U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

MIT Portugal

ANALYZING STRATEGIC AND TACTICAL DECISIONS IN THE URBAN EMERGENCY MEDICAL SERVICE (uEMS) TRANSPORT SYSTEM

DOCTORAL THESIS IN TRANSPORT SYSTEMS

Candidate: Marco Raul Soares Amorim

Supervisor: Professor António Couto

Co-supervisor: Professor Sara Ferreira

May, 2019

DOCTORAL PROGRAM IN TRANSPORT SYSTEMS

Edited by
FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
Tel. +351 22 508 1400
Fax +351 22 508 1440
feup@fe.up.p
http://www.fe.up.pt

Co-funded by

EUROPEAN UNION
European Social Fund

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

POPH
PROGRAMA OPERACIONAL POTENCIAL HUMANO

To Maria Leopoldina,
Rui Fiuza and
Maria Helena

“Somewhere, something incredible is waiting to be known.”
— Carl Sagan

ACKNOWLEDGEMENTS

More than the several years of research that are resumed in this thesis, this was a journey of self-development, learning and making a dream come true. However, alone none of this would have been possible, and this day would just remain as a dream. Several people were part of this journey and helped me to reach where I am today.

Professor Carlos Rodrigues, you were the person that first believed in me and opened the doors that allows me to follow this dream. I am, and will always be grateful for it, for believing in me and for continuously supporting and encouraging me along the way.

To my supervisors Professors António Couto and Sara Ferreira. The two of you were the pillars of my development and of my work. Professor Couto, your knowledge and advice were fundamental to my progress. I learned a lot, and I enjoyed every step of the way. Professor Sara, your guidance, especially during the LIVE project, allowed me to transit from a practical background to the academic demands.

Lobo and Cristina, you were my colleagues from the very start to the very end. You were there always, keen to help and support, or just to distract me from the stressful life of a Ph.D. student.

For all of you, and for Professor Tavares, I am thankful for meeting you and that you became part of my life not only as professors/colleagues but also as friends.

To the University of Porto (UP), in particular to the Faculty of Engineering (FEUP) and the Research Centre for Territory, Transports, and Environment (CITTA), I am thankful for providing me with all the resources and conditions that allowed me to develop my work. Furthermore, I would like to thank the MIT Portugal program in particular for the work of Professor Jorge Pinho de Sousa in directing the Doctoral Program in Transport Systems, and for all the assistance provided by Carla Monteiro. A second thank also goes to the Massachusetts Institute of Technology, to the Center for Transports and Logistics, and the Megacity Logistics Lab, particularly to Professor Edgar Blanco, Professor Eva Ponce and Dr. Matthias Winkenbach for receiving me in Cambridge, sharing your knowledge and supporting me during my stay in Boston and at the MIT.

It was in MIT where I made two new colleagues that also became friends. Sebastian and Joana, countless afternoons and nights discussing transport problems or just distracting ourselves from the long workdays in the hidden corners of Cambridge and Sommerville, thank you for all. Sebastian, I am happy we extended our friendship from Boston to Porto and from Porto to Germany, and that you kept supporting and helping me whenever needed. Pouya and Irene (and the countless people I met there), thank you for all the journeys we been through and for the remaining part of my life.

To all my other colleagues from the *Laboratório de Tráfego* (LAT), a thank you for all the moments we spent. Particularly to Murilo and Pedro for the extra support and friendship during the last months. D. Guilhermina, thank you for all the support during my stay at LAT.

I want to acknowledge the support of FCT (Portuguese national funding agency for science, research, and technology) under the grant PD/BD/52355/2013 during the development of this work. Also, I would like to express gratitude to INEM (Portuguese Institute of Medical Emergency) for providing me with relevant information and data.

However, along the way, scientific support is not enough. The pursuit of a Ph.D. is a path that goes uphill but many times downhill. Moreover, without diminishing the moral support of my professors and colleagues, it was required the moral support of family and friends.

My parents were there for me since I can remember. Always believed in me and made all the necessary sacrifices so that I could pursue my goals and dreams. I do not believe I can ever thank you enough for it. To my sister, and yes, to my brother too, thank you for helping whenever it was required, and for caring.

To the rest of my family, particularly to my closest cousins with whom we shared countless Sundays throughout the years and countless miles throughout USA state roads, thanks for being there, always.

For all my friends I am also thankful for that you have remained in my life even when I had less time to enjoy the best life has to give us. Roberto, Rui, and later Carlos, I know you since I can remember, and I am sure no words are needed to show how much I appreciate your friendship and support. Mario and Fábio, not as an old friendship as the ones before, but you been there since the start of this journey. For your support, I am grateful.

Finally, I could not forget my new colleagues at Fraunhofer IAO, particularly to my team leader Michael and unit leader Florian which, for the last months, continuously shown interest in my work and studies. Thank you.

Abstract

Models for strategic and tactical planning decisions concerning the transport system and *modus operandi* of an Emergency Medical Service (EMS) have been systematically investigated in recent years. In particular, researchers have put significant work on theoretical models and mathematical formulations, e.g. to optimize operational metrics to locate stations or allocate emergency vehicles, as well as on heuristics to solve these formulations. However, the continuous growth in complexity requires solving tools that rely on oversimplification or disregard of real-world conditions, such as the cyclic and dynamic fluctuations in time and space of people and traffic. However, in fact, these key players define two of the most important parameters of any EMS Transport System model: *demand* - which directly correlates with people location, and, *vehicles drivability* - which directly correlates with traffic conditions. These parameters become even more important when serving a dense urban area.

As per the extensive body of research on EMS related decision making when planning for the underlying transport system, the state of the art relies often on classic performance metrics. These classic performance metrics essentially constitute the core of any strategic and tactical optimization model for EMS and focus on operational measures such as average response time or coverage. However, these state-of-the-art models generally lack to incorporate and test for alternative metrics such as *victim survival* or are note taking advantage of the properties of *victim heterogeneity*, for instance when looking at classic vehicle dispatching rules.

In essence, oversimplifications, the neglecting of demand characteristics and the absence of victim survivability in theoretical formulations, would then raise questions about whether what works on paper, performed equally well in practice.

This thesis aims to address these problems by studying strategic and tactical models implemented in a dynamic urban area, specifically focusing on demand and drivability changes of temporal and spatial nature and providing empirical evidence of how each model would perform in the real world. A methodology is designed to feed theoretical models with real-life-data that afterward are further assessed in a simulation to provide empirical evidence. Using simulation rather than a real experiment is justified by the fact that experimenting emergency services in urban areas is prohibitive, on a minor degree due to time and budget constraints, but most importantly due to social, moral and ethical standards.

This thesis demonstrates how the dynamics of urban life require proper consideration within theoretical EMS models. Furthermore, we give evidence that the use of survival functions and the acknowledgement of demand heterogeneity increases the social benefit of EMS. Finally, we support claims in literature that strategic and tactical decisions should be integrated into a unique planning process and that the standard dispatching rule - dispatching the closest vehicle - is in fact not an optimal procedure.

.

Keywords: Emergency Medical Service, Transport System, Strategic decisions, Tactical Decisions, Facility location, Vehicle Dispatching

RESUMO

Modelos de decisão para o planeamento estratégico e tático do sistema de transporte do serviço de emergência medica (SEM) têm vindo a ser sistematicamente investigados nos últimos anos. Em particular, a academia têm-se focado em modelos teoricos e correspondentes formulações matemáticas. Como por exemplo na otimização do desempenho operacional a quando da localização de postos de emergência ou alocação de veículos de emergência, e também em heurísticas para resolver as formulações mais complexas. No entanto, os avanços no desenvolvimento destes modelos, que levou a aumentos da sua complexidade, requer ferramentas para a sua resolução que assentam em simplificações ou não consideram condições reais de operação, e.g. flutuações cíclicas, no tempo e no espaço, das condições da população e tráfego. População e tráfego definem dois dos mais importante parámetros de qualquer modelo do sistema de transporte do SEM. A procura – que se relaciona diretamente com a população, e a acessibilidade dos veículos – que se relaciona diretamente com as condições do tráfego. Estes parametros são ainda mais importantes em ambiente urbano.

O estado da arte, no que toca a modelos de decisão e planeamento do sistema de transporte do SEM, ainda depende em métricas de desempenho clássicas. Estás métricas clássicas constituem o corpo dos modelos de decisões estratégicas e táticas para o SEM e focam-se em medidas operacionais tais como o tempo de resposta média ou a cobertura. No entanto, os atuais modelos encontrados na literatura falham na incorporação de novas métricas tais como a sobrevivencia, ou negligenciam as propriedades heterogênicas das vitimas, como por exemplo a quando do envio de um veículo para respodner a uma emergência.

No geral, simplificações excessivas, a negligencia das características da procura e a ausência da incorporação de funções de sobrevivência nas formulações toericas põe em duvida se os modelos teoricos formulados fazem sentido na prática.

Esta tese procura resolver estes problemas ao investigar modelos estratégicos e táticos a serem implementados em zonas urbanas e oferecendo provas empíricas, com especial foco nas mudanças temporais e espaciais da procura e das condições de tráfego. É proposta uma metodologia que combina dados reais com modelos teóricos que posteriormente são avaliados numa simulação como prova empírica das suas capacidades. O uso da simulação vem substituir experiências em ambiente real pois

estas são impossíveis de se realizar devido a limites de tempo e orçamento, mas principalmente pelas implicações sociais, morais e éticas.

Os resultados presentes nesta tese demonstram como a dinâmica existente numa zona urbana requer considerações específicas nos modelos teoricos de SEM. É mostrado que o uso de funções de sobrevivência e a implementação do carácter heterogénico da procura levam a uma melhor resposta do SEM do ponto de vista social. Por fim são ainda suportadas as mais recentes afirmações na literatura no que diz respeito a integração, num plano único, de decisões estratégicas e táticas, e ao uso de regras clássicas de envio de veículos tais como o envio da unidade mais próxima, onde se demonstra que esta não são o procedimento ótimo.

.

Papavras-chave: Serviço de Emergência Médica, Sistemas de Transporte, Decisões estratégicas, Decisões táticas, Localização de estações, Envio de veículos

CONTENT

1. THESIS FRAMEWORK

1.1. INTRODUCING THE URBAN EMERGENCY MEDICAL SERVICE

1.1.1. KEY TERMS AND SCOPE

Before we elaborate on the motivation, objectives and methodology of this thesis it is important that we first describe our study subject so that the reader can have an overview of how the urban Emergency Medical Service transport system operates and that the better grasps the concepts that we will present afterwards.

We define urban Emergency Medical Service, subsequently called uEMS, as the service that responds to ‘habitual’ medical emergencies thus it can be provided by a single organization. Therefore, disaster services and specific hazard emergencies are out of this thesis scope. Furthermore, we go deeper in this definition and we use the term *urban* not only to define the uEMS as the service that responds to ‘habitual emergencies’ but as an Emergency Medical Service that operates in an urban area.

For this thesis, an *urban area* is a metropolitan area, a city or a block where there is a dense mass of population, visitors and commuters and where a road network with high demand exists. For such areas, we assume the effect of city dynamics.

We define city *dynamics* as an urban area where dynamism exists, and dynamism is described as a force that stimulates changes in short periods, such as hours or days. Therefore, the mass of population, visitors and commuters is not homogenous through the urban area nor throughout the temporal spectrum, although cyclic patterns are assumed. Similarly, the traffic conditions change according to temporal variables, but it is also assumed that a cyclical pattern exists.

When it comes to design and decide, the urban Emergency Medical Service stakeholders plan the transport system at strategic and tactical levels. The strategic level represents decisions that have consequences for the long term, e.g. definition of facility locations to store the EMS resources such as the emergency vehicles. Oppositely, the tactical level comprises of decisions that have consequences for a short term such as the next hours or days of operation, e.g. dispatching of an emergency vehicle.

In this thesis we aim to study the urban factors that mostly influence uEMS transport system decisions both at the strategic and tactical level. The focus of the research is transport system plan when it comes to locating vehicle facilities, allocating vehicles to facilities and dispatching vehicles during response time. For that we investigate different models by comparing them and infer empirical evidence to support our findings.

The next section will dive in the overall emergency response service scheme by focusing on how the service stacks in the different types of emergencies and the respective institutions that respond to them. After, we focus in the Medical Emergency Service and describe how the system responds to a call for aid from the moment it arrives in the national/international emergency number to the moment the service dispatches a vehicle.

1.1.2. THE EMERGENCY SERVICE SYSTEM

Worldwide, any person can request the emergency service (ES) through a unique telephone number. In Europe this number, 112, is part of the Global System for Mobile communications (GMS) standard, thus all the GMS-compatible telephones can dial this number even if the device is blocked or there is no SIM card, depending on the country technology. In some countries outside Europe this number redirects the caller to the national emergency number through the GMS protocol (e.g. to the 911 if it is dialed in the USA). A similar calling system exists in the USA.

In general, when the emergency number is dialed a call center answers it and redirects the caller to the proper emergency department. These can be the police department, the fire department or the emergency medical service (EMS) department. However, in some countries the system is only divided in Police and Fire departments. In this case, the fire department manages the medical emergencies, as displayed in Figure 1.

In terms of resources, either the station facilities, and consequently the medical units and its staff, can belong to the city, regional or national service, thus public, or they are owned and operated privately. This means that there might not exist a direct control of the department over all the emergency medical transport resources, though the EMS responsible department has the authority to request any idle vehicle, be it public or private, to answer to a call for aid. It is a fact that not all EMS systems can design strategic and tactical plans independently. For this work, we will assume that the uEMS is independent and the department owns all available resources.

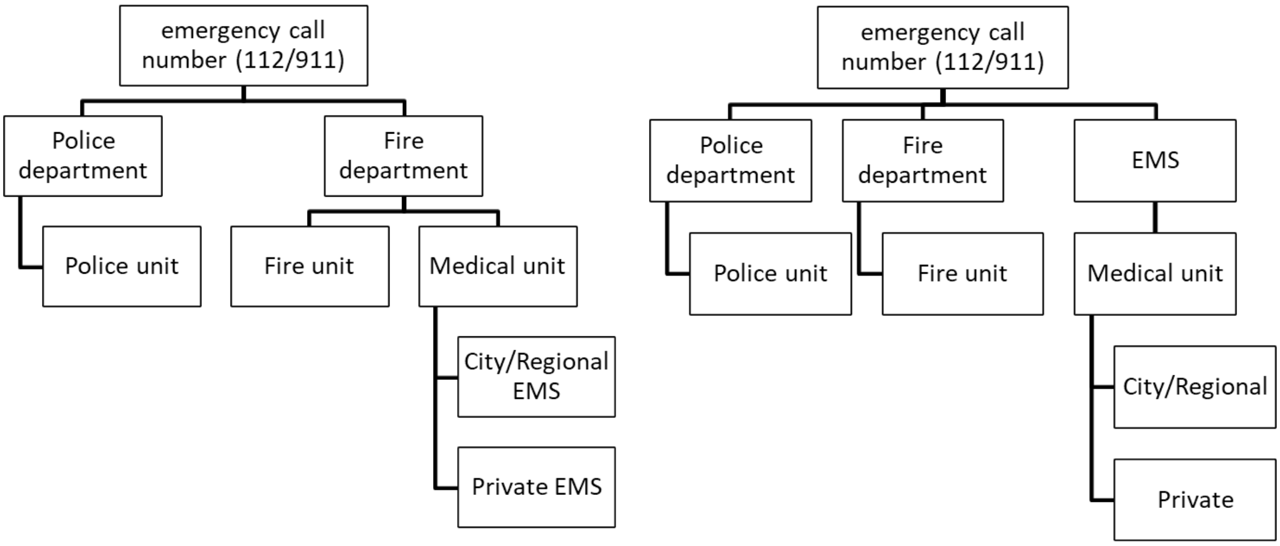


Figure 1. Emergency service structure

1.1.3. The Medical Emergency Service System

At emergency call time, as previously mentioned, the caller is pointed to the proper emergency department. In Portugal, emergency calls are answered by the police force and those that concern medical emergencies are forwarded to one of the “Urgent Patient Guidance Centers” (CODU) of the “Institute of Medical Emergencies of Portugal” (INEM).

An operator, professionally trained, processes the call using a software that consists of a predefined inquiry where a sequence of questions is asked to the caller. The answers to these questions feed a background algorithm that is responsible to assess the medical emergency and activate a request for the proper vehicle when required. The use of such algorithm allows for an unbiased assessment of the medical emergency through the use of quantitative metrics.

During the previous process, the operator instructs the caller for possible assistance techniques to delay the worsening of the victim’s status. In the meantime, if the algorithm flags an emergency priority (i.e. the victim needs urgent professional medical assistance) a request is sent to another operator who has access to the available response vehicles and can query the system for their position and availability. However, if the vehicles are not equipped with GPS and transmitting their coordinates in real-time, the operator only has access to the location of idle vehicles (assuming they are at the original facility). Figure 2 resumes the call processing and decision sequence.

The dispatching rule of INEM consists of the dispatch of the closest idle vehicle. This is a common rule worldwide spread and at practice. Furthermore, another relevant operational characteristic of most

EMS transport system is that at dispatching time, the closest idle vehicle is assessed using static metrics i.e. traffic conditions are usually not considered.

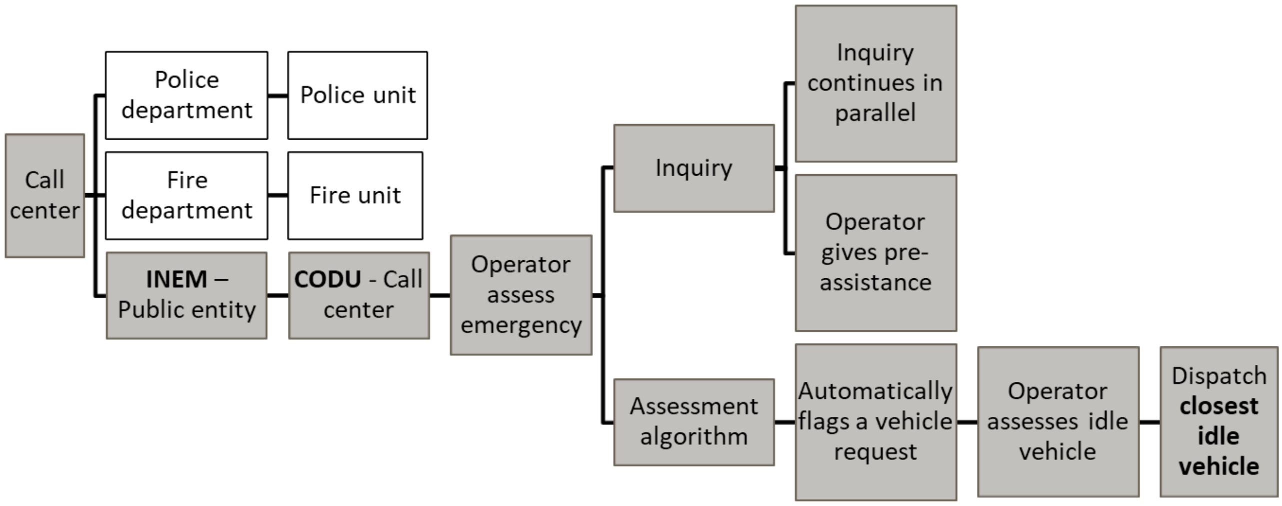


Figure 2. The Portuguese example on how a uEMS call is processed.

1.2. MOTIVATION AND RESEARCH DEVELOPMENT

1.2.1. GENERALITIES

This thesis is a ramification of the project tooLs for Injury preVEntion (LIVE)[1] (Amorim et al., 2014d, Ferreira and Amorim, 2014) which proposed several tools to assess and quantify road crash outcomes. These assessment tools were built to support further analysis when it comes to road crash injury severity reduction. A fundamental aspect of the project outcomes was to build an initial view on how improvements in the emergency medical service transport system could minimize or reduce road crash victims’ severity.

This view came in line with the World Health Organization (WHO) global plan for road safety for the decade of actions of 2011 – 2020. In this plan, WHO defines several pillars where post-crash response is pillar 5 and includes one activity that explicitly encourages the research community to improve the post-crash response.

Nevertheless, one can assume that a system that is public and aims the public health should not prioritize certain events without proof that such policy is advantageous and will not degrade the service

[1] https://citta.fe.up.pt/projects/4-5-live-tools-to-injury-prevention.
Final report available at https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/projects/live.pdf

performance to other events. This brings another interesting dilemma: what is the best way to measure such service performance? Should we keep it operational oriented, or should a public health service focus on the victims' outcomes thus its performance be measured accordingly? Moreover, after defining a metric that is able to assess accurately the uEMS performance, how can we be sure that the theoretical models used at decision time correctly translate each of the stakeholders' objectives?

This sequence of questions motivated our research and defined its developed. The next sections will further detail on it.

1.2.2. Departure Point

The last step of the LIVE project focused on how the uEMS could integrate road safety research, particularly how road safety assessment could give valuable input when designing or planning EMS strategic and tactical decisions.

Road safety assessments are highly applied and carried out mostly to help reduce the number or the severity resulting from a road crash (Elvik et al., 2009). In developed countries the standards for road safety reached such high levels that, more and more, new measures have a less visible impact, i.e. road characteristics, safety policies, vehicles safety and drivers education are so far developed that most of road crashes result from unavoidable actions or specific situations because no system is perfect. It is important to remember that we assume as of now, the safety improvements reachable through the implementation of automation in the driving process, are not yet feasible.

Worldwide, the number of people killed in road crashes each year is estimated at almost 1.24 million, and between 20 and 50 million of people sustain non-fatal injuries (WHO, 2013). However, more than 90% of road crash deaths occur in low- and middle-income countries which have just around 50% of the worldwide registered vehicles (WHO, 2013).

