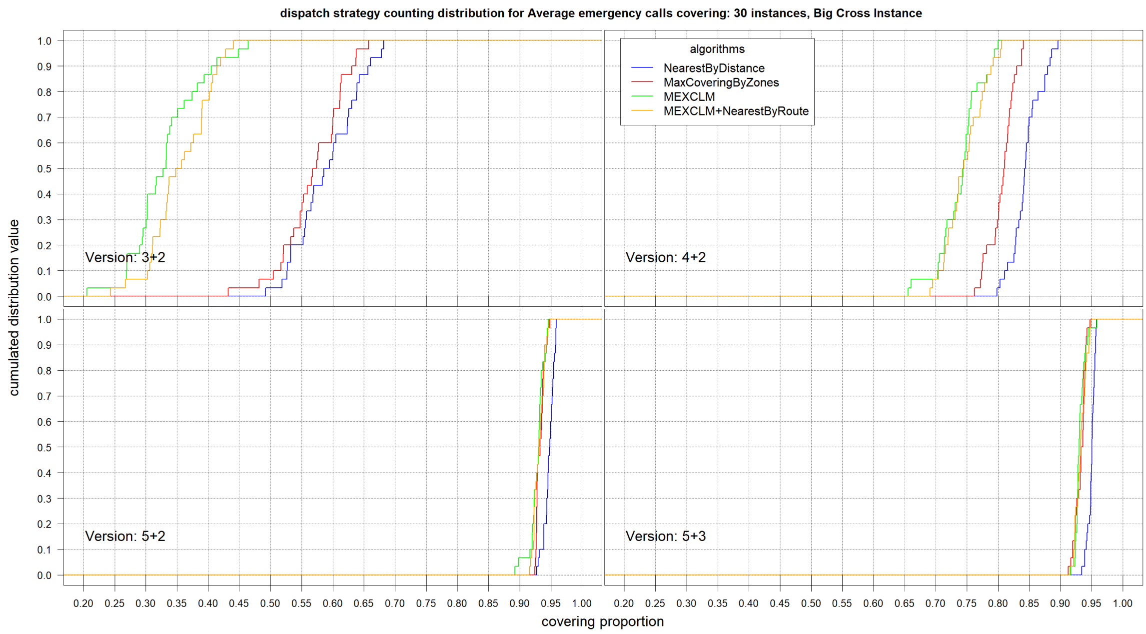


Figure A.4. Dispatch strategy counting distribution for average emergency calls covering: Big Cross scenario.

Appendix B

Tables corresponding to the results of Chapter 6

Call Group	Algorithm	Response Time [min]					
		Mean	Min	P25	P50	P75	Max
All	NearestByDistance	13.43	1.45	8.75	11.85	15.91	60.66
	NearestByRoute	**13.01**	1.50	8.82	**11.79**	15.68	**38.66**
	MaxCoveringByZones	**19.57**	1.92	11.46	**17.66**	22.98	**185.75**
	MEXCLM:aC	15.77	2.86	9.91	13.57	18.12	91.49
	MEXCLM+NearestByRoute	14.84	1.76	9.41	12.98	17.70	86.77
	MEXCLM+NearestByRoute:aC	15.00	1.79	9.38	12.84	17.60	79.41
worst cases	NearestByDistance	13.43	3.82	8.72	11.72	15.12	39.53
	NearestByRoute	12.76	4.07	8.62	11.36	15.25	**32.01**
	MaxCoveringByZones	**16.16**	4.46	10.14	**14.82**	20.70	40.39
	MEXCLM:aC	14.99	3.66	9.25	12.32	16.94	**70.72**
	MEXCLM+NearestByRoute	**12.62**	3.91	8.75	11.43	15.09	32.83
	MEXCLM+NearestByRoute:aC	12.66	3.90	8.65	**11.10**	14.89	32.76
not worst cases	NearestByDistance	13.38	1.45	8.95	12.25	16.24	42.52
	NearestByRoute	**13.35**	1.50	8.88	**12.02**	15.88	**38.02**
	MaxCoveringByZones	**20.91**	1.92	11.95	**19.08**	23.90	**185.75**
	MEXCLM:aC	16.20	3.32	10.14	14.13	18.69	81.17
	MEXCLM+NearestByRoute	15.99	1.76	9.94	14.07	19.18	86.77
	MEXCLM+NearestByRoute:aC	16.21	1.79	9.97	13.97	19.11	79.41

Table B.1. Median values (P50) of statistics for the response time over 40 instances of **Coordination Center North**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.