The departure hypothesis of this thesis states that although road safety measures are usually taken directly in the source, which makes all the sense, it is possible to reduce road crash outcomes if improvements are made *a posteriori*, i.e. improving the emergency medical service transport system. This is also the view of the World Health Organization and many road safety researchers (WHO, 2013, Sánchez-Mangas et al., 2010).

The initial approach intended to address the Emergency Medical Service, EMS, with focus on Urban Service, uEMS, and investigate how to support strategic and tactical decisions, with focus in the decisions related to the emergency vehicles and their stations. The subject falls mainly in the Operational Research framework of *facility location*, and *vehicle allocation* and *dispatching* however, it bonds interest with *road safety* and *demand modeling* topics.

Plants, distribution centers and other facilities that have an influence radius of action, and need to supply a certain demand within it, are operational for several years, thus subject to substantial temporal and physical changes of the environment they will settle in. These classical facility location problems have usually highly uncertain costs, demands, travel times as well as other, hard to correctly measure inputs. Therefore, these types of problems require, many times, decision-making tools to deal with these uncertainties or else there will be the risk of underestimate or overestimate their design, which on the other hand can return a negative impact, mainly monetary.

uEMS belong to the family of facility location problems; however, there are particularities that draft them out from the common problems. While under- or overestimation of the earlier mentioned facility location problems might have mostly a monetary impact, in uEMS problems there is also a social impact, particularly a bad decision can lead to e.g. higher response times which may seriously reduce the survival probability of the victims to be rescued. For instance, Sánchez-Mangas et al. (2010) indicated that a reduction of 10 minutes in the emergency response time could result in a 30% reduction of traffic accident fatalities. Although this number can vary depending on many factors, one can assume that a quicker medical response will result in an improved medical assistance – considering that the medical team can give a “better” assistance the “more recent” the occurrence is. Not only with road crashes, as per the presented reference, but also in any other case of medical emergency, a faster arrival at the calling site will always result in a better assessment of the problem resulting in an earlier engagement hence in most cases providing better results. However, the rate at which survival degrades is strongly associated with the type of medical emergency, thus medical emergencies have heterogeneous characteristics.

Two concepts are derived from the above statements - the demand for EMS is heterogeneous and its satisfaction is not solely operational, i.e. reduce average response times might not be optimal in the victims’ perspective; the outcomes of an EMS response is tied to the emergency type and different medical emergencies require different types of assistance which in turn lead to different success rates. To simplify, victims’ survival probability is the target performance to satisfy EMS demand.

1.2.3. Standard Tools from Literature to Assess uEMS improvements

The concepts of demand heterogeneity and victims’ survival were applied at the start of our research development motivated by uEMS strategic models that focus on road crashes (Kepaptsoglou et al., 2012). To achieve this goal, first, it was necessary to study the impact of the EMS response on road crash victims.

The literature[2] shows a lot of controversy when it comes to measure survival rates of road crash victims. In fact, the only well-defined survival functions are those that concern cardiac-arrest victims. Nevertheless, project LIVE allowed for a detailed database that comprise of hospital and police road crash records. These two databases, when combined, could describe a road crash victim in terms of

[2] The reader may at this point kindly refer to chapter 2 for an in-depth state of the art.

demographics, crash location and characteristics, and detailed injury information such as the length of stay in the hospital, the international classification of diseases and detailed severity and costs.

The development of this research started with a linkage methodology to allow connecting police and hospital road crash related datasets and with it build the necessary database to study pre-hospital time impact on road crash victims (Amorim et al., 2014c). Further development of the core ideas was developed in Amorim et al. (2014a).

With such rich database several modeling approaches were possible and insightful scientific output positively contributed to the state of the art when it comes to road safety assessment and road crash social impacts (Ferreira et al., 2015, Couto et al., 2016). Furthermore, with the study of different injury scales, it was possible to select an injury classification, easy to implement but also with the essential detail to build insightful performance metrics (Ferreira et al., 2016, Ferreira et al., 2018).

Finally, with the necessary tools at hand, we ramified our research into the study of the impact of pre-hospital time on road crash victims. The goal was to define a metric system that would equate with the existing survival functions for cardiac arrest emergencies. A study was conducted after an in-depth literature review and several modeling trials to produce the necessary input (Ferreira et al., 2019).

However, every effort made, lead to similar conclusions: such survival functions are usually not accurate for road crash injuries because these have a wide range of injuries types and such information does not arrive to the EMS department at call time. Moreover, road crash emergencies are complex situations where many times victims are at a difficult access point, which require previous intervention of other rescue teams. This complexity in addition to the wide range of possible types of injuries makes survival modeling very hard and inaccurate, invalidating its use to our objectives.

1.2.4. From a Road Safety Focus to a Generalized Approach

Road safety is a constant concern worldwide and in developed countries, authorities and agencies already pointed out that a way to reduce the outcomes of road crash victims is to improve the EMS response. Little work has been developed in this area, therefore we ask: is it possible for road safety authorities to support the EMS and have a direct benefit from it?

We saw, that to quantitatively assess performance gains (in the context presented similar to survival rates) when responding to road crash injuries it is necessary to have a robust metric that can consistently give us unbiased survival rates for each victim. However, in the previous section we demonstrated our effort and concluded that these kinds of metrics are hard to define when it comes to road crash victims.

Following the works of Kepaptsoglou et al. (2012) we developed a method to assess how road safety investments in the EMS could help reduce road crashes outcomes (chapter 3), (Amorim et al., 2017). Even though we could not apply the expected survival functions to road crashes, this study gave rich insight on what should be the next steps in this research by demonstrating that a generalized approach – taking into account all types of emergencies – is the best moral and operational choice in our view.

From this point out, we investigated strategic and tactical uEMS decisions under the scope of the complete medical range. To simplify our approach, we clustered all types of emergencies into life-threatening and non-life-threatening emergencies segregating cardiac-arrest emergencies and road crash emergencies when required. In the end, and from the results presented in chapter 3, we assume that improving the uEMS transport system as a whole will also improve the response to each medical emergency type, at least when it comes to life-threatening situations.

1.3. RESEARCH PURPOSE AND OBJECTIVES

Emergency Medical Service responds to medical aid calls with the objective of protect and ensuring public health and safety. In urban areas, as previously defined, highly dynamic environments actuate over the urban Emergency Medical Service's demand and in the drivability conditions.

For this service, demand is very hard to categorize because of its heterogeneous characteristics and its satisfaction is not easy to quantify. Nevertheless, the use of operational performance metrics such as the average response time, demand coverage and other time or space measures might not be the proper way if we want victims' outcome to be the focus of our approach – which gives value for tax payer, aligns with constitutional rights and reduces external costs for society. Therefore, survival functions are a possible metric to better assess the service performance.

Moreover, as will be shown in the state of the art, the uEMS research has been focused on the mathematical and theoretical problems to solve facilities and vehicle location. These models are many times very simplified representations of the real world. The real improvement they might achieve is hard to quantify because testing and comparing in the real world is prohibitive. It is important that research focus on providing real evidences, or at least empirical evidence, of these models' potential.

What is it that a certain approach or model is improving? – What looks optimal on paper might not correctly, or fully, translate into practice.

We claim that uEMS transport system is highly dependent on city dynamics and optimal or improved solutions require proper empirical evidences – these evidences should focus on the victims' outcomes rather than on operational metrics due to the heterogeneity of the demand.

To support our claim this thesis aims to analyze the different planning levels that define strategic and tactical decisions and how these decisions will perform in an urban environment highly dynamic, particularly, an environment where people location and traffic behavior are in constant change. To achieve this, we explored optimization models and performance metrics to support strategic and tactical decisions and applied simulation to assess these models and provide empirical evidence. With this approach, we addressed different gaps that exist in the state of the art:

- Traditional models use objective functions that intend to maximize coverage or minimize time response;
- Most of the works have no regards for the emergency calls priority or heterogeneity;
- Traffic and demand are usually rough estimations and daily changes are not accounted for;
- Most of the research focus on the “the closest vehicle” dispatching rule.

Our approach comes in line with the uEMS research tendencies and last findings:

- new approaches use survival objective functions for the optimization model
- some studies start to address specific medical emergency types, e.g. road crashes
- the use of scenario and multi-period approach to account for dynamic effects
- integrated strategical and tactical decisions model
- dispatching the closest vehicle is not always the best solution

Therefore, the objectives of this research are summarized as follows:

- Identify the possibility of isolating the demand heterogeneity by studying a service focused on road crashes;
- Identify and compare different performance metrics and assess their value when it comes to victims’ outcomes;
- Assess different location models and how to implement city dynamics to produce more robust solutions;
- Provide a platform that allows for empirical inference of solutions performance to support strategic and tactical decisions;
- Explore dispatching rules and technological advantages that can be used during call time.

1.4. RESEARCH QUESTIONS

Within the boundaries traced by the previous section the following research questions were formulated to guide the research focus and produce insightful results:

RQ1. Does it make sense to segregate EMS demand and tailor specific services for specific medical emergencies such as road crashes?

RQ2. How should we measure EMS transport system performance? Through which metrics and how to quantify them to allow comparison of solutions?

RQ3. Should strategic and tactical planning decisions be integrated in a unique model?

RQ4. Is the closest idle vehicle the best dispatching rule, and how can new technologies improve these rules?

To answer these questions a global methodological approach is proposed using data-driven theoretical models that are afterwards tested in a simulation of the real world to obtain empirical evidences, section 1.5. This approach resulted in six chapters, chapter 3 - 8, in the format of scientific papers that were submitted to peer reviewed journals of high interest for the research scope. In Table 1 the correspondence of each research question and the chapter (and corresponding scientific papers) where they are addressed.

Table 1. Map of the thesis research questions and the respective chapters where they are investigated.

	Chapter 3	Chapter 4	Chapter 5	Chapter 6	Chapter 7	Chapter 8
RQ1	X	X				
RQ2		X	X	X		
RQ3			X	X		
RQ4					X	X

Paper 1 – Chapter 3 Road Safety and The Urban Emergency Medical Service (uEms): Strategy Station Location

This paper consists of a research that analyse the possibility to segregate EMS heterogeneous demand, particularly in urban environments. The demand is modeled by emergency type, e.g. cardiac arrest and road crashes; and an optimization model is proposed to locate key vehicle stations according to operational constrains or the requirements of each emergency type and by implementing simplified survival functions.

Paper 2 – Chapter 4. How do Traffic and Demand Daily Changes define Urban Emergency Medical Service (uEMS) Strategic Decisions? A multi-period survival approach

This paper dives deeper in the uEMS demand heterogeneity and the use of survival functions as objective function of the optimization model. It researches on the topic of dynamic cities using a scenario-based approach to capture the different states of the demand and traffic cycle. A deep analysis of the requirements of stations versus average response time and the demand heterogeneity is made. Finally, different optimization models are compared for different performance objectives and levels of stochasticity.

<u>Paper 3 – Chapter 5. An Active Learning Metamodeling Approach for Policy Analysis: Application to an Emergency Medical Service Simulator</u>

This paper implements machine-learning techniques, particularly an active learning model using Gaussian Processes, to make a metamodel of the simulated environment that allows for a quick analysis of station and vehicle location using both the average response time and the survival rate as performance metrics. This model accounts for stochasticity of the traffic and demand. Its outcome is mainly theoretical and serves as the supporting tool for – Chapter 6. However, the model on its own allows for the analysis of frontier solutions for stations and vehicles location, allowing to support decision-makers during solution analysis.

<u>Paper 4 – Chapter 6. An Integrated Approach for Strategic and Tactical Decisions for the Emergency Medical Service: Exploring Optimization and Metamodel-Based Simulation for Vehicles Location</u>

This paper further develops the concepts presented in chapter 4 and adds to the optimization model the possibility to allocate vehicles to the proposed stations through an integrated approach. Afterwards, a metamodel is used to refine vehicles allocation according to empirical evidences. This approach allows to assess the myopic solutions that the desegregated approach (strategic and tactical decisions are made at different stages) produces and show the advantages of integrating these planning stages. Furthermore, the proposed methodology is applied to other types of strategic and tactical problems to show its validity and wide applicability.

<u>Paper 5 – Chapter 7. Emergency Medical Service Response: Analyzing Vehicle Dispatching Rules</u>

In this paper the dispatch of vehicles is analyzed by comparing the classical rule (always dispatch the closest idle vehicle) and using a rule that considers the system survival status. The different rules are tested for a big urban area, San Francisco city, using a very detailed simulation model to assess how each rule behaves in different periods of the day, month and year, inclusive during specific city or holiday events. An initial assessment of the use of real travel time information and vehicles configuration is done.

<u>Paper 6 – Chapter 8. Emergency Vehicles Dispatching Technological Advantages: Implementing Survival and Real-Time Information</u>

This paper finalizes our research by diving deep in tactical decisions. Particularly, it further develops the proposed dispatching rule that takes into account survival, demand heterogeneity and city dynamics. The use of technological improvements is assessed to analyze the possible contribution they can add to the uEMS performance.

1.5. PROPOSED METHODOLOGY

1.5.1. OVERVIEW

To answer the formulated research questions a general methodology was framed to bridge theory and practice. The main pillar of this methodology is the use of the real environment to empirically support theoretical knowledge provided by the mathematical models.

The main features of the methodology are the empirical evidence model and the theoretical model. Both can be implemented alone or be used together. A data module collects real data from the real system, filters it and if needed models the required parameters to feed both the theoretical and empirical models.

The idea underlining this platform, presented in Figure 3, is to provide a framework where different location or dispatching models can be implemented and afterwards tested in a simulation of the real world to provide empirical evidence of each solution performance. This brings us to one final feature of the platform: the performance metric. Operational and survival metrics are implemented to satisfy every stakeholder requirements and objectives.

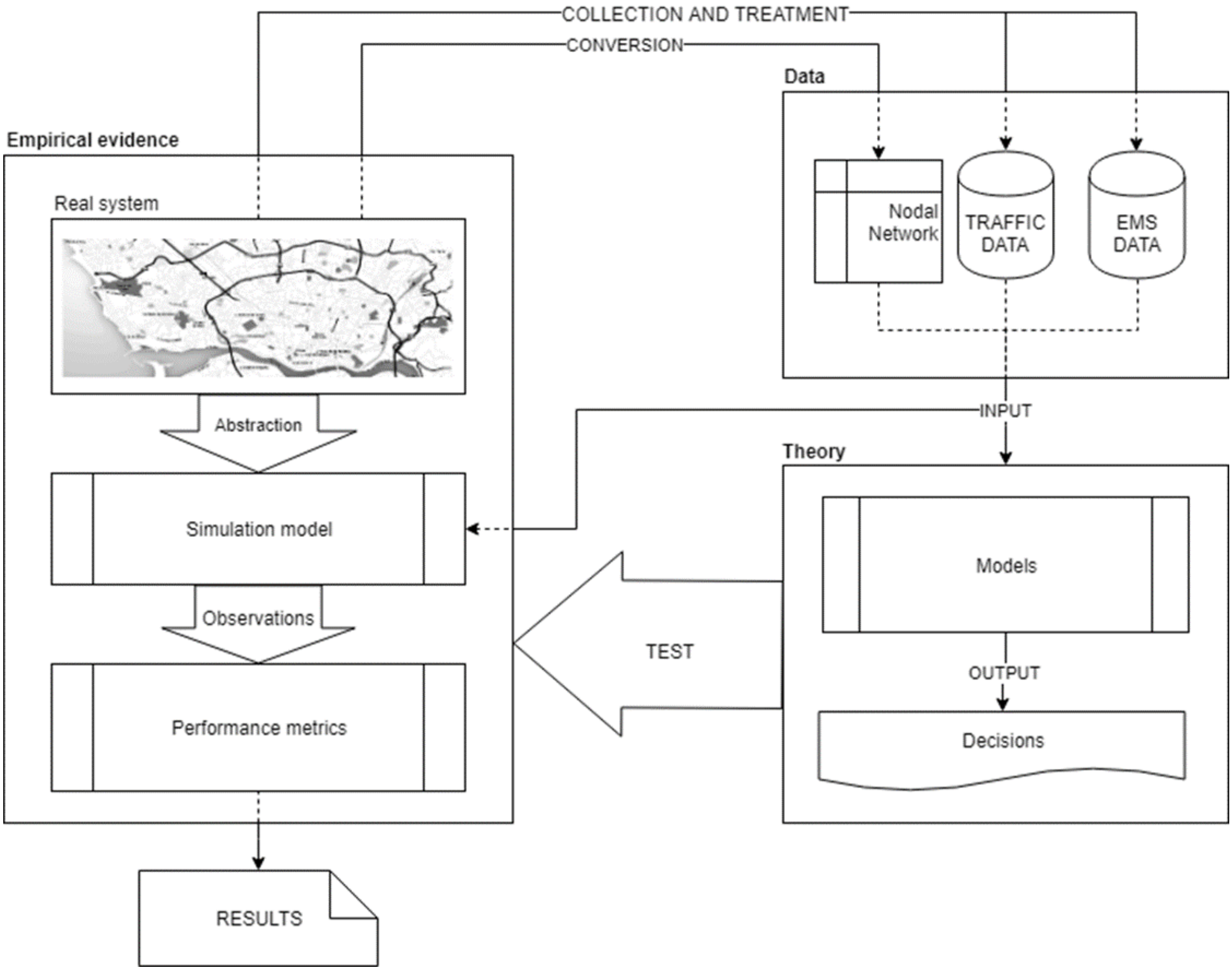


Figure 3.Methodological scheme.

1.5.2. Theoretical Models and Foundations

As previously mentioned a location problem usually assumes that the system will be up for many years thus its success depends on how it will adapt to the many changes that the urban area goes through the years, e.g. continuous changes in the spatial dimension and its distribution.

The line of research that focus on the management of uEMS regards two fundamental aspects: vehicle station location – a garage that shelters a group of medical emergency vehicles – and vehicle dispatching.

The *station location problem* usually aims to cover all the influence area in a manner that there is no single demand unit that cannot be reached within a specific time window – maximum response time limit. However, many studies start to deviate from this rigid rule and focus on a maximal coverage aspect, with the intention of covering the largest possible area within certain maximum response thresholds, leaving certain demand points over the maximum response time limit. There are many situations where the physical and economic conditions of the city simply do not allow a fair use of resources if full coverage is implemented; therefore, many studies assess performance measures to define standards of response.

In the last years, as Erkut et al. (2008) point out that the research direction is to substitute the covering concept as the base to optimize uEMS with concepts that account for survival probability. Furthermore, the temporal changes of the city as well as the daily inhabitants' movements through the city points to dynamic models for uEMS. However, uEMS facilities have a strong static aspect if we think of them as a group of garages for emergency vehicles. Therefore, a question rises whether these two concepts can coexist. The primary idea that rises is to have the facilities statics and reallocate assistance units between facilities in order to fit the service to the continuous morphological change of the city, firstly studied by Berman and Odoni (1982) as a scenario-based approach. Another example is the campus problem addressed by Carson and Batta (1990) where the movements of students through the day lead to a solution of dividing the day in four time periods. If we have in mind a metropolis area with a center that is daily fed by satellite zones that are used as city dorms, we can imagine the spatial variation of inhabitants along the day, thus dynamical concepts are more than plausible.

City dynamics have a big impact on traffic; arterial, collector and local roads have different purposes and with-it different drivability conditions for the emergency vehicles. Time and traffic are highly correlated in high populated cities or urban areas - unless everyone's origins and destinations would be in the same spatial unit. Vehicle dispatching is then affected by how other vehicles are dispersed in the road infrastructure. The subject is widely studied and many works have already implemented stochastic traffic scenario and real time traffic information together with dynamic routing, e.g. we can point out Xiang et al. (2008) but a deep review to dynamic and stochastic routing problems is further addressed through Pillac et al. (2013) and pickup and delivery routing problems through Berbeglia et al. (2010). The point here is that with the evolution and introduction of intelligent transport systems (ITS) in this problem, there is a uEMS dispatching problem that can be improved considering that we have the tools to predict traffic and demand conditions at key operational times.

Moreover it has been proven that although under light traffic conditions using a myopic allocation policy[3] will lead to an optimal solution, when the objective is to minimize the long run average cost for heavy traffic the optimal policy can deviate from the myopic policy (Katehakis and Levine, 1986, Jarvis, 1981).

[3] myopic allocation policy – this policy states that to each demand request the model will assign always the closest available server, e.g. the closest ambulance or request the closest facility.

1.5.3. EMPIRICAL EVIDENCE

To demonstrate how each of the theoretical models or assumptions would perform in a real environment an experiment should take place. However, experimenting in a real uEMS situation is prohibitive, thus the second-best option is to create a simulation of the real environment.

We developed a simulation that is based on a multi-agent model and reflects the key aspects of this research: demand and traffic changes.

Dynamic environments are in constant change; in cities this translates into land demographic occupation changes, such as daily periods when people cluster in business and industrial areas, and night periods when people cluster in residential and nightlife areas. Moreover, when residential areas exist and cluster far from the business and industrial areas, traffic flow differs during commuting times; in the morning towards the city´s business, industrial and commercial areas and in the evening towards the residential areas.

When studying vehicle allocation and dispatching, the assistance time and driving conditions are usually unknown. To cope with variables where their distribution is unknown or that vary in a random way, simulation allows us to introduce randomness in our model. The main idea behind the proposed simulation model is to feed a simulated environment where a uEMS system exists with an infinite number of possible vehicle and station configurations can be requesting through different dispatching policies and using different available technologies.

To simulate the system, an agent-based model is used, where an authority agent, the *city agent*, controls lower level agents: the *event agent*, road *network agent*, *ambulance agent*, and *node agent*. These *agents* coexist in an environment that simulates a spatial area defined by nodes, key locations, and a set of arcs connecting those nodes (Algorithm 1).

Algorithm 1 General simulation algorithm

<u>Definitions:</u>
T = simulation period
t = timestamp
j = step
j = 60 s
t = 0

<u>While $t < T$</u>

1. Update city

- Sets the environment conditions, s, from possible status $S = \{s_1, s_2, \ldots, s_n\}$, where $s = f(time)$
- Move events from events waiting list $E^w = \{e_1, e_2, \ldots, e_m\}$ to events active list E^a if the timestamp of event $e_m(t) < time$, and generate assistance time required, e_n^{atime}

2. For all vehicles in the network:

- Vehicle time to destination, a_d^t, is updated → $a_d^t = a_d^t - j$
- If $a_d^t = 0$ → transfer vehicle to destination

---

3. For all active events $e_n^a \in E^a$:
- if no vehicle is allocated → run Vehicle dispatching algorithm, **Algorithm 2**
- If vehicle is at the occurrence location → Update assisting timer, $e_n^{atime} = e_n^{atime} - j$
- If $e_n^{atime} \leq 0$, assisting time ended → run vehicle to hospital routing algorithm, **Algorithm 3**

4. Update nodes of type Hospital
- If vehicle arrived → Transfer event to hospital
- Ask network to return vehicle to its station → set new a_d^t

5. Update results dictionary, $R_{\{i\}\{j\}}$, with i = t and j = a^g
- For all vehicle in the network → if not in original station, a^g, $R_{\{i\}\{j\}} = R_{\{i\}\{j\}} + 1$, with $i = t$ and $j = a^g$

6. If $t < T$ go back to 1.

Algorithm 2 Vehicle dispatching algorithm

Definitions:

Station $s_p \in S = \{s_1, s_2, \ldots, s_p\}$WHO (2011)

$S_p^a = \{ s_p^1, s_p^2, \ldots, s_p^a\}$ is a list of ambulances parked at s_p

$C = \{t_v, t_{v+j}, \ldots, t_b\}$ is a set of timestamps t

e_m^{max} is the maximum allowed response time for e_m

$Time(s_p, e_m)_c$ is the minimum time travel between station s_p and e_m at scenario $c \in C = \{c_1, c_2, \ldots, c_m\}$

1. For all c *in* C: if t *in* $c \rightarrow q = c$
2. For all s_p *in* S: order S by $Time(s_p, e_m)_q$ or other dispatching rule function F(a, b, c…N) in ascending order
3. For all s_p *in* S: if $S_p^a \neq \{\emptyset\}$ and $Time(s_p, e_m)_q \leq e_m^{max} \rightarrow a = s_p^1$, proceed to 5
4. Select $s_1 \rightarrow$ create s_1^1, $a = s_1^1$ or wait for a vehicle to become available
5. Allocate a to e_m and return to **Algorithm 1**

Algorithm 3 Vehicle to hospital routing algorithm

Definitions:

Node of type Hospital $h_r \in H = \{h_1, h_2, \ldots, s_r\}$

$C = \{t_v, t_{v+j}, \ldots, t_b\}$ is a set of timestamps t

$Time(e_m, h_r)_c$ is the minimum time travel between e_m and h_r at scenario c from list $C = \{c_1, c_2, \ldots, c_m\}$

a is the vehicle allocated to e_m, and a^d is the destination of vehicle a

1. For all c *in* C: if t *in* $c \rightarrow q = c$
2. For all h_r *in* H: order H by $Time(e_m, h_r)_q$ in ascending order
3. Select $h_1 \rightarrow a^d = h$
4. Return to **Algorithm 1**

The *city agent* is responsible for generating and dispatching vehicles when required and activating the events at the right time. The city is also accountable for storing all other *agents* and gives update orders to them.

The *event agent* is responsible for feeding the *city agent* with events and informing the *city agent* of its current state, asking for a *vehicle agent* to be allocated when it is activated (Algorithm 2). When being assisted, the *event agent* is responsible for generating a random assisting time and when this time terminates it will request the *network* agent to be transported to the closest *node agent* of type hospital (Algorithm 3).

Algorithm 2 step 3 goes through a list of ordered stations and chooses the one with an inactive vehicle if the time between this station and the event is lower than the maximum time allowed to assist the event or through other pre-defined dispatching rule or policy. When there is no available vehicle, step 4 creates a new vehicle or puts the service at hold.

Algorithm 3 simply chooses the closest hospital (in terms of trip time) by ordering a vector of available hospitals, step 2, and then selecting the first member of the ordered vector, step 3.

The *network agent* is responsible for routing all *vehicle agents* and choosing the closest hospital when a *vehicle agent* is transporting an *event agent*. It is also responsible for computing the fastest real time Origin-Destination (OD) route.

The *vehicle agent* keeps track of its position in the *network agent* and informs the network when it arrives at any destination. It travels to the node where the event occurs, assists the event, brings the event to the closest hospital and returns to its base. It is completely dependent on orders given by other agents.

The node agent can be of three types: simple node, hospital node and station node. This agent assists the network and city agents by storing vehicles and events.

1.5.4. PERFORMANCE METRIC

Usually uEMS strategic and tactical decisions simply care to minimize response time or/and maximize coverage. The measurements of the improvements they serve are most of the time quantified by relative arguments facing what other models have reached. To complete our methodology, we give a better understanding of the improvements we might reach by providing metrics that allow to measure the social and economic impact of each solutions.

The extended variability of medical emergencies that might arise makes it a long task to try to categorize, in an economic and social manner, the impact of the response time for the different medical emergency call types. The task would require many studies over different medical pathologies and how a quick response might affect them, falling within the scope of medical research rather than in the transport system view. Therefore, our research focused solely on road crashes, cardiac arrest and other life-threatening emergencies, and how the improved medical response might affect the injuries survival.

To implement such measure, besides any other intermediary measures intrinsic to each of the planned tasks and tools, it is required to make a thorough treatment of the database through a standardized methodology. These kinds of metrics exist in the literature for cardiac arrest events, however none was found for road crash injuries. To cover this problem, we had to develop a linkage algorithm that links police reports with hospital reports of road crash victims to connect each crash with its social and economic impact and assess the improvements of a faster medical response on those. Because of the difficulties previously presented and because this is not the primary focus of this thesis, we ended up relying in the tools provided by the state of the art. Nonetheless we developed a linkage methodology (Amorim et al., 2014b) and researched on the topic of pre-hospital time impact on road crash injuries (Ferreira et al., 2019).

1.5.5. DATA BASE

1.5.5.1. GENERALITIES

The work produced through the European Commission co-funded project *LIVE* allowed a compilation of road crash and EMS related datasets.

The resulting database from the project is divided in four datasets: road crash injuries information, uEMS calls, and traffic and population demographics. These data refer to the municipalities of Porto, Vila Nova de Gaia, Matosinhos and Maia and regards the period between 2006 and 2011. Not all the four databases exist for every municipality thus we focused on the Porto city region, which was the only complete set. Furthermore, a dataset with the “Calls-For-Service” of all the fire units’ responses to calls to medical emergencies corresponding to the city of San Francisco is also part of our database.

Finally, we built a SQL dataset with the traffic and demographics of the study areas using the Googles Direction API and the Statistics Portugal (INE) web database. The next sections further detail these datasets.

1.5.5.2. ROAD CRASH INJURIES DATABASE - PORTO

The road accidents injuries database was obtained through different datasets from the Portuguese road safety authority ASNR and several hospitals that covers Porto’s metropolitan area: Hospital São João, Hospital and Santo António, covering a 6-year period (2006-2011). Hospital São João has two different datasets, one for emergency entries and other for inpatient time, therefore these two datasets required a linkage process.

To connect the various datasets a linkage methodology was produced (Amorim et al., 2014b) which resulted in a total of 2,802 links. This result leads to a matching success of 42% (2,802/6,741). For the Hospital São João, the linkage between the emergency and inpatient entries resulted in 1,114 matches out of 9,370 emergency records. From the latter, only 1,182 required inpatient time, bringing the linkage rate success to 85%. To confirm a true match, it was assumed that if the destination field reports an inpatient, the match is true and false otherwise. A total of 1,001 true positive matches, 181 false negative matches and 114 false positive matches were assessed, leading to a positive predictive value of 90%.

Finally, when the three data sets are connected, a linkage success rate of 40% was obtained. The ambulance service data set of 2010 with road crash victims transport timestamped was used for validation. After the data treatment, it was possible to verify 98% of matched records, denoting that the linkage process has a potential of 98% of true matches. Detailed description on the methodology is annexed and published as Amorim et al. (2013).

The GPS coordinates of the accidents were computed through Google Maps API using the police crash address information. These coordinates were then attached to each injury after the linkage process.

1.5.5.3. uEMS Calls Database - Porto

The uEMS calls database was collected from the Instituto “Nacional de Emergência Médica”, INEM, “Centros de Orientação de Doentes Urgentes”, CODU, of Porto. The database includes all the calls with source location on Porto, Gaia, Gondomar and Maia municipalities between the 10th May 2012 and the 10th May 2013.

The database contains information on the Date and time of the call, call ID, Type of occurrence (a total of 42 types including cardiac arrest and road crashes), the facility from where the ambulance was dispatched, the priority (the type of vehicle dispatched, e.g. INEM vehicle, assistance unit), and the address of the occurrence (sometimes not complete).

There is a total of 87 481 occurrences whereas 1 125 are resulting from a cardiac arrest and 3 285 from road crashes. It is important to remember that a road crash call might refer to various injuries with different degrees of severity.

1.5.5.4. uEMS Calls Database – San Francisco

A second uEMS calls database was collected. This database refers to the city of San Francisco and is part of the U.S. Government’s open data of the strategic American resources and consists of a collection of Calls-For-Service datasets, which includes all the fire units’ responses to calls in a total of 4.4 million vehicles dispatched between 2000 and 2017. The fire department is responsible for managing the EMS calls and responses, requesting a private unit when required. Thus, the database also records the dispatch of private units (to 911 calls).

Each data entry was characterized by a unique ID, an event ID, the GPS location of the block where the call originated, the type of vehicle dispatched, the priority and the timestamp.

The data were processed and filtered into a SQL database for easy access and data manipulation. From the same open data source, fire station locations were acquired and added to the SQL database.

The city was divided in a grid of 500 m × 500 m cells and each unit is represented by a node corresponding to its center, totaling 518 nodes. 46 fire stations were identified and assigned to the closest node as well as three fictional hospitals to represent the major San Francisco hospitals. The lack of information on the destination hospital for each call required a random allocation of calls to hospitals based on their proximity.

1.5.5.5. TRAFFIC AND DEMOGRAPHICS

To complement the uEMS databases and enable a data-driven simulation of the study cases it was necessary to collect traffic and demographic data for each study case.

Traffic data is required in order to tailor the optimization models to real case studies. Moreover, travel-times are required to investigate how vehicle dispatching behave within a real traffic situation using a simulation model. The simulation model also allowed for an empirical assessment of the system performance by using travel-time as the response time in the survival function of each medical emergency.

The collection of such data is very hard because cities do not have a detailed record of traffic density for each road and at different times of the day. Nevertheless, the goal was to obtain the real travel time for each period in analysis thus we can directly collect travel times instead. Here, the IoT (Internet of Things) comes into place and several phone apps collect driving characteristics such as driving speed or route times. Google has a powerful app, Google maps, which has been collecting data for many years. Using Google Direction API it is possible to access travel times for specific days and times.

We designed a Python script to access Google Direction API and collect the travel-times for the city OD matrices for different periods of the week and different periods of the day. This data was, afterwards, processed into a SQL database for easy access and manipulation.

The demographics of the study zone are important to infer relations between the randomness of an uEMS call and the population-land use characteristics. The INE (National Institute of Statistics) keeps demographic data in open access. This data was collected directly from the INE web databased and imported to the GIS software qGIS.

1.6. RESEARCH SCOPE, LIMITATIONS, ASSUMPTIONS AND BENEFITS

The outcomes of this research consist of a platform to assess strategic and tactical decisions in dynamic environments through a data-driven simulation model that provides empirical proof for theoretical models with the goal to study the heterogeneity of the demand, the use of survival metrics and the implementation of city dynamics in the modeling process.

We provide three optimization models that focus in different levels of planning:

- A global static long-term optimization model that defines key locations for EMS stations supported by a heterogeneous demand model;
- A model that implements discrete time scenarios to allow for a more robust solution on station positioning; and
- A model that introduces an integrated approach on strategic and tactical planning, i.e. location of stations together with the location of vehicles.

With the analysis of the proposed models and when compared to other models that do not account for heterogeneous demand, survival metrics or city dynamics, we were then able to provide useful insight on the weakness of theoretical models and to identify where EMS research should focus.

For real-time tactical decision, we focused on the dispatching policies and used the methodology to test the classical dispatching policy and propose a new one, which focuses on victims' survival and city dynamics. We go further in our research and complement the study of dispatching rules with the study of new technologies by assessing the advantages of using intelligent predictive models both for demand and for travel times.

The methodology and proposed models have the potential to assist stakeholders during the decision-making process to balance or compare different solutions or policies. The different assessments carried out through this research provide educated insight on three important and actual problems in EMS (research):

- A non-integrated strategic and tactical approach produces myopic solutions thus effort should focus on models that integrate the two levels of decision;
- The classic dispatching policy, heavily focused on dispatching the closest idle vehicle, is not optimal and it is not even clear that a single policy or rule is enough to reach optimality; and
- Operational performance metrics are far from satisfying the victims' needs and effort should be put on the use of survival approach.

Our research tackles a very wide subject and at very different planning levels. However, for obvious reasons, we had to limit our research to the following essential assumptions and simplifications;

- It was assumed that we know the real travel-time of the emergency vehicles;

- The problem of vehicle routing was left out of this research. Vehicle reallocation or rerouting was also not part of the scope because we only focused on the decision making till the vehicle dispatching;
- The survival functions were simplified as well as the demand heterogeneity because this is still an active research topic waiting for new inputs; and
- The proposed models are intended to be solved in short time for different runs without the use of approximated heuristics. The goal of this research is to study the consequences of certain decisions or approaches, not the mathematical or computational aspect of finding solutions.

Overall, the output of this thesis can benefit primarily the emergency medical service institution, public health organizations, a wide range of research areas, and eventually the public, i.e. the urban citizens.

Research-wise- as of today, the contributions from this work lead to the creation of a dedicated session in the TRB Annual meeting, which, due to its success, expanded to the creation of a new subcommittee solely focused on the EMS transport system and the implementation of new technologies and approaches. The national emergency medical service, INEM, is also benefiting from our results and guidelines particularly at dispatching time.

1.7. THESIS OUTLINE

This thesis is divided in 9 chapters. Chapter 1 introduces the thesis scope and the transport system here at study: The Emergency Medical Service. A brief description of the different concepts used along our research is presented and a résumé of the research development is provided. We also present of methodology that follows this thesis objectives and research questions.

Chapter 2 gives a global literature review on the main concepts of this thesis; these are the EMS demand characteristics, EMS performance and EMS strategic and tactical decisions.

Chapter 3 through chapter 8 -, in line with Table 1 - present the six scientific papers that resulted from our research. A brief abstract of each of the papers is presented in subchapter 1.4.

Finally, chapter 9 wraps up the research by highlighting the main research results, findings and conclusions. Some guidelines for further development of this research are proposed.

1.8. REFERENCES

Amorim, M., Ferreira, S. & Couto, A. An accident record linkage study: The first step towards an injury strategy. *In:* Citta, ed. CITTA 6th Annual Conference on Planning Research: RESPONSIVE TRANSPORTS FOR SMART MOBILITY, 2013 Coimbra, Portugal. CITTA.

Amorim, M., Ferreira, S. & Couto, A. A conceptual algorithm to link police and hospital records based on occurrence of values. Transportation Research Procedia, 2014 2014a. 224-233.

Amorim, M., Ferreira, S. & Couto, A. 2014b. Linking police and hospital road accident records. *Transportation Research Record: Journal of the Transportation Research Board,* 2432**,** 10-16.

Amorim, M., Ferreira, S. & Couto, A. 2014c. Linking police and hospital road accident records: How consistent can it be? *Transportation Research Record: Journal of the Transportation Research Board***,** 10-16.

Amorim, M., Ferreira, S. & Couto, A. 2017. Road safety and the urban emergency medical service (uems): Strategy station location. *Journal of Transport & Health,* 6**,** 60-72.

Amorim, M. R. S., Falcão, L., Couto, A. F., Rodrigues, C. & Tavares, J. P. 2014d. 2nd progress report-live project.

Berbeglia, G., Cordeau, J.-F. & Laporte, G. 2010. Dynamic pickup and delivery problems. *European Journal of Operational Research,* 202**,** 8-15.

Berman, O. & Odoni, A. R. 1982. Locating mobile servers on a network with markovian properties. *Networks,* 12**,** 73-86.

Carson, Y. M. & Batta, R. 1990. Locating an ambulance on the amherst campus of the state university of new york at buffalo. *Interfaces,* 20**,** 43-49.

Couto, A., Amorim, M. & Ferreira, S. 2016. Reporting road victims: Assessing and correcting data issues through distinct injury scales. *Journal of safety research,* 57**,** 39-45.

Elvik, R., Vaa, T., Erke, A. & Sorensen, M. 2009. *The handbook of road safety measures*, Emerald Group Publishing.

Erkut, E., Ingolfsson, A. & Erdogan, G. 2008. Ambulance location for maximum survival. *Naval Research Logistics,* 55**,** 42-58.

Ferreira, S. & Amorim, M. 2014. Data collection, analysis and recommendations-live project (deliverable 1).

Ferreira, S., Amorim, M. & Couto, A. 2016. Risk factors affecting injury severity determined by the mais score. *Traffic Injury Prevention,* 18**,** 515--520.

Ferreira, S., Amorim, M. & Couto, A. 2018. Exploring clinical metrics to assess the health impact of traffic injuries. *International journal of injury control and safety promotion,* 25**,** 119-127.

Ferreira, S., Amorim, M. & Couto, A. 2019. The pre-hospital time impact on traffic injury from hospital

fatality and inpatient recovery perspectives. *Journal of Transportation Safety & Security*. 1-21.

Ferreira, S., Falcao, L., Couto, A. & Amorim, M. 2015. The quality of the injury severity classification by the police: An important step for a reliable assessment. *Safety Science,* 79**,** 88-93.

Jarvis, J. P. 1981. Optimal assignments in a markovian queueing system. *Computers & Operations Research,* 8**,** 17-23.

Katehakis, M. N. & Levine, A. 1986. Allocation of distinguishable servers. *Computers & Operations Research,* 13**,** 85-93.

Kepaptsoglou, K., Karlaftis, M. & Mintsis, G. 2012. Model for planning emergency response services in road safety. *Journal of Urban Planning and Development ASCE,* 138**,** 18-25.

Pillac, V., Gendreau, M., Guéret, C. & Medaglia, A. L. 2013. A review of dynamic vehicle routing problems. *European Journal of Operational Research,* 225**,** 1-11.

Sánchez-Mangas, R., García-Ferrrer, A., De Juan, A. & Arroyo, A. M. 2010. The probability of death in road traffic accidents. How important is a quick medical response? *Accident Analysis & Prevention,* 42**,** 1048-1056.

Who 2011. Global plan for the decade of action for road safety 2011-2020. *World Health Organization*.

Who 2013. Global status report on road safety 2013: Supporting a decade of action. *World Health Organization: Geneva*.

Xiang, Z., Chu, C. & Chen, H. 2008. The study of a dynamic dial-a-ride problem under time-dependent and stochastic environments. *European Journal of Operational Research,* 185**,** 534-551.

2. THEORETICAL MODELS – A LITERATURE REVIEW

2.1. SURVIVAL AND MEDICAL RESPONSE

One of the greatest impacts of planning uEMS is the medical response time and how that can affect victims' survival. Most of the research on survival rates due to uEMS response times focuses on cardiac arrests (Erkut et al., 2008). However, as previously presented, there is also an interest in understanding how these survival rates work for road crashes (or any other type of medical emergency) and how to implement it in uEMS with road crash focus (Kepaptsoglou et al., 2012).

Focusing in cardiac arrest, which is where survival functions are deeper developed, there is a study by Eisenberg et al. (1990) who evaluates the survival rates of cardiac arrest in out-of-hospital individuals. Hypothetical survival curves suggest that the ability to resuscitate is a function of time, type, and sequence of therapy, and early cardiopulmonary resuscitation (CPR) permits definitive procedures, including defibrillation, medications, and intubation, to be more effective (Eisenberg et al., 1990). There is a big advantage in rapidly assisting such medical conditions. The deeper analysis of the previous study indicates that without any intervention the survival rate of a cardiac collapse drops, linearly, to zero after 10 minutes. With CPR techniques the linear slope decreases however keeping its negativity. Stabilization of the patient is only assumed when paramedics administrate medication and intubation. But if there is no local assistance, stabilization is assumed when the patient arrives at the hospital. In the latter, that would mean to assume that the period between the cardiac arrest and the arrival at the hospital must not exceed 10 minutes plus the gain of using CPR techniques.

Another study that shows the differences mentioned before is the one from Valenzuela et al. (2000) held in a casino where the security officers were trained for CPR and defibrillation. The author concludes that the survival rate for those who received their first defibrillation no later than three minutes after a witnessed collapse was of 74 percent, and for those who received their first defibrillation after more than three minutes the rate dropped to 49 percent.

Erkut et al. (2008) point out four relevant studies that estimate such survival functions and that we will address next.

Larsen et al. (1993) use data from a cardiac arrest surveillance system in place since 1976 in King County, Washington, where they selected 1,667 cardiac arrest patients with a high likelihood of survival: they had underlying heart disease, were in ventricular fibrillation, and had arrested before arrival of emergency medical services (EMS) personnel. The authors provided us with a survival rate s as per equation (2.1):

$$s = 0.670 - 0.023 I_{CPR} - 0.011 I_{Defib} - 0.021 I_{ACLS} \tag{2.1}$$

Where:

I_{CPR} is the duration from collapse to CPR,

I_{Defib} is the duration from collapse to defibrillation, and

I_{ACLS} is the duration from collapse to Advance Cardiac Life Support (ACLS)

The authors made proofs that there was little or none correlation between the independent variables thus the addictive equation presented a good approach.