Call Group	Algorithm	Response Time [min]					
		Mean	Min	P25	P50	P75	Max
All	NearestByDistance	18.11	3.49	11.13	**15.48**	22.71	51.96
	NearestByRoute	**17.99**	3.57	11.19	15.52	22.61	**51.29**
	MaxCoveringByZones	20.88	4.26	12.68	**18.36**	27.23	56.03
	MEXCLM:aC	**20.95**	3.74	13.80	17.89	24.86	82.43
	MEXCLM+NearestByRoute	20.51	3.55	12.98	17.50	24.26	81.46
	MEXCLM+NearestByRoute:aC	20.65	3.79	13.27	17.89	24.26	**91.33**
worst cases	NearestByDistance	11.67	4.77	8.79	11.16	13.74	25.42
	NearestByRoute	11.93	5.10	8.95	11.36	13.80	25.19
	MaxCoveringByZones	12.44	5.18	9.18	11.82	14.63	23.67
	MEXCLM:aC	**12.88**	5.55	9.71	**12.25**	14.73	**28.69**
	MEXCLM+NearestByRoute	**11.38**	4.86	8.72	**10.77**	13.67	**21.54**
	MEXCLM+NearestByRoute:aC	11.50	5.06	8.85	11.10	13.70	24.66
not worst cases	NearestByDistance	19.19	3.75	12.18	**16.67**	24.56	51.96
	NearestByRoute	**19.16**	3.80	11.95	16.71	24.26	**51.29**
	MaxCoveringByZones	**22.42**	4.78	14.23	**20.37**	29.44	56.01
	MEXCLM:aC	22.36	5.34	15.15	19.11	26.37	82.43
	MEXCLM+NearestByRoute	22.25	4.03	14.69	19.05	26.04	81.46
	MEXCLM+NearestByRoute:aC	22.23	4.51	15.22	19.21	26.31	**91.33**

Table B.2. Median values (P50) of statistics for the response time over 40 instances of **Coordination Center Karlsruhe**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.

Call Group	Algorithm	CDF [%]	Covering [%]					
		(15 min)	Mean	Min	P25	P50	P75	Max
All	NearestByDistance	69.61	84.70	0.00	89.00	**98.70**	99.80	100.00
	NearestByRoute	**71.91**	**84.45**	0.00	89.00	**98.70**	99.80	100.00
	MaxCoveringByZones	**39.54**	88.73	0.00	89.00	99.50	99.80	100.00
	MEXCLM:aC	59.35	**94.11**	0.00	96.30	**99.60**	99.80	100.00
	MEXCLM+NearestByRoute	63.04	91.11	0.00	96.30	99.50	99.80	100.00
	MEXCLM+NearestByRoute:aC	63.63	90.82	0.00	96.30	99.50	99.80	100.00
worst cases	NearestByDistance	74.14	86.40	0.00	89.05	**99.55**	99.93	100.00
	NearestByRoute	73.15	85.60	0.00	89.05	99.16	99.93	100.00
	MaxCoveringByZones	**51.54**	90.83	0.00	96.36	**99.55**	99.93	100.00
	MEXCLM:aC	64.70	**94.35**	0.00	96.36	**99.55**	99.93	100.00
	MEXCLM+NearestByRoute	74.83	85.65	0.00	89.05	**98.78**	99.91	100.00
	MEXCLM+NearestByRoute:aC	**75.20**	**85.22**	0.00	89.05	**98.78**	99.93	100.00
not worst cases	NearestByDistance	67.45	84.04	0.00	89.05	**98.78**	99.93	100.00
	NearestByRoute	**68.31**	**83.30**	0.00	89.05	**98.78**	99.93	100.00
	MaxCoveringByZones	**34.78**	87.76	0.00	89.05	**99.55**	99.93	100.00
	MEXCLM:aC	56.45	**93.39**	0.00	96.36	**99.55**	99.93	100.00
	MEXCLM+NearestByRoute	56.38	93.00	0.00	96.36	**99.55**	99.93	100.00
	MEXCLM+NearestByRoute:aC	57.14	92.94	0.00	96.36	**99.55**	99.93	100.00