Moreover, Valenzuela et al. (1997) indicate that the time interval needed for EMTs or paramedics to attach the defibrillator and clear the patient for defibrillation once CPR was in progress was estimated to be 2 minutes past EMT arrival or 1-minute past time of initiation of CPR by EMTs. They then present a Logistic Regression Survival Model, equation (2.2), to calculate the survival rate:

$$s = \frac{1}{1 + \exp\left\{-0.260 + 0.106 I_{CPR} + 0.1390 I_{Defib}\right\}} \tag{2.2}$$

It is however interesting to point out that the survival function overestimates the probability of survival when the response time is large, as reported by the authors.

The third study mentioned by Erkut et al. (2008) is the one from Waalewijn et al. (2001) using a dataset of cases from out-of-hospital nontraumatic cardiac arrests of patients older than 17 years of age between 1 June 1995 and 1 August 1997. The authors included in their logistic regression a binary variable to indicate whereas the cardiac arrest was witnessed or not by EMS staff. For our analysis, this variable is set to zero because we assume that there is a delay between the cardiac arrest event and the arrival of the uEMS ambulance. Therefore the model built by Waalewijn et al. (2001) is transformed into the equation (2.3):

$$s = \frac{1}{1 + \exp\left\{0.040 + 0.300 I_{CPR} + 0.140 (I_{response} - I_{CPR})\right\}} \tag{2.3}$$

Where:

$I_{response}$ is the time response in minutes.

Finally the last presented model belongs to De Maio et al. (2003) with the use of stepwise logistic regression to estimate survival at various defibrillation response intervals. The data was from January 1, 1991, to December 31, 1997, containing 392 (4.2%) survivors among an overall of 9 273 patients treated. The model construction had several steps in order to find a final model with only the response time as a dependent variable, as per equation (2.4):

$$s = \frac{1}{1 + \exp\left\{0.679 + 0.262 I_{response}\right\}} \tag{2.4}$$

In a recent study of uEMS response to cardiac arrest Gold et al. (2010) observed data showed survival declined, on average, by 3% for each minute that EMS was delayed, following the collapse. However, survival rate did not decline significantly if the time between collapse and arrival of EMS was 4 min. or less but they declined by 5.2% per minute between 5 and 10 minutes. EMS arrival between 11–15 minutes after collapse showed a less steep decline in survival, 1.9% per minute.

Interestingly, when looking at trauma incidents, Newgard et al. (2010) indicate that there was no significant association between time and mortality for any uEMS interval: activation, response, on-scene, transport, or total uEMS time, using multivariable analyses of a set of trauma patients with field-based physiologic abnormality. The dataset corresponded of transported victims by 146 EMS agencies to 51 Level I and II trauma hospitals in 10 sites across North America from December 1, 2005, through March 31, 2007. This study indicates that certain injury profiles do not benefit from an earlier medical treatment within the range of uEMS arrival.

In general terms, Wilde (2013) finds that a one minute increase in response time causes an 8% change in survival within one day of the initial incident using a dataset from the 2001 Utah Pre Hospital Incident Dataset. .

Wilde (2013) concludes that response times are very important for survival from cardiac arrest but less important for survival from other conditions. This is based on the fact that most of the studies fall on the cardiac arrest condition and in the results from the works of Newgard et al. (2010), Pons and Markovchick (2002) and Esposito et al. (1995) which found no association between response times and survival in other types of conditions. However, it is interesting to note that the study by Pons and Markovchick (2002) only clarifies that there was no difference in survival after traumatic injury when the 8 minutes real-time ambulance criteria were exceeded (mortality odds ratio 0.81, 95% CI 0.43–1.52). There was also no significant difference in survival when patients were stratified by injury severity score group. Moreover, Pons and Markovchick (2002) used a database that comprises all types of uEMS calls, where each ambulance of the system is equipped with advanced life support (ALS) and where victims had significant trauma requirements. Backing up this view is the study made by Pepe et al. (1987) whose results indicate that, even in a geographically large urban EMS system, the time factor involved in managing and transporting hypotensive penetrating injury victims directly to a regional trauma center does not appear to be related to an adverse outcome, at least during the first hour after injury. The study comprises of a 30-month-period and 498 consecutive victims of penetrating injury. Moreover, Jones and Bentham (1995), with police data on serious and fatal road crashes between the period of 1987 and 1991, complete these claims stating that although elevated probability of death was found among old, pedestrians, casualties involved in multiple crashes, and casualties on roads with higher speed limits, no relationship was found between outcomes and the estimated time taken to reach victims and convey them to hospital, either before or after adjustment for other factors.

On the other hand, to simply evaluate uEMS response by looking at the survival rate on hospital outcomes would bias the view for our work, where the intention is to reduce the road crash social impact, which starts with the survival at site. In other words: Would a death before arrival be avoided if the emergency vehicle would arrive quicker? In a more recent study on road crash outcomes and their relation with the uEMS response time, Sánches-Mangas et al. (2010) show that the medical response time appears as a significant variable to explain the probability of death for both types of roads (conventional roads and motorways). The authors go further and even indicate that the partial effect of a 10 minutes response time reduction, from 25 to 15 minutes, in motorway road crashes, lead to an increase of the survival ratio of around 33%. For conventional roads, a similar value is obtained (32%).

The list of authors that found a positive relationship between higher distance or time for road crash assistance and higher probability of dead is long (Brodsky, 1990, Brodsky, 1992, Brodsky, 1993, Gonzalez et al., 2009, Li et al., 2008, Durkin et al., 2005, Zwerling et al., 2005, Muelleman and Mueller, 1996, Clark and Cushing, 2002, Evanco, 1999) .

This review concludes that the cardiac arrest survival rate is clearly dependent on the uEMS time response and empirical equations, (2.1), (2.2), (2.3) and (2.4) are provided and widely used to calculate survival. For other types of conditions, there are not enough evidence that correlate survival rates with uEMS time response. However, road crash injuries and their survival rates have been demonstrated to have some correlation with uEMS response time. Most of the authors support the latter claims despite some studies that show otherwise.

2.2. DEMAND ON UEMS

uEMS demand varies spatially and is fluctuating temporally throughout the week, depending on the day of the week, and the time of day (Channouf et al., 2007). Usually, authors assume that demand follows a Poisson process, either by showing theoretical proof (Henderson, 2005) or by empirical evidence, e.g., the works of Brown et al. (2005), Gunes and Szechtman (2005) and Zhu et al. (1992).

Henderson (2005) proves that Gaussian and Poisson random fields have an important role to play in simulation models of spatial phenomena, due to their relative tractability and physical interpretations. The use of the Poisson properties tracks back to the Palm-Khintchine Theorem (see Arthur, 1985) which states that the arrival process that arises from a large number of independent sources, where no source contributes too much to the arrivals, is approximately a Poisson process (Cinlar, 1972).

Another approach to such studies lies in the demand patterns analysis, e.g., moving average. Autoregressive integrated moving average (ARIMA) models are very used in forecasting arrivals to call centers. The bibliography is rich of works that predict daily volumes and:

- incorporate advertising effects (Andrews and Cunningham, 1995); that show the benefits of outlier elimination (Bianchi et al., 1998);
- with unobserved components such as the dynamic harmonic regression (Tych et al., 2002); and
- that predict arrival rates over short intervals in a day via linear regression on the previous day's call volume (Brown et al., 2005).

Such forecasts have helped uEMS managers make more effective resource allocation decisions (Setzler et al., 2009). However these aggregated forecasts, when applied to entire regions, are not sufficient to effectively deploy the often-limited transportation resources in order to minimize response time to an emergency call (Goldberg, 2004).

Setzler et al. (2009) provide an interesting division on the demand model requirements according to the use they will have: dynamic deployment or real-time repositioning. The authors state that for the dynamic deployment the uEMS manager uses call volume forecasts for the next few hours to reallocate the fleet in anticipation of space and amount shifting. For the real-time repositioning, the fleet is reallocated when an ambulance is dispatched in order to maximize the coverage. Moreover, the authors add that in both cases, it is a common practice that when an ambulance finishes a call, it can be dispatched immediately from its current position to a new or pending call. In case of no pending call, the vehicle can be sent to a location that may not be its original base but rather is in the area of the city that currently has the biggest "coverage hole". For dynamic allocation demand forecast, further information can be seen in Rajagopalan et al. (2008) and Channouf et al. (2007).

One of the earliest works in modeling demand is the work of Aldrich et al. (1971) using least squares regression and socioeconomic variables. The model uses dependent variables that address total demand and type of incident (road crashes, accidents, cardiac problems, poisonings, other illnesses, and dry runs) and 31 independent variables describing the study area demographics. The author concludes that demand for public ambulances appears to be highly predictable when using socioeconomic, land-use and service variables. Areas with elderly people or of low age seem to be more demanding. The calls pattern seems to be stable over time for each call type.

Another earlier works in modeling demand for pre-hospital care is the one of Kamenetzky et al. (1982). The authors provide models for uEMS based on population demographics such as inhabitants and employment. The age demographic is deeper analyzed in the work of McConnel and Wilson (1998). The work indicates that the pattern of utilization associated with age can be divided into three modes. Further figures in the same work show that compared to the age group 45 to 64 years of age, rates of utilization for those aged 85 years and older were 3.4 times higher for total uEMS incidents, 4.5 times higher for emergency transports and 5.2 times higher for incidents of a life-threatening nature.

Other representative studies with similar methodologies, but with just slight changes in the dependent and independent variables, are those of Siler (1975), Kvålseth and Deems (1979), Kamenetzky et al. (1982) and Cadigan and Bugarin (1989).

Within the short breaks of the previously mentioned models such as autocorrelation, multicollinearity, and the difficulty of finding meaningful explanatory variables, new models that implement exponential smoothing to the previous ones or use ARIMA model are studied (Winters, 1960, Trudeau et al., 1988).

Channouf et al. (2007) collected several time series models and applied it to the real case study of the city of Calgary in Alberta, Canada. The authors concluded that it is possible to generate a forecast of the number of calls few days in advance when using autoregressive models. Using a devised conditional distribution approach, it was possible from the daily predicted volume to estimate hourly demands. Moreover, the authors addressed the importance of spatial studies to strengthen the forecasts.

Nevertheless, real-time management requires a detailed description of the demand, which these models lack to address. State of the art forwards then to the use of Artificial Neural Networks (ANN) to overcome such disadvantages (Cao et al., 2005, Denton, 1995).

Setzler et al. (2009) provided a comparative study using ANN and a moving average formula. The authors conclude that the moving average only overperforms the ANN for high resolution grids. They indicate that future studies could investigate the use of population and demographic variables, perhaps in addition to historical call volumes, and a better description of the population demographics such as population shifts hour-by-hour and key demographic elements such as age, employment status, and income level.

Finally, Henderson (2005) indicates that we must attempt to calibrate models with very little (relatively speaking) data. The author justifies this claim by stating that we are trying to model a multidimensional random vector (or even a full-time series) rather than a univariate random vector. The curse of dimensionality is the key problem. The author concludes that this difficulty with calibration suggests that methods for addressing input uncertainty will play an important role in simulations involving random vectors with complicated joint distributions.

2.3. uEMS Optimization Models

2.3.1. Classical Models

The optimization of emergency response (ER) and emergency medical services (EMS) is tightly connected to operational research (OR), and usually, it is the driver that conducts OR trends. The common topics of study are generally divided into three major groups, Urban Services, Disaster Services, and Hazard-Specific. The study by Simpson and Hancock (2009) *Fifty years of operational research and emergency response* indicates that over the years, from 1965 to 2007, the OR community focus lies on the latter two mentioned groups. Specific historic events drive the attention from the operational research academy in particular to those subthemes of disaster and hazard-specific.

The foundational stream of research for emergency response tracks back to the year of 1955 with fire station location planning studies by Valinsky (1955). Additionally, Hogg (1968) together with Savas (1969) fill the base archetypes for this theme, being the latter focused on the EMS and the two former ones concerning fire-fight facilities.

Further developments in the generalized OR problems lead to the so called Hakimi property which states that there is an optimal solution to a network location problem in which the facilities are located on the nodes of the network and not along the edges (Hakimi, 1965, Hakimi, 1964). This statement might be one of the most well-known properties presented in all the upcoming OR for facility location in nodal networks.

However, the two most relevant works that truly drove the OR community interest in EMS were those of Toregas et al. (1971) and Church and Velle (1974). The former one presents a solution to solve the location set covering problem (LSCP) making sure all demand is covered within a time or distance maximum radius, equations (2.5) and (2.6):

$$Minimize \rightarrow z = \sum_{j=1}^{j=J} y_j \tag{2.5}$$

Subject to:

$$\sum_{j \in N_i} y_j \geq 1 \tag{2.6}$$

With:

i being any node of the network node set I ,

j being any possible node of the node set J where a facility can be established,

y_j being the decision variable, which takes value 1 if a facility is established at point j and 0 otherwise,

$N_i = \{ j \mid d_{ij} \leq s \}$,

s being the acceptable time distance from j to i and

d_{ij} being the response time or distance from any node i to node j.

Nonetheless, full coverage is hard to reach especially when resources are limited, which is the case of any practical problem. Li et al. (2011) reviewed covering models for uEMS in their work "*Covering models and optimization techniques for emergency response facility location and planning: a review*." The authors point out many future models that relax some of the Toregas et al. (1971)'s assumptions. The discrete points network of Toregas et al. (1971) is changed to a continuous region in Aly and White (1978) work on probabilistic formulations for the emergency service problem using stochastic response time. A hierarchical vision of the LSCP is proposed by Daskin and Stern (1981) with the objective of minimizing the number of facilities, providing full coverage within a distance standard first and then maximizing the number of demand points with multiple coverages. A probabilistic version of the LSCP is formulated by ReVelle and Hogan (1989) with the requirement that all the demand points must be covered with a reliability level α. Further specifications of this model will be presented in due time.

To integrate the concept of ambulance service capacity into the LSCP and consider the road condition and the population distribution, Shiah and Chen (2007) propose an Ambulance Allocation Capacity Model (AACM). The new approach presented considerable improvements with increases from 49% to 91% in coverage rate and decrease from 48% to 18% in overlapping rate within the study area and using almost the same number of ambulances.

Church and Velle (1974) point out a solution for a maximal coverage location problem (MCLP) that intends to overcome the resources limitation of the Toregas et al. (1971) problem. For this problem, Church and Velle (1974) add a new decision variable,

x_i, which defines whereas a demand node is served or not by any facility, knowing that:

P is the number of facilities to be located and

a_i is the population on i.

Therefore the earlier solutions (equations (2.5) and (2.6)) become as equations (2.7), (2.8) and (2.9):

$$\max imize \rightarrow z = \sum_{i=1}^{i=I} a_i x_i \tag{2.7}$$

Subject to:

$$\sum_{j \in N_i} y_j \geq x_i \ \forall i \in I \tag{2.8}$$

$$\sum_{j=1}^{j=J} y_j = P \tag{2.9}$$

The objective function, equation (2.7), holds now a maximization objective while in Toregas et al. (1971) the goal was a minimization. The difference relies on the fact that one wants to minimize the number of facilities to be used in order to cover all demand points while the other knows that resources are limited thus the number of facilities is known *a priori*, therefore, the goal is to maximize the population served. Although both works have no references to the archetype works earlier mentioned, later work by Toregas and ReVelle (1972) "*Optimal location under time or distance constraint"* brings references to Savas (1969) paper, which is a simple development of his initial idea.

Li et al. (2011) point out the studies of Jia et al. (2007) and Dessouky et al. (2006) as an extension of the MCLP. These studies fall in large scale EMS using multiple quality levels and multiple quantities of facilities at each quality level for demand points. The earlier mentioned study suggests that the minimum number of facilities that must be allocated to demand point i to achieve a certain quality level of coverage should be determined by population, a weighted factor, and an emergency occurrence likelihood at each demand point.

Other extensions of the MCLP are addressed by Schilling et al. (1979) where coverage is provided by two distinct types of servers, one of which is the Tandem Equipment Allocation Model (TEAM). In the same framework falls the Backup Double Covering Model, BDCM (Başar et al., 2009). Moreover, the work of Hogan and ReVelle (1986) introduces the notion of Backup Coverage Problem by maximizing the population coverage with more than two facilities while forcing all demand points to be covered once. Alsalloum and Rand (2003) and (2006) developed Goal Programming models and extend the MCLP. Initially, by determining the locations of facilities to maximize expected demand coverage and subsequently by adjusting the capacity of each station while meeting the minimum performance requirements. Marianov and Serra (1998) propose a queuing version of the MCLP and calls it Maximal Covering Location-Allocation Problem (MCLAP). They propose new linear models for locating service centers in a congested situation. These models explicitly include a constraint on service quality, specifically the waiting time or queue length at each center, and are solved through heuristic solutions which are compared to the solutions obtained by commercial optimization packages. Finally, it is important to refer the work of Erkut et al. (2008) which incorporates a survival function into the covering model and formulates the Maximum Survival Location Problem (MSLP). The author's model falls as per equations (2.10), (2.11), (2.12) and (2.13):

$$Maximize \rightarrow \sum_{i=1}^{I} \lambda_i \sum_{j=1}^{J} s(t_{ji} + t_\lambda) y_{ij} \tag{2.10}$$

Subject to:

$$\sum_{i=1}^{I} y_{ij} \leq I \times y_j,\ j = 1,\dots,\mathrm{J} \tag{2.11}$$

$$\sum_{j}^{J} y_{ij} = 1,\ i = 1,\dots,I \tag{2.12}$$

$$\sum_{j=1}^{n} y_j \leq P \tag{2.13}$$

Where:

λ_i is the demand on i,

t_{ji} is the travel time from j to demand node i,

t_d is the pretravel delay,

$s(t_{ji} + t_\lambda)$ is a function of time and can be obtained from equation (2.4), and

y_{ij} is equal to 1 if node i is served by an EMS in position j.

The survival function is a monotonic decreasing function, mapping response time to survival rate and the model is tested with out-of-hospital cardiac arrest emergencies (Li et al., 2011).

Bringing closer attention to the first approach, maximal coverage models, we can conclude that this method is more a generalized facility location problem solution rather than specifically intended to solve EMS problem, i.e. in an EMS network every single demand point – every household – must be covered. It is, of course, understandable that some households will be of quicker reach than others will, but there must always be a minimum assistance time. Thus every network node must be within that predefined reach time of a facility.

The classical interpretation of the facility location problem, in particular to urban emergency services, soon was overcome by uncertainty approaches leading to double coverage, scenario approach, stochastic and robust optimization problems as well as dynamic location. Some were already presented as extensions of the two classic models, MCLP and LSCP. However, Cooper (1974) explored the stochastic approach by assuming a bivariate normal distribution to solve the Weber problem. In this approach, the location of the demand points may be random, and an iterative algorithm was developed to solves the first-order conditions; Sheppard (1974) followed a scenario approach to facility location, although the first rigorous approach was from Mirchandani and Oudjit (1980); It was not up until the ’80s, between 1981 – 1984 up to 1990, that OR focus on Urban Services reached its peak (Simpson and Hancock, 2009).

Following the previous models and each one's problems, new views from different authors were proposed. Focusing on the fact that once a facility is called for service the demand points under its coverage are no longer covered, Daskin and Stern (1981), (1983) and Hogan and ReVelle (1986), (1989) account for facility busy probability and reliability. The former ones solving the maximum expected covering location problem (MEXCLP) and the latter defining the maximum availability location problem (MALP).

Daskin and Stern (1981) and (1983) propose extending the original LSCP (Toregas et al., 1971) to a hierarchical objective problem by keeping the first objective, minimize the number of facilities, and adding a new objective with the intention to maximize the number of times a node is covered. Equations (2.5) and (2.6) of the LSCP (Toregas et al., 1971) problem now become equations (2.14) and (2.15) of the MEXCLP (Daskin, 1983):

$$\begin{aligned} &Minimize \rightarrow z_1 = \sum_{j=1}^{j=J} y_j, \\ &Maximize \rightarrow z_2 = \sum_{i=1}^{i=I} \mathrm{M}_i \end{aligned} \tag{2.14}$$

Subject to:

$$\sum_{j \in N_i} y_j - \mathrm{M}_i \geq 1 \quad \forall \boldsymbol{i} \in \boldsymbol{I} \tag{2.15}$$

With:

M_i being the number of times node i is covered in addition to 1.

Nevertheless, it is important to remember that demands are not evenly distributed temporally and spatially, thus the busy probability varies from facility to facility leading to the maximal expected coverage location model with time variation (TIMEXCLP) from Repede and Bernardo (1994), where varying temporal demands are incorporated.

Again, the review from Li et al. (2011) indicated several other extensions of the MEXCLP. Fujiwara et al. (1987) and Fujiwara et al. (1988) applied simulation to make further analysis on the optimality of an EMS location problem in Bangkok with the use of MEXCLP. Saydam and McKnew (1985) reformulated the MEXCLP into a nonlinear form using a separable programming approach. Later, Rajagopalan et al. (2007) employ a statistical experimental design to guide and evaluate the development of four meta-heuristics applied to a probabilistic location model, specifically to solve the MEXCLP. Finally, and not long ago, with the idea of incorporating local reliability estimation, Sorensen and Church (2010) formulated Local Reliability-based MEXCLP (LR-MEXCLP). A hybrid model that combines the local business estimates of Maximum Availability Location Problem (MALP) with the maximum coverage objective of MEXCLP.

Contrasting with the MEXCLP the Maximum Availability Location Problem (MALP) proposed by Hogan and ReVelle (1986) and ReVelle and Hogan (1989), developed with roots from the MCLP (Church and Velle, 1974) and the LSCP, seeks to position P facilities in such a way that maximizes coverage within a distance or time-of-travel standard S and a reliability of α . MALP has two versions, MAPL I and MALP II, however for the sake of generalization we will just contemplate the second version as this one assumes that the busy fraction of the facilities may differ across different city sections while MALP I assumes that all facilities are equally busy. The definition of busy fraction q leads to the creation of a chance constraint on service availability originally presented by Charnes and Cooper (1959) in order to determine the service requirements of the demand areas as per inequality (2.16):

$$\sum_{j \in N_i} x_i \geq b_i \tag{2.16}$$

Where:

$$b_i = \left\lceil \frac{\log(1-\alpha)}{\log(q_i)} \right\rceil \tag{2.17}$$

Moreover,

$$q_i = \frac{t \cdot \sum_{k \in M_i} a_k}{24 \sum_{j \in N_i} y_j} \tag{2.18}$$

With:

α being the reliability of the facilities,

t being the average duration of a call (hours), including all the time the vehicle is out,

q_i being the busy fraction of a facility and

M_i being the set of demand nodes within S travel time of node i .

The goal of MALP II is to maximize the population of demanding areas which have b_i facilities within S , in other words, to maximize the population with α reliability. The MCLP problem of Church and Velle (1974) presented in equations (2.7), (2.8) and (2.9) becomes the MALP II solution presented by ReVelle and Hogan (1989) as per equations (2.19), (2.20) and (2.21):

$$Maximize \rightarrow z = \sum_{i=1}^{i=I} a_i x_{ib_i} \tag{2.19}$$

Subject to:

$$\sum_{j \in N_i} y_j \geq \sum_{k=1}^{b_i} x_{ik} \quad \forall i \in I \tag{2.20}$$

$$x_{ik-1} \geq x_{ik} \quad \begin{cases} \forall i \in I \\ k = 2, \ldots, b_i \end{cases} \tag{2.21}$$

With:

x_{ib_i} being 1 if server b_i is placed in N_i, and 0 otherwise, and

x_{ik} being 1 if demand area i has at least k facilities within S.

Ball and Lin (1993) established a new version of the probabilistic LSCP, MALP. In their model, the uncovered probability of each demand point must be below a preset value. Marianov and ReVelle (1996) based on the MALP introduces the Queuing Probabilistic Location Set Covering Problem (QPLSCP) by relaxing the assumption that servers were operated independently. The main difference between the new proposed model and the MALP resides in the way b_i is calculated. Also, the travel distances or times are seen as random and consequently derives possible different sets of N_i.

Galvão et al. (2005) dropped the simplified assumptions of the original models MALP I and MEXCLP, and embedded Larson's hypercube model[4] (Larson, 1974) into local search methods to which they called the Extended Maximum Availability Location Problem (EMALP). As per Li et al. (2011) review, the authors state that it was necessary to identify which server was located at which site, therefore they changed the decision variable y_j into y_{kj}, which is equal to 1 if and only if facility k is located at node j and 0 otherwise.

Another important model for double coverage is the Double Standard Model (DSM). DSM aims to allocate facilities among potential sites in order to fully cover the entire study region within a longer distance standard while maximizing the coverage within a shorter distance standard. Gendreau et al. (1997) proposed a tabu search to solve this problem and in between develops its own DSM. The model proposes to maximize the demand covered by two facilities within a radius r_1. Gendreau et al. (1997) define the problem on a graph with two vertices sets representing the demand points and the potential local sites where:

γ_{ij} is a binary coefficient that takes 1 if $t_{i,\,n+j} \leq r_1$ (i is covered within the smaller radius r_1) and 0 otherwise,

[4] The Larson's hypercube model analyzes the behaviors of a multi-server queuing system with distinguishable servers. The study region is partitioned into several cells or geographical atoms with a certain fraction of region wide workload.

δ_{ij} is a binary coefficient that takes 1 if $t_{i,\,n+j} \le r_2$ (i is covered within the larger radius r_2) and 0 otherwise,

α is the proportion of total demand that must be covered within r_1 (in the MALP II model it was the reliability of the facility, i.e., having or not an ambulance to satisfy the demand), and

x_i^k is a binary variable equal to 1 if demand node i is covered at least k times with $k = 1 \vee 2$.

The problem is solved by equations (2.22), (2.23), (2.24), (2.25), (2.26) and (2.27):

$$Maximize \rightarrow z = \sum_{i=1}^{I} \lambda_i x_i^{k=2} \tag{2.22}$$

Subject to:

$$\sum_{j=1}^{J} \delta_{ij} y_j \ge 1 \quad \forall i \in I \tag{2.23}$$

$$\sum_{i=1}^{I} \lambda_i x_i^{k=1} \ge \alpha \sum_{i=1}^{I} \lambda_i \tag{2.24}$$

$$\sum_{j=1}^{J} \gamma_{ij} y_j \ge x_i^{k=1} + x_i^{k=2} \quad \forall \boldsymbol{i} \in \boldsymbol{I} \tag{2.25}$$

$$x_i^{k=2} \le x_i^{k=1} \quad \forall \boldsymbol{i} \in \boldsymbol{I} \tag{2.26}$$

$$\sum_{j=1}^{J} y_j = P \tag{2.27}$$

The main difference between the DSM and the MALP is that the DSM directly assumes that two rules must be sastified when setting a uEMS. First rule is that a maximum time response must be fulfilled within a certain radius, as per the previous studied models. Second rule is that a more tight time response exists to double cover the demand points being it the goal of the maximization problem; On the other hand, MALP assumes that facilities might be busy at certain periods thus a different facility must cover the busy facility demand points. As we will see further, the interest of DSM is that we might be able to adapt it to improve the time response to road crashes which can be seen here as the tighter radius fulfillment.

Based on the later model, the DSM, Doerner et al. (2005) and (2008) propose some models where they augmented the penalty terms to the objective function to avoid unmet coverage requirements and uneven workload.

This concludes the basics of the classical approaches for uEMS facilities optimizations. The drive for this sub-section was to build the base for the upcoming specific approaches of uncertainty to deal with the demand and time. The problem at hand enters in field of randomization where demand occupies an important role, as previously mentioned.

Rosenhead et al. (1972) presents three possible scenarios for each problem at hand:

- Certainty – all parameters are deterministic and known
- Risk - uncertain parameters, values governed by known probability distributions (**Stochastic optimization problems**)
- Uncertainty – uncertain parameters, unknown probability distributions (**robust optimization problems**)

In stochastic optimization problems the goal usually orbits around the optimization of the expected value of some objective function, while in a robust optimization problem the goal is to optimize the worst-case performance of the system.

Both Stochastic and Robust optimization have the common goal of finding the solution that will perform well under any possible realization of the random parameters (Snyder, 2006). To do so both optimization formulations require the choice of appropriated performance measures as part of their modeling processes. Continuous and discrete random parameters are considered in different solution approaches, usually using probabilistic functions or scenarios-based approaches.

Snyder (2006) points out some disadvantages for both approaches, which are interesting in the way that they show possible expected problems and how to accommodate them. For the scenario approach Snyder (2006) points out two main drawbacks: the first (and the obvious one) is that identifying scenarios, and, even more, the probabilities tied to them, is an overwhelming and difficult task; the second drawback is that usually we would prefer to limit our number of scenarios due to computational reasons, however, this would limit the range of future states under which decisions are evaluated. For the drawbacks of continuous approaches, Snyder (2006) indirectly presents them by pointing out the scenario approach main advantages: resulting in more manageable models and allowing statistically dependent parameters. Dependency is of course important particularly for us if we want to further model demand according to time and city demographics, thus correlation with time periods and geographic locations is necessary.

Before we proceed further with deeper analysis on the different solutions presented along the years by a vast group of researcher and faculties, it is important to mention the complexity of these types of problems. Usually the stochastic and robust facility location problems fall in the category of NP-hard as per their base construction settled in classical facility location problems (Snyder, 2006). The latter research over the vast existing bibliography, up until 2004, indicates that "*minisum"* models such those of P-Median Problem (PMP) (Hakimi, 1964) and the *"uncapacitated"* fixed-charge location problem (UFLP) (Balinski, 1965), both stochastic problems, are relatively easy to solve. In the other hand resides the minimax structured problems, usually robust location problems, which are more difficult to solve to their optimality. Snyder (2006) sustains that the later type of robust problems can be solved in the same amount of time of a stochastic similar problem but with one order of magnitude more. These conclusions show the parallel difference in difficulty between deterministic *minisum* and *minimax* problems.

2.3.2. STOCHASTIC OPTIMIZATION PROBLEMS

2.3.2.1. GENERALITIES

Most of the stochastic problems that have been solved during the last decades have the objective to minimize expected cost or maximize expected profit (of the system), while others may take a probabilistic approach (Snyder, 2006). The probabilistic approach consists of maximizing a certain qualitative parameter that defines the solution, e.g. the solution is "good".

One of the first attempts on uncertainty tracks back to the 1970's where Cooper (1974) considers the locations of the demand points may be random within Weber problem. The Weber problem, capacitated multisource Weber problem (CMWP), results in the location of a certain number of facilities in a Euclidian plane and allocates them to the customers in order to satisfy their demand at a minimum total costs knowing the location and demand (Brimberg et al., 2000). A branch of this problem is when we put aside the deterministic assumption and consider customer locations are randomly distributed, the so called probabilistic CMWP (PCMWP).

Cooper (1974) resolves the problem by assuming a two dimensional plan with Cartesian coordinates (x^c, y^c) where P_j (P being the number of facilities, P_j is the facility located at *j*) has associated a probability density function $f_i(x_j^c, y_j^c)$. If $r_j = (x_j^c, y_j^c)$ and $r = (x^c, y^c)$ then the expected value of a function $H(r, r_j)$ is given by equation (2.28):

$$E\left[H(r, r_j)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(r, r_j) f_j(r_j) dx_j dy_j \tag{2.28}$$

If we assume that the influence area of a facility P_j is circular, this is, every point distant from P_j by a certain radius is at the same time or distant reach by P_j, the function that is of interest is the same as the one in Cooper (1974) and takes the form of equation (2.29):

$$H(r, r_j) = \beta_j \left[\left(x_j^c - x^c\right)^2 + \left(y_j^c - y^c\right)^2\right]^{\frac{1}{2}} \tag{2.29}$$

Where $\beta_j \geq 0$ and is a known weight.

Obvious this is more interesting for a long-term model where a general view is set over the area in analysis and the only goal is to settle a certain number of facilities thus not requiring an analysis of the city transport infrastructure if we assume traffic homogeneity. For mid-term models where the goal is not only to cover the city but to reduce as much as possible the time response to emergency medical calls then more sophisticated functions that account for the existent infrastructures might be of interest.

Cooper (1974) presents the density function for the bivariate normal distribution as per equation (2.30) and assuming that x_j^c and y_j^c are not correlated:

$$f_i(x_j^c,\, y_j^c) = \frac{1}{2\pi\sigma_{x_j^c}\sigma_{y_j^c}} \times \exp\left\{-\frac{1}{2}\left[\left(\frac{x_j^c-\mu_{x_j^c}}{\sigma_{x_j^c}}\right)^2 + \left(\frac{y_j^c-\mu_{y_j^c}}{\sigma_{y_j^c}}\right)^2\right]\right\} \tag{2.30}$$

All results from Cooper (1974) are valid to a very wide range of density functions as proven by Katz and Cooper (1974). We can then assume that other types of density functions are plausible to be implemented in such type of mathematical model.

The final objective function to be minimized goes as equation (2.31):

$$\underset{x,\,y}{Minimize} \rightarrow z(x^c,\, y^c) = \sum_{j=1}^{J}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(r,\, r_j)\, f_j(r_j)dx_j dy_j \tag{2.31}$$

In the end, Cooper (1974) proves that the objective function is convergent and presents computationally effective method to solve the probabilistic version of Weber problem. This solution is a good starting point for the most common stochastic problem objective - optimize the average outcome of the system.

Altınel et al. (2009) computes $E\left[H(r,\, r_j)\right]$ for specific distance functions and probabilistic distributions, among which the Euclidian, squared Euclidian, rectilinear and $Weighted\ l_{1,2}-norm$ distances and the bivariate symmetric normal and exponential distribution. The authors implemented a location-allocation heuristic and conclude that exact expected distances evaluations are only possible for few cases. Therefore, he proposes an average distance approximation to solve with the most accuracy and simplicity any of the distance function and customers location distribution.

Mousavi and Niaki (2013) work on a capacitated location allocation problem, of the type PCMWP, with fuzzy customer demands and stochastic locations of the customers. The authors use fuzzy demands in order to escape the difficulties to assign probability distributions to the demands in real environment and solve the problem with a simplex algorithm, a fuzzy simulation, and a modified genetic algorithm which they integrate in a hybrid intelligent algorithm. Previously Zhou and Liu (2007) considered the PCMWP with fuzzy demands in which the customers' locations were deterministic, while Wen and Iwamura (2008a) proposed a fuzzy facility location allocation model under the Hurwicz criterion[5] (more about the Hurwicz criterion can be found on Jaffray et al. (2007)) and in another work (2008b) utilizes a random fuzzy environment. Also Abiri and Yousefli (2010) proposed an application of the probabilistic programming approach to model the fuzzy PCMWP where demands were fuzzy and locations were deterministic.

2.3.2.2. DYNAMIC LOCATION PROBLEM

Another problem on the stochastic location for uEMS is the dynamic location problems. Mostly these problems arrive to solve the problem of relocating facilities/ambulances.

Maxwell et al. (2009) classified research on dynamic allocation problems into three categories:

- First category - solving the model in real-time each time a redeployment decision is to be made (Brotcorne et al., 2003, Kolesar and Walker, 1974, Gendreau et al., 2001, Nair and Miller-Hooks, 2006),
- Second category - involves computing optimal ambulance positions for every number of available ambulances via a similar integer programming formulation in an offline preparatory phase (Ingolfsson, 2006, Gendreau et al., 2005).
- Third category - intends to incorporate system randomness into the model by:
    - Modeling the problem as a Markov decision process (Berman, 1981c, Berman, 1981b, Berman, 1981a, Zhang et al., 2008, Alanis et al., 2013, Berman and Odoni, 1982, Jarvis, 1981) .
    - Making decisions under particular system configurations (Andersson and Varbrand, 2006, Andersson, 2005).

To enter in the dynamic formulations let us first address static facility location in which a Single Facility Location Problem (SIFLP) is assumed. One example is the problem addressed by Wesolowsky (1973) which falls in a generalization of the Weber problem where the objective function is to minimize the cost to satisfy the demand by locating a new facility from a set of existing facilities sites, as per equation (2.32) :

$$Minimize \rightarrow Z = \sum_{i=1}^{m} w_i d(P,\ x_i) \tag{2.32}$$

Where:

m is the total number of candidate destinations for the facility,

[5] Hurwicz criterion: the value of a decision is a weighted sum of its lowest possible expected values (pessimist evaluation) and of its highest ones (optimistic evaluation).

w_i is a weight transforming distances into costs for the existing node i, and

$d(P, x_i)$ is distance between node x_i and the facility P.

The problem can afterwards be extrapolated to a dynamic model. Wesolowsky (1973) proposes an optimal location in which p time periods l are considered instead of a single period, as per equation:

$$Minimize \rightarrow \sum_{l=1}^{p} \sum_{i=1}^{m_p} f_{li}(x_l^c, y_l^c) + \sum_{l=2}^{p} c_l z_l \tag{2.33}$$

Subject to:

$$z_l = \begin{Bmatrix} 0 \text{ if } d_{l-1,l} = 0, \\ 1 \text{ else if } d_{l-1,l} > 0 \end{Bmatrix} \text{ for } l = 1, \ldots, p \tag{2.34}$$

Where:

m_p is the total number of candidate destinations in period l,

$f_{li}(x_l^c, y_l^c)$ is the shipping cost between a facility located at (x_l^c, y_l^c) and destination i,

c_l stands for the moving cost in period l and

$d_{l-1,l}$ is the distance by which the facility is transited in period l.

Berman and Odoni (1982) develop a dynamic model based on the generalization of the p-median problem by allowing facilities to be moved at a certain cost in order to better accommodate to network changes. The same idea might be applied to the reallocation of ambulances, therefore the solution Berman and Odoni (1982) takes as assumption that the network, at any instant, can be at a finite number of states and the state transitions are made dynamically with Markovian[6] dependence among the states of the network.

Another assumption is that whereas in a p-median facility problem the facilities are to be located once and for all, in their problem there is the option to relocate them with a certain cost associated. The parallelism to ambulance allocation can easily be made if we change the decision variable. The model presented by Berman and Odoni (1982) assumes that in a network there is:

i nodes from the set I,

[6] A stochastic process has the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present values) depends only upon the present state, not on the sequence of events that preceded it.

l links of the set L,

r representing the network state,

S representing the network state $\neq r$ and

$t_r(i, j)$ representing the travel time between i and j in the network state r.

Then there is at least one $l(i, j)$ where equation (2.35) applies:

$$t_r(i, j) \neq t_s(i, j) \tag{2.35}$$

Without going into deeper details, we bring from Berman and Odoni (1982) the required notion of strategy, which can be viewed as a vector $K = (K(1), K(2), ..., K(m))$ of m elements $K(r), r \in M$ that provides the set of p_j locations where the facilities/ambulances will be located/allocated when the network is at the state r. Finally, we have:

$D_s(K(r), x)$ as the shortest travel time between any point in the set $K(r)$ and a specific point X,

$d_l(K_\alpha(r), K_\gamma(s))$ as the shortest travel time between the α-th point in set $K(r)$ and the γ^{th} element in set $K(s)$ for $\alpha \wedge \gamma = 1, 2, ..., p$,

l as a new state that complies with the inequality in equation (2.35),

$W_s(K_\alpha(r), K_\gamma(s))$ as a binary variable that takes 1 if ambulance at $K_\alpha(r)$ is reallocated to $K_\gamma(s)$ when the state of the network changes from r to s, and

P as an ergodic[7] Markov transition matrix with

$p_{rs} \in P$ as the probability of a transition from a state r to a state s, and

$\pi(\pi P = \pi, \sum_{r=1}^{m} \pi = 1)$ as the steady-state probability vector of the matrix P

Within the previously mentioned assumptions and notations, Berman and Odoni (1982) grant the following solution for the problem as per equations (2.36), (2.37) and (2.38):

$$Minimize \rightarrow Z = A + B \tag{2.36}$$

Subject to:

[7] In mathematics, the term ergodic is used to describe a dynamical system which, broadly speaking, has the same behavior averaged over time as averaged over the space of all the system's states (phase space).

$$\sum_{\alpha=1}^{p} W_l(K_\alpha(r),\ K_\gamma(l)) = 1, \text{ for } \gamma = 1,\ 2,\ ...,\ p;\ \forall r,\ l \in M;\ r \neq l \tag{2.37}$$

$$\sum_{\gamma=1}^{p} W_l(K_\alpha(r),\ K_\gamma(l)) = 1, \text{ for } \alpha = 1,\ 2,\ ...,\ p;\ \forall r,\ l \in M;\ r \neq l \tag{2.38}$$

Where:

$$A = \sum_{r=1}^{m} \sum_{i=1}^{I} \pi_r h_i D_r(K(r),\ i) \tag{2.39}$$

h_i is the conditional probability that a demand comes from node i given that a demand was generated, and

$$B = \sum_{r=1}^{m} \sum_{\substack{l=1 \\ l \neq r}}^{m} \pi_r \cdot p_{rl} \left[\sum_{\alpha=1}^{p} \sum_{\gamma=1}^{p} W_l(K_\alpha(r),\ K_\gamma(l)) \cdot f\left(d_l(K_\alpha(r),\ K_\gamma(l)) \right) \right] \tag{2.40}$$

The quantity A gives the long term (steady-state) expected travel time to facilities on the network per transition epoch.

The same problem is studied further by Carson and Batta (1990). In this study the authors face the problem of reallocating a single ambulance in the Amherst campus of SUNY Buffalo as the population moves throughout the day. Due to the difficulty in identifying probabilistic distributions and estimating relocation costs in practice, the authors propose a discrete dynamic facility location model with four uneven day periods and solves a 1-median problem in each.

Moreover we can point out the more recent works on facility/ambulance location-reallocation elaborated by Alanis et al. (2013) who analyze an uEMS system by a two-dimensional Markov chain model that repositions ambulances using a compliance table policy, a common operational practice. The model has the same data requirements and can produce the same outputs as the Hypercube Queueing Model (HQM) (Larson, 1974) but models repositioning policies are not considered by the HQM. The authors also develop procedures to estimate the parameters of the analytical model; validate the model against a realistic simulation model and, among others, find that the Markov chain model provides a good approximation to several performance measures. Moreover, they demonstrate that the Markov chain model can be used to identify solutions that are near-optimal, as measured by a realistic simulation model. Finally Alanis et al. (2013) results show that different compliance tables may lead to large variations in performance, which demonstrates the importance of using a well-designed compliance table.

In fact, when addressing dynamic location models the bibliography tend to show its relation with multi-period location models (discrete time models) which are much more useful than single period (continuous time) models. This is proved by Miller et al. (2007) and resumed in three points by Boloori Arabani and

Farahani (2012) which characterizes the achievements that are possible to reach when using multi period location models:

- the appropriate timing of location decision,
- clarifying the best location(s), and
- allowing a firm to better anticipate any favorable/unfavorable fluctuations in market demand in the corresponding time horizon,

In contrast, single-period models (continuous time horizon) do not show the mentioned characteristics. Further statements by Boloori Arabani and Farahani (2012) refer to the advantage of multi-period models over single-period models since in each subordinate planning horizon a decision maker can deal with changing parameters more effectively in comparison with single-period models in which the decision maker is hardly able to cope with the uncertain essence of changing parameters. As proof, the authors point the works of Hale and Moberg (2003), Şahin and Süral (2007), ReVelle and Eiselt (2005), Melo et al. (2009), Klose and Drexl (2005) and Snyder (2006).