Table B.3. Median values (P50) of statistics for covering over 40 instances of **Coordination Center North**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.

Call Group	Algorithm	CDF [%] (15 min)	Covering [%] Mean	Min	P25	P50	P75	Max
All	NearestByDistance	47.31	84.37	0.00	89.00	99.70	99.80	100.00
	NearestByRoute	**47.48**	**84.21**	0.00	89.00	99.70	99.80	100.00
	MaxCoveringByZones	35.21	85.58	0.00	89.00	99.70	99.80	100.00
	MEXCLM:aC	**31.55**	**86.59**	0.00	96.30	99.70	99.80	100.00
	MEXCLM+NearestByRoute	35.69	86.12	0.00	89.00	99.70	99.80	100.00
	MEXCLM+NearestByRoute:aC	33.00	86.04	0.00	89.00	99.70	99.80	100.00
worst cases	NearestByDistance	83.13	96.61	67.00	99.55	99.93	99.93	100.00
	NearestByRoute	84.29	97.19	67.00	99.55	99.93	99.93	100.00
	MaxCoveringByZones	76.46	**95.14**	0.00	99.55	99.93	99.93	100.00
	MEXCLM:aC	**76.17**	97.31	67.00	99.55	99.93	99.93	100.00
	MEXCLM+NearestByRoute	**85.24**	**97.38**	67.00	99.55	99.93	99.93	100.00
	MEXCLM+NearestByRoute:aC	84.62	96.56	67.00	99.55	99.93	99.93	100.00
not worst cases	NearestByDistance	40.91	81.79	0.00	89.05	99.83	99.93	100.00
	NearestByRoute	**41.03**	**81.52**	0.00	89.05	99.55	99.93	100.00
	MaxCoveringByZones	27.16	83.59	0.00	89.05	99.83	99.93	100.00
	MEXCLM:aC	**22.99**	**84.45**	0.00	89.05	99.83	99.93	100.00
	MEXCLM+NearestByRoute	26.61	84.03	0.00	89.05	99.55	99.93	100.00
	MEXCLM+NearestByRoute:aC	23.63	83.98	0.00	89.05	99.83	99.93	100.00

Table B.4. Median values (P50) of statistics for covering over 40 instances of **Coordination Center Karlsruhe**. P25, P50 and P75 mean percentile 25, 50 and 75 respectively.

C. Center	Algorithm	Utilization [%] ALS	NEF	Call-Taker
North	NearestByDistance	16.20	**23.25**	14.46
	NearestByRoute	16.24	**22.56**	14.26
	MaxCoveringByZones	**17.38**	22.74	**14.00**
	MEXCLM:aC	**16.08**	23.21	14.07
	MEXCLM+NearestByRoute	16.26	22.67	**14.85**
	MEXCLM+NearestByRoute:aC	16.17	23.05	14.40
Karlsruhe	NearestByDistance	30.84	19.77	8.18
	NearestByRoute	**30.47**	**19.02**	8.09
	MaxCoveringByZones	**32.40**	**20.20**	**8.47**
	MEXCLM:aC	31.70	20.14	8.03
	MEXCLM+NearestByRoute	31.57	19.28	7.87
	MEXCLM+NearestByRoute:aC	31.30	18.34	**7.80**

Table B.5. Summary of the median (P50) utilization for each resource after 40 instances. (ALS: ambulance, NEF: emergency doctor)

List of Figures

List of Tables