Another interesting view is the dynamic DMS (DDMS) from Gendreau et al. (2001) which considers real time redeployment of uEMS ambulances. The model is based in the previous mentioned DSM of the same author in section 2.3.1 Classical Models. In the aforementioned work, the authors implement to the DSM some extra variables required to allow dynamism in the model:

M_{jl}^{t} is the penalty coefficient of reallocating ambulance l of the set of L ambulances from its current site to new site j at time t, and

y_{jl} is a binary variable that takes value 1 if and only if ambulance l is located at j and 0 otherwise.

The new model objective function subtracts to the DSM equation (2.22) the penalization of reallocating an ambulance, as per function (2.41):

$$Maximize \rightarrow \sum_{i=1}^{I} \lambda_i x_i^{k=2} - \sum_{j=1}^{J} \sum_{l=1}^{L} M_{jl}^{t} \times y_{jl} \tag{2.41}$$

Subject to the same equations of the DSM, equations (2.23), (2.24), (2.25), (2.26) and (2.27). The only difference is the variable y_j which now is y_{jl}. The author solves the problem using a tabu search heuristic and applies it to a real-life case.

One extra view that can be made in dynamic models is to incorporates the hypercube theory, with facilities working independently with different busy probabilities, and dividing the time horizon into clusters based on significant change of demands. Rajagopalan et al. (2008) developed the Dynamically Available

Coverage Location (DACL) model for dynamic redeploying facilities to time-varied demands as per the assumptions mentioned.

2.3.3. Robust Optimization Problems

Sometimes the uncertain parameters cannot be described with probability functions, or it is simply not possible to determine a parameter probability because the event might be random. In our study we look for the separation of uEMS calls in medical occurrences, such of cardiac arrest, and road crash occurrences. While cardiac arrest might have a visible correlation with the population demographics over the study area, road crash casualties might not be that obvious, or even be random. The study LIVE by CITTA at University of Porto, demonstrates difficulties in correlating severity of road crash injuries and geographical location, as per Amorim et al. (2014). Therefore, robust optimization comes at hand as a relatively studied tool applied in facility location and that might give a measure of robustness for several solutions probabilities are unknown.

Snyder (2006) studied facility location under uncertainty as a review of the existing work up to 2005. He described robust location problems as the type of problems where no probability information is known about uncertain parameters. Robust problems rely on measures of robustness and usually the two more common ones are *minimax cost* and *minimax regret.* Other types of robustness measures have been studied by Kouvelis and Yu (1997) but appear to be comparatively less common.

To better understand robust measures we rely on the definitions by Snyder (2006). Minimax cost solution is the one that minimizes the maximum cost across all scenarios. The author states that this measure is overly conservative and emphasizes the worst possible scenario. However, minimax cost may be the appropriate measure for a situation in which it is critical for the system to perform well even in the worst case, as is a uEMS. Another measure considers the regret solution which is described as the opportunity loss – the difference between the quality of a given strategy and the quality of the strategy that would have been chosen if we had known what the future holds. These types of models that seek to minimize the maximum regret across all scenarios are the *minimax absolute regret* and *minimax relative regret* models. Moreover, with just the difference of a constant, *minimax cost* problems can be transformed into *minimax regret* problems, and vice-versa.

To understand this type of problems let take the example from Snyder (2006) where:

c_{is} is the objective function coefficient,

z_s^* is the optimal scenario objective,

q_s is the scenario probabilities, and

R_s is the regret measure as per equation (2.42):

$$R_s = \sum_{i=1}^{I} c_{is} \times x_i - z_s^* \;\; \forall s \in S \tag{2.42}$$

Thus the objective function comes as in (2.43):

$$Minimize \rightarrow \sum_{s}^{S} q_s \times R_s \tag{2.43}$$

The focus of the work (placed in robust optimization) is to find ways to solving problems as stated in equation (2.43). Several authors worked in deterministic solutions which are possible for a single facility location, be it 1-median or 1-center, although the 1-center problem proves to be harder - Chen and Lin (1998) presents an $O(n^3)$ algorithm for the 1-median minimax-regret problem while Averbakh and Berman (2000) reaches an $O(n^6)$ algorithm for the minimax-regret for the 1-center problem. Nevertheless, for special cases with certain restrictions the complexity could be reduced.

However for multiple-facility location problems, on general, networks under minimax objectives, the difficulty of the problem significantly raises, and usually solution can only be reached heuristically (Snyder, 2006). Still, Averbakh and Berman (2003), Averbakh and Berman (1997) demonstrate that in some cases minimax can be used and solve deterministically for multiple-facility problems if:

- For the minimax cost problem all the uncertain parameters are set to their upper bounds and solution is reached by obtaining the resulting deterministic problem;
- For the minimax absolute and relative regret problems we solve *m* deterministic problems, in which each of them we set one parameter to its upper bound and the others to their lower bounds, plus one more deterministic problem. With *m* being the number of uncertain parameters.

Moreover, the author proves that a polynomial-time algorithm for the deterministic problem implies a polynomial-time algorithm for the minimax cost and minimax relative regret problems, but not necessarily for the minimax absolute regret problem.

Focusing on the general cases where the points mentioned above are not applicable or might make no sense or have appear not to be of interest, the literature finds several heuristic approaches to solve such problems. For instance Snyder (2006), Serra and Marianov (1998) who solve the minimax cost and minimax regret problems for the p-median problem (PMP), also under scenario-based demand uncertainty. Further, when the number of facilities or ambulances is uncertain, Current et al. (1998) propose a scenario based approach and solve the problem with a general-purpose mixed integer programming (MIP) solver.

2.4. REFERENCES

Abiri, M. B. & Yousefli, A. 2010. An application of probabilistic programming to the fuzzy location-allocation problems. *Int. J. Adv. Manuf. Technol.,* 52**,** 1-7.

Alanis, R., Ingolfsson, A. & Kolfal, B. 2013. A markov chain model for an ems system with repositioning. *Production and Operations Management,* 22**,** 216-231.

Aldrich, C. A., Hisserich, J. C. & Lave, L. B. 1971. An analysis of the demand for emergency ambulance service in an urban area. *American journal of public health,* 61**,** 1156-1169.

Alsalloum, O. I. & Rand, G. K. 2003. A goal-programming model applied to the ems system at Riyadh City, Saudi Arabia.

Alsalloum, O. I. & Rand, G. K. 2006. Extensions to emergency vehicle location models. *Computers & Operations Research,* 33**,** 2725-2743.

Altınel, İ. K., Durmaz, E., Aras, N. & Özkısacık, K. C. 2009. A location–allocation heuristic for the capacitated multi-facility weber problem with probabilistic customer locations. *European Journal of Operational Research,* 198**,** 790-799.

Aly, A. A. & White, J. A. 1978. Probabilistic formulation of the emergency service location problem. *Journal of the Operational Research Society***,** 29, 1167-1179.

Amorim, M., Ferreira, S. & Couto, A. 2014. Linking police and hospital road accident records. *Transportation Research Record: Journal of the Transportation Research Board,* 2432**,** 10-16.

Andersson, T. 2005. Decision support tools for dynamic fleet management. *Doktorsavhandling, Linköpings Universitet, Sverige*.

Andersson, T. & Varbrand, P. 2006. Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society,* 58**,** 195-201.

Andrews, B. H. & Cunningham, S. M. 1995. L. L. Bean improves call-center forecasting. *Interfaces,* 25**,** 1-13.

Arthur, J. L. 1985. Stochastic models in operations research, volume ii. Stochastic optimization (Daniel P. Heyman and Matthew J. Sobel). *SIAM Review,* 27**,** 87-88.

Averbakh, I. & Berman, O. 1997. Minimax regret p-center location on a network with demand uncertainty. *Location Science,* 5**,** 247-254.

Averbakh, I. & Berman, O. 2000. Algorithms for the robust 1-center problem on a tree. *European Journal of Operational Research,* 123**,** 292-302.

Averbakh, I. & Berman, O. 2003. An improved algorithm for the minmax regret median problem on a tree. *Networks,* 41**,** 97-103.

Balinski, M. L. 1965. Integer programming: Methods, uses, computations. *Management Science,* 12**,** 253-313.

Ball, M. O. & Lin, F. L. 1993. A reliability model applied to emergency service vehicle location. *Operations Research,* 41**,** 18-36.

Başar, A., Çatay, B. & Ünlüyurt, T. 2009. A new model and tabu search approach for planning the emergency service stations. *In:* Fleischmann, B., Borgwardt, K.-H., Klein, R. & Tuma, A. (eds.) *Operations research proceedings 2008.* Springer Berlin Heidelberg.

Berman, O. 1981a. Dynamic repositioning of indistinguishable service units on transportation networks. *Transportation Science,* 15**,** 115-136.

Berman, O. 1981b. Repositioning of distinguishable urban service units on networks. *Computers & Operations Research,* 8**,** 105-118.

Berman, O. 1981c. Repositioning of two distinguishable service vehicles on networks., *IEEE Transactions on Systems, Man and Cybernetics,* 11**,** 187-193.

Berman, O. & Odoni, A. R. 1982. Locating mobile servers on a network with markovian properties. *Networks,* 12**,** 73-86.

Bianchi, L., Jarrett, J. & Choudary Hanumara, R. 1998. Improving forecasting for telemarketing centers by arima modeling with intervention. *International Journal of Forecasting,* 14**,** 497-504.

Boloori Arabani, A. & Farahani, R. Z. 2012. Facility location dynamics: An overview of classifications and applications. *Computers & Industrial Engineering,* 62**,** 408-420.

Brimberg, J., Hansen, P., Mladenović, N. & Taillard, E. D. 2000. Improvements and comparison of heuristics for solving the uncapacitated multisource Weber problem. *Operations Research,* 48**,** 444-460.

Brodsky, H. 1990. Emergency medical service rescue time in fatal road accidents. *Transportation Research Record*.

Brodsky, H. 1992. Delay in ambulance dispatch to road accidents. *American Journal of Public Health,* 82**,** 873-875.

Brodsky, H. 1993. The call for help after an injury road accident. *Accident Analysis & Prevention,* 25**,** 123-130.

Brotcorne, L., Laporte, G. & Semet, F. 2003. Ambulance location and relocation models. *European Journal of Operational Research,* 147**,** 451-463.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. & Zhao, L. 2005. Statistical analysis of a telephone call center. *Journal of the American Statistical Association,* 100**,** 36-50.

Cadigan, R. T. & Bugarin, C. E. 1989. Predicting demand for emergency ambulance service. *Annals of Emergency Medicine,* 18**,** 618-621.

Cao, Q., Leggio, K. B. & Schniederjans, M. J. 2005. A comparison between fama and french's model and artificial neural networks in predicting the chinese stock market. *Computers & Operations Research,* 32**,** 2499-2512.

Carson, Y. M. & Batta, R. 1990. Locating an ambulance on the amherst campus of the state university of new york at buffalo. *Interfaces,* 20**,** 43-49.

Channouf, N., L'ecuyer, P., Ingolfsson, A. & Avramidis, A. 2007. The application of forecasting techniques to modeling emergency medical system calls in calgary, alberta. *Health Care Management Science,* 10**,** 25-45.

Charnes, A. & Cooper, W. W. 1959. Chance-constrained programming. *Management Science,* 6**,** 73-79.

Chen, B. & Lin, C.-S. 1998. Minmax-regret robust 1-median location on a tree. *Networks,* 31**,** 93-103.

Church, R. & Velle, C. R. 1974. The maximal covering location problem. *Papers in Regional Science,* 32**,** 101-118.

Cinlar, E. (1972). Superposition of point processes, Stochastic Point Processes: Statistical Analysis, Theory, and Applications, P. Lewis.

Clark, D. E. & Cushing, B. M. 2002. Predicted effect of automatic crash notification on traffic mortality. *Accident Analysis & Prevention,* 34**,** 507-513.

Cooper, L. 1974. A random locational equilibrium problem. *Journal of Regional Science,* 14**,** 47-54.

Current, J., Ratick, S. & Revelle, C. 1998. Dynamic facility location when the total number of facilities is uncertain: A decision analysis approach. *European Journal of Operational Research,* 110**,** 597-609.

Daskin, M. S. 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. *Transportation Science,* 17**,** 48-70.

Daskin, M. S. & Stern, E. H. 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. *Transportation Science,* 15**,** 137-152.

De Maio, V. J., Stiell, I. G., Wells, G. A. & Spaite, D. W. 2003. Optimal defibrillation response intervals for maximum out-of-hospital cardiac arrest survival rates. *Annals of Emergency Medicine,* 42**,** 242-250.

Denton, J. 1995. How good are neural networks for causal forecasting? *Journal of Business Forecasting Methods and Systems,* 14**,** 17-20.

Dessouky, M., Ordóñez, F., Jia, H. & Shen, Z. 2006. Rapid distribution of medical supplies. *In:* Hall, R. (ed.) *Patient flow: Reducing delay in healthcare delivery.* Springer US.

Doerner, K. & Hartl, R. 2008. Health care logistics, emergency preparedness, and disaster relief: New challenges for routing problems with a focus on the austrian situation. *In:* Golden, B., Raghavan, S. & Wasil, E. (eds.) *The vehicle routing problem: Latest advances and new challenges.* Springer US.

Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Karall, M. & Reimann, M. 2005. Heuristic solution of an extended double-coverage ambulance location problem for austria. *Central European Journal of Operations Research,* 13**,** 325-340.

Durkin, M., Mcelroy, J., Guan, H., Bigelow, W. & Brazelton, T. 2005. Geographic analysis of traffic injury in wisconsin: Impact on case fatality of distance to level i/ii trauma care. *Wisconsin Medical Journal,* 104**,** 26-31.

Eisenberg, M. S., Horwood, B. T., Cummins, R. O., Reynolds-Haertle, R. & Hearne, T. R. 1990. Cardiac arrest and resuscitation: A tale of 29 cities. *Annals of Emergency Medicine,* 19**,** 179-186.

Erkut, E., Ingolfsson, A. & Erdogan, G. 2008. Ambulance location for maximum survival. *Naval Research Logistics,* 55**,** 42-58.

Esposito, T. J., Maier, R. V., Rivara, F. P., Pilcher, S., Griffith, J., Lazear, S. & Hogan, S. 1995. The impact of variation in trauma care times: Urban versus rural. *Prehospital and Disaster Medicine,* 10**,** 161-166.

Evanco, W. M. 1999. The potential impact of rural mayday systems on vehicular crash fatalities. *Accident Analysis & Prevention,* 31**,** 455-462.

Fujiwara, M., Kachenchai, K., Makjamroen, T. & Gupta, K. 1988. An efficient scheme for deployment of ambulances in metropolitan bangkok. *Operational Research,* 87**,** 730-741.

Fujiwara, O., Makjamroen, T. & Gupta, K. K. 1987. Ambulance deployment analysis: A case study of bangkok. *European Journal of Operational Research,* 31**,** 9-18.

Galvão, R. D., Chiyoshi, F. Y. & Morabito, R. 2005. Towards unified formulations and extensions of two classical probabilistic location models. *Computers & Operations Research,* 32**,** 15-33.

Gendreau, M., Laporte, G. & Semet, F. 1997. Solving an ambulance location model by tabu search. *Location Science,* 5**,** 75-88.

Gendreau, M., Laporte, G. & Semet, F. 2001. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing,* 27**,** 1641-1653.

Gendreau, M., Laporte, G. & Semet, F. 2005. The maximal expected coverage relocation problem for emergency vehicles. *Journal of the Operational Research Society,* 57**,** 22-28.

Gold, L. S., Fahrenbruch, C. E., Rea, T. D. & Eisenberg, M. S. 2010. The relationship between time to arrival of emergency medical services (ems) and survival from out-of-hospital ventricular fibrillation cardiac arrest. *Resuscitation,* 81**,** 622-625.

Goldberg, J. B. 2004. Operations research models for the deployment of emergency services vehicles. *EMS Management Journal,* 1**,** 20-39.

Gonzalez, R. P., Cummings, G. R., Phelan, H. A., Mulekar, M. S. & Rodning, C. B. 2009. Does increased emergency medical services prehospital time affect patient mortality in rural motor vehicle crashes? A statewide analysis. *The American Journal of Surgery,* 197**,** 30-34.

Gunes, E. & Szechtman, R. A simulation model of a helicopter ambulance service. Simulation Conference, 2005 Proceedings of the Winter, 4-4 Dec. 2005 2005. 7 pp.

Hakimi, S. L. 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research,* 12**,** 450-459.

Hakimi, S. L. 1965. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research,* 13**,** 462-475.

Hale, T. & Moberg, C. 2003. Location science research: A review. *Annals of Operations Research,* 123**,** 21-35.

Henderson, S. G. 2005. Should we model dependence and nonstationarity, and if so how? *Proceedings of the 37th conference on Winter simulation.* Orlando, Florida: Winter Simulation Conference.

Hogan, K. & Revelle, C. 1986. Concepts and applications of backup coverage. *Management Science,* 32**,** 1434-1444.

Hogg, J. M. 1968. The siting of fire stations. *Journal of the Operational Research Society,* 19**,** 275-287.

Ingolfsson, A. The impact of ambulance system status management. Presentation at 2006 INFORMS Conference, 2006.

Jaffray, J.-Y., Jeleva, M., Gains, U. & Paris, C. Information processing under imprecise risk with the hurwicz criterion. 5th international symposium on imprecise probability: theories and applications, 2007. Citeseer, 233-242.

Jarvis, J. P. 1981. Optimal assignments in a markovian queueing system. *Computers & Operations Research,* 8**,** 17-23.

Jia, H., Ordóñez, F. & Dessouky, M. 2007. A modeling framework for facility location of medical services for large-scale emergencies. *IIE Transactions,* 39**,** 41-55.

Jones, A. P. & Bentham, G. 1995. Emergency medical service accessibility and outcomefrom road traffic accidents. *Public Health,* 109**,** 169-177.

Kamenetzky, R. D., Shuman, L. J. & Wolfe, H. 1982. Estimating need and demand for prehospital care. *Operations Research,* 30**,** 1148-1167.

Katz, I. & Cooper, L. 1974. An always-convergent numerical scheme for a random locational equilibrium problem. *SIAM Journal on Numerical Analysis,* 11**,** 683-692.

Kepaptsoglou, K., Karlaftis, M. & Mintsis, G. 2012. Model for planning emergency response services in road safety. *Journal of Urban Planning and Development ASCE,* 138**,** 18-25.

Klose, A. & Drexl, A. 2005. Facility location models for distribution system design. *European Journal of Operational Research,* 162**,** 4-29.

Kolesar, P. & Walker, W. E. 1974. An algorithm for the dynamic relocation of fire companies. *Operations Research,* 22**,** 249-274.

Kouvelis, P. & Yu, G. 1997. *Robust discrete optimization and its applications*, Springer.

Kvålseth, T. O. & Deems, J. M. 1979. Statistical models of the demand for emergency medical services in an urban area. *American Journal of Public Health,* 69**,** 250-255.

Larsen, M. P., Eisenberg, M. S., Cummins, R. O. & Hallstrom, A. P. 1993. Predicting survival from out-of-hospital cardiac arrest: A graphic model. *Annals of Emergency Medicine,* 22**,** 1652-1658.

Larson, R. C. 1974. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research,* 1**,** 67-95.

Li, M. D., Doong, J. L., Chang, K. K., Lu, T. H. & Jeng, M. C. 2008. Differences in urban and rural accident characteristics and medical service utilization for traffic fatalities in less-motorized societies. *Journal of Safety Research,* 39**,** 623-630.

Li, X., Zhao, Z., Zhu, X. & Wyatt, T. 2011. Covering models and optimization techniques for emergency response facility location and planning: A review. *Mathematical Methods of Operations Research,* 74**,** 281-310.

Marianov, V. & Revelle, C. 1996. The queueing maximal availability location problem: A model for the siting of emergency vehicles. *European Journal of Operational Research,* 93**,** 110-120.

Marianov, V. & Serra, D. 1998. Probabilistic, maximal covering location—allocation models forcongested systems. *Journal of Regional Science,* 38**,** 401-424.

Maxwell, M. S., Henderson, S. G. & Topaloglu, H. 2009. Ambulance redeployment: An approximate dynamic programming approach. *Winter Simulation Conference.* Austin, Texas: Winter Simulation Conference.

McConnel, C. E. & Wilson, R. W. 1998. The demand for prehospital emergency services in an aging society. *Social Science & Medicine,* 46**,** 1027-1031.

Melo, M. T., Nickel, S. & Saldanha-Da-Gama, F. 2009. Facility location and supply chain management – a review. *European Journal of Operational Research,* 196**,** 401-412.

Miller, T., Friesz, T., Tobin, R. & Kwon, C. 2007. Reaction function based dynamic location modeling in Stackelberg–Nash–Cournot competition. *Networks and Spatial Economics,* 7**,** 77-97.

Mirchandani, P. B. & Oudjit, A. 1980. Localizing 2-medians on probabilistic and deterministic tree networks. *Networks,* 10**,** 329-350.

Mousavi, S. M. & Niaki, S. T. A. 2013. Capacitated location allocation problem with stochastic location and fuzzy demand: A hybrid algorithm. *Applied Mathematical Modelling,* 37**,** 5109-5119.

Muelleman, R. L. & Mueller, K. 1996. Fatal motor vehicle crashes: Variations of crash characteristics within rural regions of different population densities. *Journal of Trauma and Acute Care Surgery,* 41**,** 315-320.

Nair, R. & Miller-Hooks, E. A case study of ambulance location and relocation. Presentation at 2006 INFORMS Conference, 2006.

Newgard, C. D., Schmicker, R. H., Hedges, J. R., Trickett, J. P., Davis, D. P., Bulger, E. M., Aufderheide, T. P., Minei, J. P., Hata, J. S., Gubler, K. D., Brown, T. B., Yelle, J.-D., Bardarson, B. & Nichol, G. 2010. Emergency medical services intervals and survival in trauma: Assessment of the “golden hour” in a north american prospective cohort. *Annals of Emergency Medicine,* 55**,** 235-246.e4.

Pepe, P. E., Wyatt, C. H., Bickell, W. H., Bailey, M. L. & Mattox, K. L. 1987. The relationship between total prehospital time and outcome in hypotensive victims of penetrating injuries. *Annals of Emergency Medicine,* 16**,** 293-297.

Pons, P. T. & Markovchick, V. J. 2002. Eight minutes or less: Does the ambulance response time guideline impact trauma patient outcome? *The Journal of Emergency Medicine,* 23**,** 43-48.

Rajagopalan, H. K., Saydam, C. & Xiao, J. 2008. A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers & Operations Research,* 35**,** 814-826.

Rajagopalan, H. K., Vergara, F. E., Saydam, C. & Xiao, J. 2007. Developing effective meta-heuristics for a probabilistic location model via experimental design. *European Journal of Operational Research,* 177**,** 83-101.

Repede, J. F. & Bernardo, J. J. 1994. Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky. *European Journal of Operational Research,* 75**,** 567-581.

Revelle, C. & Hogan, K. 1989. The maximum availability location problem. *Transportation Science,* 23**,** 192-200.

Revelle, C. S. & Eiselt, H. A. 2005. Location analysis: A synthesis and survey. *European Journal of Operational Research,* 165**,** 1-19.

Rosenhead, J., Elton, M. & Gupta, S. K. 1972. Robustness and optimality as criteria for strategic decisions. *Journal of the Operational Research Society,* 23**,** 413-431.

Şahin, G. & Süral, H. 2007. A review of hierarchical facility location models. *Computers & Operations Research,* 34**,** 2310-2331.

Sánches-Mangas, R., García-Ferrer, A., Juan, A. D. & Arroyo, A. M. 2010. The probability of death in road traffic accidents. How important is a quick medical response? *Accident Analysis and Prevention,* 42**,** 1048-1056.

Savas, E. S. 1969. Simulation and cost-effectiveness analysis of new york's emergency ambulance service. *Management Science,* 15**,** B-608-B-627.

Saydam, C. & Mcknew, M. 1985. Applications and implementation a separable programming approach to expected coverage: An application to ambulance location. *Decision Sciences,* 16**,** 381-398.

Schilling, D., Elzinga, D. J., Cohon, J., Church, R. & Revelle, C. 1979. The team/fleet models for simultaneous facility and equipment siting. *Transportation Science,* 13**,** 163-175.

Serra, D. & Marianov, V. 1998. The p-median problem in a changing network: The case of Barcelona. *Location Science,* 6**,** 383-394.

Setzler, H., Saydam, C. & Park, S. 2009. Ems call volume predictions: A comparative study. *Computers & Operations Research,* 36**,** 1843-1851.

Sheppard, E. S. 1974. A conceptual framework for dynamic location - allocation analysis. *Environment and Planning A,* 6**,** 547-564.

Shiah, D.-M. & Chen, S.-W. Ambulance allocation capacity model. e-Health Networking, Application and Services, 2007 9th International Conference on, 19-22 June 2007 2007. 40-45.

Siler, K. F. 1975. Predicting demand for publicly dispatched ambulances in a metropolitan area. *Health Services Research,* 10**,** 254-263.

Simpson, N. C. & Hancock, P. G. 2009. Fifty years of operational research and emergency response. *Journal of the Operational Research Society,* 60**,** S126-S139.

Snyder, L. V. 2006. Facility location under uncertainty: A review. *IIE Transactions,* 38**,** 547-564.

Sorensen, P. & Church, R. 2010. Integrating expected coverage and local reliability for emergency medical services location problems. *Socio-Economic Planning Sciences,* 44**,** 8-18.

Toregas, C. & Revelle, C. 1972. Optimal location under time or distance constraints. *Papers in Regional Science,* 28**,** 133-144.

Toregas, C., Swain, R., Revelle, C. & Bergman, L. 1971. The location of emergency service facilities. *Operations Research,* 19**,** 1363-1373.

Trudeau, P., Rousseau, R., Ferland, J. A. & Choquette, J. 1988. An operations research approach for the planning and operation of an ambulance service.

Tych, W., Pedregal, D. J., Young, P. C. & Davies, J. 2002. An unobserved component model for multi-rate forecasting of telephone call demand: The design of a forecasting support system. *International Journal of Forecasting,* 18**,** 673-695.

Valenzuela, T. D., Roe, D. J., Cretin, S., Spaite, D. W. & Larsen, M. P. 1997. Estimating effectiveness of cardiac arrest interventions a logistic regression survival model. *Circulation,* 96**,** 3308-3313.

Valenzuela, T. D., Roe, D. J., Nichol, G., Clark, L. L., Spaite, D. W. & Hardman, R. G. 2000. Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos. *New England Journal of Medicine,* 343**,** 1206-1209.

Valinsky, D. 1955. Symposium on applications of operations research to urban services—a determination of the optimum location of fire-fighting units in new york city. *Journal of the Operations Research Society of America,* 3**,** 494-512.

Waalewijn, R. A., De Vos, R., Tijssen, J. G. P. & Koster, R. W. 2001. Survival models for out-of-hospital cardiopulmonary resuscitation from the perspectives of the bystander, the first responder, and the paramedic. *Resuscitation,* 51**,** 113-122.

Wen, M. & Iwamura, K. 2008a. Facility location–allocation problem in random fuzzy environment: Using (α,β) -cost minimization model under the hurewicz criterion. *Computers & Mathematics with Applications,* 55**,** 704-713.

Wen, M. & Iwamura, K. 2008b. Fuzzy facility location-allocation problem under the hurwicz criterion. *European Journal of Operational Research,* 184**,** 627-635.

Wesolowsky, G. O. 1973. Dynamic facility location. *Management Science,* 19**,** 1241-1248.

Wilde, E. T. 2013. Do emergency medical system response times matter for health outcomes? *Health Economics,* 22**,** 790-806.

Winters, P. R. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science,* 6**,** 324-342.

Zhang, O., Mason, A. & Philpott, A. Simulation and optimisation for ambulance logistics and relocation. Presentation at the INFORMS 2008 Conference, 2008.

Zhou, J. & Liu, B. 2007. Modeling capacitated location–allocation problem with fuzzy demands. *Computers & Industrial Engineering,* 53**,** 454-468.

Zhu, Z., Mcknew, M. A. & Lee, J. 1992. Effects of time-varied arrival rates: An investigation in emergency ambulance service systems. *Proceedings of the 24th conference on Winter simulation.* Arlington, Virginia, USA: ACM.

Zwerling, C., Peek-Asa, C., Whitten, P. S., Choi, S.-W., Sprince, N. L. & Jones, M. P. 2005. Fatal motor vehicle crashes in rural and urban areas: Decomposing rates into contributing factors. *Injury Prevention,* 11**,** 24-28.

3. Road Safety and The Urban Emergency Medical Service (uEms): Strategy Station Location[8]

Marco Amorim[9], Sara Ferreira[9], Antonio Couto[9]

Abstract

This paper provides a methodology on how to contribute to road safety by improving the vehicle response of urban emergency medical services (uEMS) using road safety investments. The methodology embodies two steps. The first step includes a demand assessment through a model that calculates the frequency of urban emergency events and their priority in a spatial area per different population demographics and urban characteristics. These events are categorized by type, which we separated as road crashes, cardiac arrests and other emergency events that require the dispatch of an ambulance with a medical team. The second step proposes an optimization model developed to maximize the uEMS vehicle coverage, considering road crashes by locating ambulance stations in the urban area, giving priority to high-priority emergencies and strategically double-covering road crashes. The applicability and practical interest of the proposed models are proven by applying them to the real-world case of the metropolitan area of Porto, where data on the emergency service response, land use and population demographics are available. The final conclusions indicate that areas prone to high volumes of traffic and fast roads lead to higher volumes of road crash emergencies, and stations should be located closer to fast roads. Moreover, further investigations should entail micromanagement improvements to assist in road crashes.

Keywords: Emergency medical service, static optimization, road safety, demand modelling

---

8 Amorim, M., Ferreira, S., & Couto, A. (2017). Road safety and the urban emergency medical service (uEMS): Strategy station location. Journal of Transport & Health, 6, 60-72.

Final version at: https://doi.org/10.1016/j.jth.2017.04.005

9 CITTA, University of Porto – Faculty of Engineering, Porto, Portugal

3.1. INTRODUCTION

Around the world, 1.24 million people die each year on roads, and between 20 and 50 million sustain non-fatal injuries. Half of those who die are 'vulnerable road users': pedestrians, cyclists and motorcyclists. Young adults aged between 15 and 44 years account for 59% of global road traffic deaths.

This world problem, although severe, has straightforward solutions that can be implemented *a priori* through effective interventions such as urban and transport planning, designing safer roads, requiring independent road safety audits for new construction projects, and setting and enforcing internationally harmonized laws (WHO, 2011).

In countries where such interventions have been implemented, it is expected that thorough additional *a priori* actions will have a less visible social impact in preventing road crashes or even improving their outcomes. Therefore, for the remaining road crashes that cannot be prevented, the available solutions are either to improve vehicle safety measures or to intervene *a posteriori* by improving the emergency medical services and minimize the socioeconomic impact of road crashes.

In fact, post-crash response is pillar 5 of the WHO (2011) global plan for road safety for the decade of action of 2011-2020. The post-crash response is divided into several activities; the last one, Activity 7, explicitly encourages research and development on improving the post-crash response. One way of meeting this plan is to put research effort into the vehicle response of emergency medical services, e.g., by assessing ambulance station locations, a well-known station location problem.

Station location problems usually aim to cover all the influence areas in a manner such that there is no single demand unit that cannot be reached within a specific time window – a maximum response time limit. The optimization of emergency responses (ER) and/or emergency medical services (EMS) is tightly connected to operational research (OR) and usually is the driver that conducts OR trends. The common topics of study are generally divided into three major groups, Urban Services (we will call this urban emergency medical response, or uEMS), Disaster Services and Hazard Specific.

The foundational stream of research for emergency response dates back to the year of 1955 with fire station location planning studies by Valinsky (1955). Additionally, Hogg (1968) together with Savas (1969) filled the base archetypes for this theme, with the latter focusing on ambulance services rather than fire station allocation as per the two former ones.

However, the two most relevant works, which truly sparked the OR community interest in EMR, were those of Toregas et al. (1971) and Church and Velle (1974). The former presents a solution to solve the location set covering problem (LSCP), making sure all demand is covered within a maximum time or distance radius, and the latter provide a solution for the maximal coverage location problem (MCLP) that overcomes the resources limitation of the problem of Toregas et al. (1971).

The classical interpretation of the station location problem, in particular emergency urban services, is quickly being overcome by uncertainty approaches, i.e., those leading to double coverage, scenario approaches, stochastic and robust optimization problems and dynamic location. Some have already been presented as extensions of the two classic models.

Focusing on the fact that once a station is called for service, demand points under its coverage are no longer covered, Daskin (1983), Daskin and Stern (1981) and Hogan and ReVelle (1986), (1989) account for the busy probability and reliability of a station. The former ones (Daskin, 1983, Daskin and Stern, 1981) solve the maximum expected covering location problem (MEXCLP), whereas the latter moves into a maximum availability location problem (MALP).

Another important model for double coverage is the Double Standard Model (DSM), which aims to allocate stations among potential sites to fully cover the study region within a longer distance standard while maximizing the coverage within a shorter distance standard. Gendreau et al. (1997) proposed a metaheuristic (tabu search) to solve this problem and eventually developed their own DSM.

Urban Emergency Medical Services (uEMS) is a station location problem; however, there are particularities that distinguish them from the usual location problems. Whereas underestimating or overestimating the earlier-mentioned station location problems will have a mostly monetary impact, in uEMS problems, there is also a social impact, and a bad decision can lead to, e.g., higher response times, which may seriously reduce the survival probability of the victims to be rescued. For instance, Sánchez-Mangas et al. (2010) indicated that a reduction of 10 minutes in the emergency response time could result in a 30% reduction of road crash fatalities.

Nevertheless, demand characterization posts an important role when defining station locations. The emergency medical service is requested in a random way, contrary to other station problems; for example, in a food supplier, service market studies allow for demand estimation. In uEMS, the demand is not known, and the most common way to predict it is by pattern analyses, which consider the local history of the study area. Various statistical models have been produced to predict such demand (Setzler et al., 2009, Channouf et al., 2007, McConnel and Wilson, 1998, Cadigan and Bugarin, 1989, Kvålseth and Deems, 1979, Siler, 1975, Aldrich et al., 1971). However, what better separates uEMS from other station problems is the fact that in uEMS, it is not just the demand that is stochastic but also its priority. Recalling the food supplier example, whereas we know a posteriori that certain products are perishable (thus we know their priority), in uEMS, the priority of an occurrence is assessed only during the emergency event. In some cases, the severity of the victim is found only when the medical team arrives at the incident site. Therefore, two questions arise: how often will the service be required and, what is the likelihood of a request, within certain characteristics, being of higher priority?

The previous statements about demand on uEMS address general types of emergency events; however, when the focus falls on road crashes, the demand proves to be even more complex. Although there are many studies focused on modelling and predicting the occurrence of a road crash, with some identifying hot spots, such as the ones most recently produced by Geedipally and Lord (2010), Lan and Persaud (2011), Couto and Ferreira (2011), Ferreira and Couto (2013a), (2013b), there is not much work on the definition and prediction of road crashes outcomes per victim.

In the last few years, as Erkut et al. (2008) note, the research direction has been to substitute the concept of coverage with concepts that account for survival probabilities – making a comparison with other types of station locations, such as food suppliers; the idea is to prioritize the most important clients knowing that demand will not be affected because there is no other competitor in the marketplace.

One recent study implementing patient survival is that of McCormack and Coates (2015). The authors prove that it is possible to increase cardiac arrest victims' survival without the need of additional resources. Nevertheless, the proposed model divides medical emergencies only into cardiac arrest or other types. No considerations are given to road crash victims and their specificities. In contrast, Kepaptsoglou et al. (2012) focused their work on a model for planning uEMS for the special case of road crashes; however, they address this type of emergency solely, and disregard other types of urban medical emergencies. Knight et al. (2012) propose a Maximal Expected Survival Location Model for Heterogeneous Patients where a decaying survival function is used for cardiac arrests and step functions for other types of medical emergencies. A weight parcel is added to capture emergency type priority. The work conclusions stress that multiple outcome measures lead to lives saved compared to hard targets and/or a single patient-type optimization goals.

Survival functions and the heterogeneous nature of the population lead to a higher number of lives saved (Knight et al., 2012, Mayorga et al., 2013). Nevertheless, no complete integrated methodology exists that addresses survival probabilities and the impact of different types of urban medical emergencies and implements demand prediction. To do so, we address the Emergency Medical Service, EMS, with a focus on Urban Service, uEMS, investigating how to better manage this service in a long-term scenario, with a focus on the ambulance station location. The objective is to cover all the demand within a maximum time limit so that all events can be answered within a reasonable response time, reinforce the demand with higher priority, and treat road crashes separately. In the end, we provide a methodology that can be used together by uEMS and road safety authorities to understand how road safety investments can be allocated in the uEMS.

3.2. METHODOLOGY

We propose a model to locate the uEMS ambulance stations considering that some medical emergencies occurrences have a higher priority than others (particularly cardiac arrest emergencies) and that road crashes can be separated from other cases of uEMS. The special requirement of a road crash is twofold. First, there might be a requirement of multiple ambulances if there is more than one injury, and second, a road crash might cause traffic jams that interrupt part of the road network, influencing the response time in that area. Within this consideration, a two-step methodology was developed. First, a generalized linear model is proposed to assess a demand indicator for each type of event and its correlation with land use, demographics and social factors of a certain area. Second, an ambulance station location model is proposed to analyze possible improvements toward road safety.

3.2.1. GENERALIZED LINEAR MODEL

For the first step, a generalized linear model is used with three main components: A random component to address the demand randomness, a linear predictor to capture the empirical relation between demand and the predictors, and a link function to adjust the non-linearity of the observations.

For the random component, the demand indicator is assumed to follow either a Poisson distribution or a Negative Binomial distribution. Many authors usually assume that the EMS demand follows a Poisson process, based on theoretical proofs (Henderson, 2005) or empirical evidence, e.g., the works of Brown et al. (2005), Gunes and Szechtman (2005) and Zhu et al. (1992). However, when we do not assume that all cases have an equal probability of experiencing a rare event, but rather that events may cluster (because the variance is larger than the mean), the Negative Binomial is preferable and keeps the essential Poisson properties. Nevertheless, Henderson (2005) proves that Gaussian and Poisson random fields have an important role to play in the simulation models of spatial phenomena.

For each type of event, our hypothesis assumes that the linear predictor, η_i , follows the local population demographics, the land use and social factors as per equation (3.1):

$$\eta_i = \sum_{demo=1}^{DEMO} x_{i,demo}\beta_{demo} + \sum_{lu=1}^{LU} x_{i,lu}\beta_{lu} + \sum_{social=1}^{SOCIAL} x_{i,social}\beta_{social} \quad (3.1)$$

where

x_{ij} is the value of the *j*th variable for the *i*th observed section,

β_j is a vector of unknown parameters to be estimated,

demo are the variables in the set *DEMO* that represent the population demographics,

lu are the variables in the set *LU* that represent the land use, and

social are the variables in the set *SOCIAL* that represent the social factors.

With these linear predictors, the demand indicators are determined by equation (3.2).

$$d_i^{type} = \exp(\eta_i) \quad (3.2)$$

We propose three different applications for this model: the cardiac arrest model, the road crash model and the non-cardiac arrest and non-road crash model.

With the demand characterized by type, the next step is to assess the share of high priority events at each city section. As noted previously, cardiac arrests always require the quickest possible response; thus, the priority characterizer model falls exclusively in the non-cardiac arrest events. With this goal, we developed a hotspot model, equation (3.3), which allows us to define if a section is of higher priority regarding non-cardiac arrest events using a logistic distribution:

$$\text{Prob}\left(Y=1 \mid x_{ij}\right)=\frac{\exp\left(\gamma_j x_{ij}\right)}{1+\exp\left(\gamma_j x_{ij}\right)}, \tag{3.3}$$

where

Y=1 defines section i as a hotspot (e.g., the sections with high percentage of priority events) and

γ_j is a vector of unknown parameters to be estimated.

3.2.2. STATION LOCATION MODEL

For the second step – the ambulance station location model – we propose an optimization model based on the double standard location model (DSM) developed by Gendreau et al. (1997), using different maximum response time thresholds as a way to implement survival step functions for the different emergency types (Knight et al., 2012). This is the same formulation used in the model for planning emergency response services in road safety by Kepaptsoglou et al. (2012). The base of this model is as follows:

$$\text{maximize} \rightarrow \sum_{s \in S} d_s^{type} x_{s,r_2}, \tag{3.4}$$

subject to

$$\sum_{j \in J} y_j R_{s,j}^{r_1} \geq 1, \quad \forall s \in S, \tag{3.5}$$

where the decisions variables are as follows:

x_{s,r_2} is a decision variable that has a value of 1 if section *s* of the set of sections *S* is covered within a time coverage of r_n, and 0 otherwise and

y_j is a decision variable that has a value of 1 if a station is set in the available station position *j* of the set of available stations positions *J* and 0 otherwise.

In addition,

d_s^{type} is the demand indicator by *type* for section *s*,

$R_{s,j}^{r_1}$ is equal to 1 if section *s* is covered by *j* within time coverage r_1

The set of time coverage is $r[r_1, r_{2, \ldots,} r_N]$ where $r_1 \geq r_2 \geq \ldots \geq r_N$.

Equation (3.4) maximizes the demand that is covered by the more restricted time coverage r_2, whereas equation (3.5) enforces all demand to be covered by time coverage r_1.

We propose an extension of the latter model by dividing the maximizing function by the type of emergency and adding double coverage. The expanded Double Standard Double Coverage location model per type of demand ($\text{DSDCM}_{\text{type}}$) transforms the objective function (3.4) into the following:

$$\text{maximize} \rightarrow C^{cardiac} + C^{road} + C^{general} \tag{3.6}$$

with

$$C^{cardiac} = \sum_{s \in S} d_s^{cardiac} x_{\mathrm{s,r_3}} \tag{3.7}$$

$$C^{road} = \sum_{s \in S} d_s^{road} x_{\mathrm{s},double\ r_1} \tag{3.8}$$

$$C^{general} = \sum_{s \in S} d_s^{non\ cardiac} x_{\mathrm{s,r_2}} W_s \tag{3.9}$$
,

and expanding the constraints to the following:

$$\sum_{j \in J} y_j R_{s,j}^{r_1} \geq 1, \quad \forall s \in S \tag{3.10}$$

$$\sum_{j \in J} y_j R_{s,j}^{r_3} \geq x_{s,r_3}, \quad \forall s \in S \tag{3.11}$$

$$\sum_{j \in J} y_j R_{s,j}^{r_1} \geq 2 x_{s,\text{double } r_1}, \quad \forall s \in S \tag{3.12}$$

$$\sum_{j \in J} y_j R_{s,j}^{r_2} \geq x_{s,r_2}, \quad \forall s \in S \tag{3.13}$$

$$\sum_{j \in J} y_j \leq n \tag{3.14}$$
,

where the decisions variables are as follows:

x_{s,r_2} , which has a value 1 if section *s* of the set of sections *S* is covered within a time coverage of r_2 and 0 otherwise;

$x_{s,double\ r_1}$, which has a value of 1 if section *s* of the set of sections *S* is double covered within a time coverage of r_1 and 0 otherwise;

x_{s,r_3} , which has a value of 1 if section *s* of the set of sections *S* is covered within a time coverage of r_3 and 0 otherwise; and

y_j , which has a value of 1 if a station is set in the available station position *j* of the set of available stations positions *J* and 0 otherwise.

The parameters are as follows:

d_s^{type} is the demand of section *s* calculated by equations (3.1) and (3.2) for each considered *type* of demand, and

W_s is the percentage of demand with higher priority in section *s* based on equation (3.3)

The set of time coverage is $r[r_1, r_2, r_3]$ with $r_1 \geq r_2 \geq r_3$, and *double* r_1 indicates that there is at least two stations covering the section within r_1.

Equation (3.7) maximizes the time coverage of the cardiac arrest demand points within a time limit (in accordance with the findings of Erkut et al. (2008) and the notion of survival step functions from Knight et al. (2012)), equation (3.8) maximizes the double coverage of the demand from road crashes, and equation (3.9) maximizes the time coverage of any other (non-cardiac arrest) high priority demands within a relaxed time limit.

Equation (3.10) makes sure all demand is covered by the overall time coverage limit. Equation (3.11) and equation (3.13) guarantee that the higher priority emergencies and cardiac arrests are scored in the objective function only if covered by the correspondent time coverage limits. Finally, equation (3.12) ensures that road crashes are scored in the objective function only if they are at least double covered.

The maximum number of stations to be set is controlled by equation (3.14). Additionally, we want to reduce the number of stations; thus, a minmax optimization model is required. To avoid adding this extra complexity, we opt to build a simple algorithm that will run the optimization model for an infinite set of integer values of n = [1, 2, 3, …, ∞] until a solution is found. This algorithm will increase the optimization complexity linearly, depending on the size of n. At most, n will be as large as the number of possible station locations.

To assess the possible contributions of uEMS to road safety, several scenarios with and without the incorporation of the road crash objective will be built to compare the solutions. A change in the station location together with the score of the demand indicator will produce ground to infer confirmation on this work claim. Furthermore, the demand assessment models will give directions on how uEMS demand, by type and severity, can be correlated with the several demographic, land use and social factors.

3.3. ASSESSMENT OF PORTO CITY - DATA

The proposed methodology was applied to the city of Porto. We defined the demand indicator as the number of uEMS events for the period of 1 year, divided by three types: cardiac arrests, road crashes and other. This division comes from the fact that cardiac arrests are always a high priority emergency event (De Maio et al., 2003). Road crash emergency occurrences have advantages if they are double covered (i.e., two or more ambulance stations are within the maximum response time). The other type attempts to cover all the remaining cases. A higher level of detail could be achieved; however, research on how time might influence medical emergencies other than cardiac arrests is not detailed enough or has proven to be non-significant (Pons and Markovchick, 2002, Newgard et al., 2010, Culley et al., 1994).

The data is composed of the following:

- The INEM (National Institute of Medical Emergency - Portugal) emergency events within Porto during the year of 2011 which has information on the time, type, local and priority of each event;
- The INE (National Institute of Statistic – Portugal) 2011 census which has the information of the population demographics, number of buildings and social demographics with a resolution of 441 sections coded in GIS;
- The Urban Atlas database by the European Environment Agency which provides pan-European comparable land use and land cover data for Large Urban Zones with more than 100.000 inhabitants as defined by the Urban Audit. It has information on the land use of the city such as Continuous and dense urban fabric, industrial, commerce and public units, fast transit roads and sports and leisure stations (Meirich, 2008).

With these three datasets, three separated types of variables plus the independent term (demand indicator), to be used in equation (3.1) and per INE geographic sections, are presented:

- Population Demographic variables – Population, population density, population age, population gender and population per dwelling.
- Land Use variables – Section area, percentage of high dense urban areas, fast transit areas, other transit areas, percentage of employment areas, percentage of areas without use, number of dwelling buildings and number of non-dwelling buildings.
- Social variables – percentage of population employed, percentage of population without economic activity and percentage of retired population.

- Demand – number of cardiac arrest events, number of road crash events, number of non-cardiac arrest non-road crash events, and percentage of high priority non-cardiac arrest events.

Nevertheless, for the road crash emergency indicator, traffic is an essential predictor (Ferreira and Couto, 2013a, Ferreira and Couto, 2013b, Lan and Persaud, 2011, Couto and Ferreira, 2011, Geedipally and Lord, 2010). Unfortunately, these data were not available for our year of study or for all the observations we have. Thus, we used a study from Ferreira (2010) to estimate the annual average daily traffic, AADT.

Moreover, it is important to stress that for these kinds of predictive models, usually it is good practice the use of 3 years of observations (Aguero-Valverde and Jovanis, 2006, Bonneson, 2010, Hauer, 1997). For this study, there were observations available for a period of only one year. Therefore, for future research, we advise a recalculation of the models parameters.

The characteristics of all the considered variables are presented in Table 2 and a demand heat map of the study area is represented in Figure 4.

Table 2. Variables statistics

Covariate	Nº Sections	Minimum	Maximum	Average	Std. Deviation
Area (hectare)	441	1.181	69.477	9.409	9.466
dPopulation: Population density (/1000/hectare)	441	0.000	0.359	0.092	0.059
Young population density (/1000/hectare)	441	0.000	0.084	0.015	0.011
Adult population density (/1000/hectare)	441	0.000	0.240	0.056	0.037
dElderPopulation: Elder population density (/1000/hectare)	441	0.000	0.084	0.021	0.014
Ratio of agriculture area	441	0.000	0.598	0.043	0.109
Ratio of construction area	441	0.000	0.752	0.009	0.053
AreaUrban80: Ratio of ADSS* >80%	441	0.000	1.000	0.421	0.302
Ratio of ADSS* 50%-80%	441	0.000	0.944	0.173	0.228
Ratio of ADSS* 30%-50%	441	0.000	0.277	0.003	0.019
Ratio of ADSS* 10%-30%	441	0.000	0.595	0.020	0.068
Ratio of ADSS* <10%	441	0.000	0.019	0.000	0.001
FastRoads: Fast transit extension (km)	441	0.000	1.384	0.007	0.073
Ratio of forest area	441	0.000	0.211	0.002	0.018
Ratio of green urban area	441	0.000	0.631	0.032	0.090
AreaWork: Ratio of commerce area	441	0.000	0.917	0.124	0.158
NoLandUse: Ratio of non-use area	441	0.000	0.240	0.009	0.031
Other roads extension (km)	441	0.186	5.059	1.197	0.736
ResidentBuildings: Residential buildings (/1000)	441	0.000	0.293	0.086	0.056
Semi residential buildings (/1000)	441	0.000	0.104	0.013	0.017
NonResidentBuildings: Nonresidential buildings (/1000)	441	0.000	0.044	0.001	0.004
Population (/1000)	441	0.000	1.527	0.538	0.189
Ratio females	441	0.000	0.649	0.545	0.036
Ratio under 20 years	441	0.000	0.292	0.162	0.044

Ratio adults	441	0.000	0.865	0.598	0.060
Ratio over 64 years	441	0.000	0.441	0.237	0.071
Ratio retired people	441	0.000	0.500	0.272	0.076
PopWorking: Ratio population with job	441	0.000	0.759	0.376	0.081
PopNoEcoActiv: Ratio population no economic activity	441	0.000	0.651	0.429	0.083
Traffic (AADT)	441	380	63 205	13 325	10 604

Note:*ADSS - Average degree of soil sealing

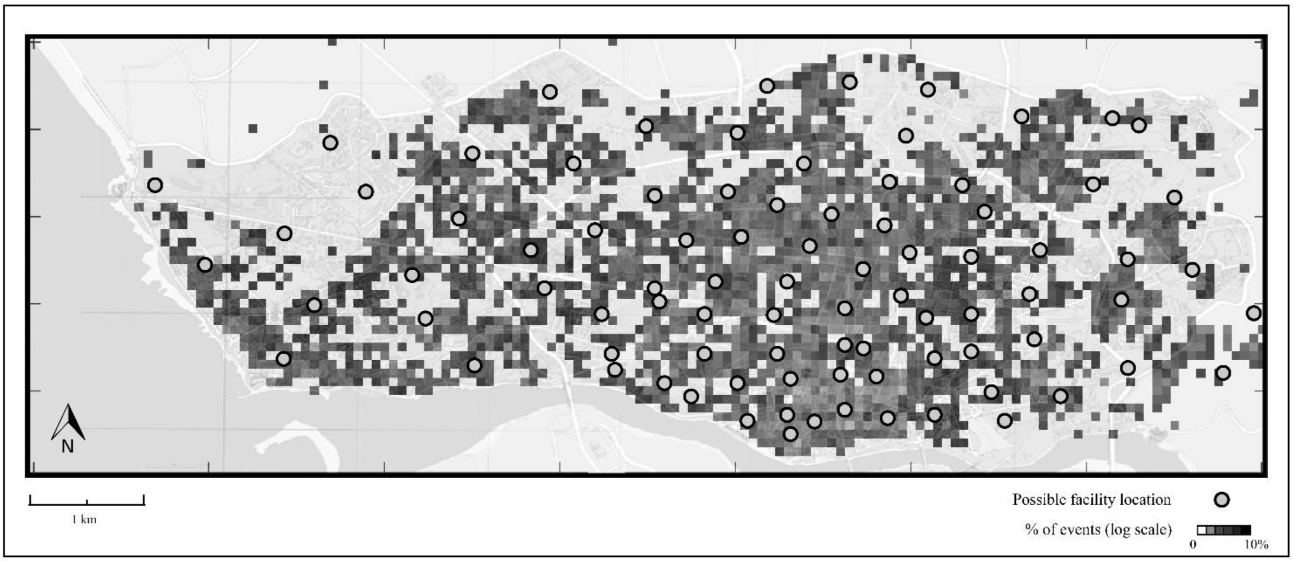


Figure 4. Porto heat map of emergency events and possible station location.

3.4. Porto City Results and Analysis

3.4.1. Assessment of the Demand Indicators

For each demand predictor model, the beta parameters for equation (3.1) were calibrated using each INE geographic section as an observation and tested with the different available variables.

The results are presented in Table 3 for the cardiac arrest demand indicator (Model C), Table 4 for the road crash demand indicator (Model R) and Table 5 for the non-cardiac and non-road crash demand indicator (Model N).

Within all available variables (Table 2) the significant ones are:

- AreaUrban80 - Predominant residential use: areas with a high degree of soil sealing, independent of their housing scheme (single family houses or high rise dwellings, city centre or suburb). Included are downtown areas and city centres, and central business districts (CBD) as long as there is partial residential use. In percentage of the section area;
- dPopulation – Population density in habitants per hectare in the section;
- Population – habitants in the section;
- PopNoEcoActiv – Ratio of population without an economic activity per total population in the section;
- Area – Area of the section in kilometres square;
- PopWorking – Ratio of population with a job per total population in the section.
- NoLandUse - Areas in the vicinity of artificial surfaces still waiting to be used or re-used. The area is obviously in a transitional position, “waiting to be used”. No actual agricultural or recreational use. Areas where the street network is already finished, but actual erection of buildings is still not visible. In percentage of the section area;
- FastRoads - Roads defined as “motorways” in the COTS navigation data, and motorway rest and service areas and parking areas, only accessible from the motorways. In linear meters;
- NonResidentBuildings – Number of buildings where the main use is non-residential;
- ResidentBuildings – Number of building that are exclusively for residential use;
- AreaWork - Industrial, commercial, public, military or private units. The administrative boundaries of the production or service unit are mapped, including associated features larger than the MinMU (e.g. sports areas or transport structures). In percentage of the section area;
- dElderPopulation – Elder population (>64 years old) density in habitants per hectare in the section.
- Ln(Traffic) – Napierian logarithm of the estimated Annual average daily traffic (AADT).

The three models present a significance higher than 99% for the omnibus test, which compares the fitted model with the intercept-only model. In addition, all three models pass the Lagrange multiplier test within a significance of 99%. This means that the negative binomial model is significantly different from the Poisson model. The binomial model's ancillary (dispersion) parameter, k, is different from 0, so we conclude that there is over dispersion, and the negative binomial model is preferable over the Poisson model, supporting our hypothesis that uEMS road crashes and cardiac arrest events may cluster. The Likelihood Ratio Chi-Square for each of the models is 54 for Model C, 620 for Model R and 91 for Model N. Although these values are lower than expected, the significance of each explanatory variable being different from zero is over 95%.

The variables were tested for correlation using Person’s correlation and according to Evans (1996). We found that all our variables but two can be verbally described as having a weak correlation effect. Nonresidential buildings and residential buildings are moderately correlated (absolute value of *r* of 0.542).

This moderate correlation might be explained by the fact that Porto is a homogeneous city and has many buildings that have multiple uses (residential and nonresidential use); thus, each variable might capture some of the other variable effect. Unfortunately, no better variables were available to describe the land use as residential or nonresidential; therefore, we were obligated to keep these two.

Finally, for each model, sensitivity and residual analyses were conducted. For each of the demand indicators, the predictors proved to be stable when removing or adding new variables (if correlation was not present in the predictors). For the residual analysis, a cumulative residual (CURE) method (Hauer and Bamfo, 1997) was applied. Globally, the CURE analysis showed that the estimated models fit well the data with respect to each individual explanatory variable. Only for some of the higher values of few variables, such as population density, is the cumulative residuals curve larger than the ±2.0σ boundary and does not end near the zero-residual line.

With these models, it was possible to calculate the demand indicator vector d_s^{type} for equation (3.7) using Model C, equation (3.8) using Model R and equation (3.9) using the sum of the results of Model R and Model N.

Table 3. Cardiac arrest events parameters estimation, Negative Binomial Model C.

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-2.911	0.554	-3.997	-1.824	27.561	1	0.000
AreaUrban80	1.110	0.297	0.529	1.692	13.996	1	0.000
dPopulation	-6.470	1.646	-9.696	-3.244	15.451	1	0.000
Population	1.039	0.444	0.169	1.909	5.480	1	0.019
PopNoEcoActiv	5.252	1.016	3.261	7.242	26.732	1	0.000

Table 4. Road crash events parameters estimation, Negative Binomial Model R.

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercepto)	-20.772	1.405	-23.525	-18.018	218.645	1	0.000
dPopulation	-7.478	1.415	-10.251	-4.705	27.934	1	0.000
NoLandUse	-5.969	2.541	-10.949	-0.989	5.519	1	0.019
Ln(Traffic)	2.271	0.143	1.990	2.552	251.165	1	0.000
fastRoads	2.260	0.686	0.916	3.605	10.853	1	0.001

Table 5. Non-cardiac arrest and non-road crash events parameters estimation, Negative Binomial Model N.

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	1.738	0.322	1.107	2.368	29.180	1	0.000
NonRes.Buildings	40.081	14.775	11.123	69.039	7.359	1	0.007
PopNoEcoActiv	4.134	0.621	2.918	5.351	44.348	1	0.000
ResidentBuildings	9.148	3.223	2.832	15.465	8.057	1	0.005
dPopulation	-1.665	0.845	-3.322	-0.008	3.881	1	0.049
Population	1.108	0.275	0.570	1.646	16.288	1	0.000

Regarding the correlation between the demand indicators and the demographic, land use and social factor predictors, population has a positive influence on its growth. Nevertheless, for road crash emergencies, motorway extensions and traffic substitute population as the growing factor. Clearly, areas with extended motorways and/or high volumes of traffic are likely to lead to higher rates of this type of emergency. Moreover, areas with a dense population contribute negatively for the demand indicators. It is important to remember that areas with a high population density are mostly residential areas; thus, the working share of the population is not at these locations during most of the day. Rather, they occupy city areas where commercial and economic activities exist.

Cardiac arrest emergencies show a strong correlation with areas with a high population with no economic activity. This indicates that older people or people unable to work, most likely due to health of physical problems, are more prone to cardiac arrests.

Regarding road crashes, it is important to indicate that higher traffic volume areas in an urban context mean that not only vehicle crashes exist but also vehicle-pedestrian collisions make an important account – crossovers.

For Vector W_s of equation (3.9), it is essential to assess the percentage of high priority non-cardiac arrest events. To do so, the probability of each section being a hotspot (Y=1 if percentage of higher priority non-cardiac arrest emergencies ≥ percentage of higher priority non-cardiac arrest emergencies of the whole study area) is calculated as per equation (3.3). The results, in Table 6, show that emergencies of higher priority are correlated with less dense urban areas and areas where the elder population is more concentrated. Clearly, elderly people are more susceptible to severe medical conditions due to their fragile nature. Dense urban areas and working areas have a higher concentration of healthier and younger people; thus, they are less likely to have severe emergency medical problems.

Furthermore, the results are converted to a coefficient of high priority using the cumulative distribution as a link function (Figure 5).

Table 6. non-cardiac arrest high priority events logistic model.

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
AreaUrban80	-1.114	0.355	-3.139	0.002	0.421
AreaWork	-1.081	0.532	-2.031	0.042	0.124
dElderPopulation	0.022	0.007	3.070	0.002	21.460

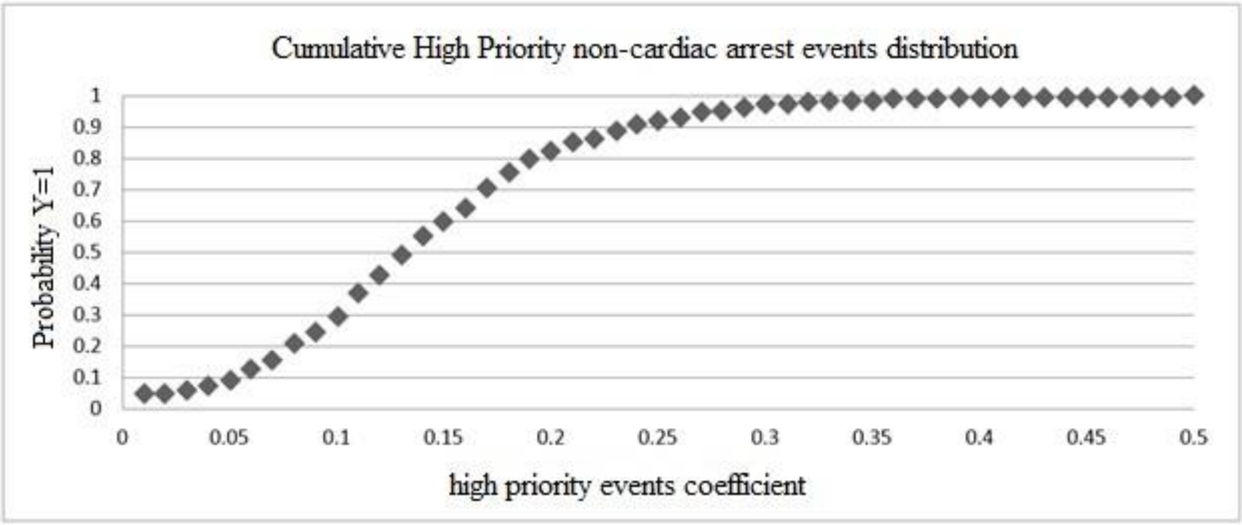


Figure 5. Cumulative High Priority non-cardiac arrest events distribution.

3.5. Optimization Scenarios and Results for the Location of Ambulance Stations

The optimization model has three different components, road crash emergencies constrained by a double coverage, cardiac arrest emergencies constrained by coverage of a maximum of eight minutes, and other priority emergencies constrained by coverage of a maximum of ten minutes. To test the sensibility of the road crashes and the impact of possible road safety funds in the uEMS, two different strategies are proposed. These two strategies define a twofold starting point: one, the network is planned with no special consideration for road crashes (Strategy 1) and two, the network is planned stressing road crash needs (Strategy 2).

Strategy 1 starts with a scenario (scenario 1.0) where road crashes are not differentiated from other events. This allows us to study uEMS planning without the intention of better response to road crash victims. To do that, we must eliminate the C^{road} parcel of equation (3.6); thus, equations (3.8) and (3.12) must also be eliminated from our model as well.

Strategy 2 starts with a scenario (scenario 2.0) that contains the full model; therefore, we explicitly assume that road crash assistance needs to be improved.

For both strategies, a first planning model is considered where the minimum number of stations required are located. To understand the impact of adding an additional station to an existing baseline station location, an upgrade of the uEMS is computed by adding one more station to the baseline station location. This upgrade can be done in two ways: improving the response to road crashes (scenarios 1.1 and 2.1) or improving the overall response including road crash needs (scenarios 1.2 and 2.2).

Thus, a total of 6 scenarios will be analyzed (Figure 6):

- Scenario 1.0 – Strategy 1 baseline station location. The minimum number of required stations are located based on cardiac arrest emergencies and other priority emergencies constraints, without focus on road crashes.
- Scenario 1.1 – station update. An additional station is proposed in scenario 1.0 with the intention only of improving the response to road crashes.
- Scenario 1.2 – station update. An additional station is proposed in scenario 1.0 taking into consideration the response to road crashes, cardiac arrests and other priority emergencies.
- Scenario 2.0 – Strategy 2 baseline station location. The minimum number of required stations are located based on cardiac arrests, road crashes and other priority emergencies constraints.
- Scenario 2.1 – station update. An additional station is proposed in scenario 2.0 with the intention only of improving the response to road crashes.
- Scenario 2.2 - station update. An additional station is proposed in scenario 2.0 taking into consideration the response to road crashes, cardiac arrests and other priority emergencies.

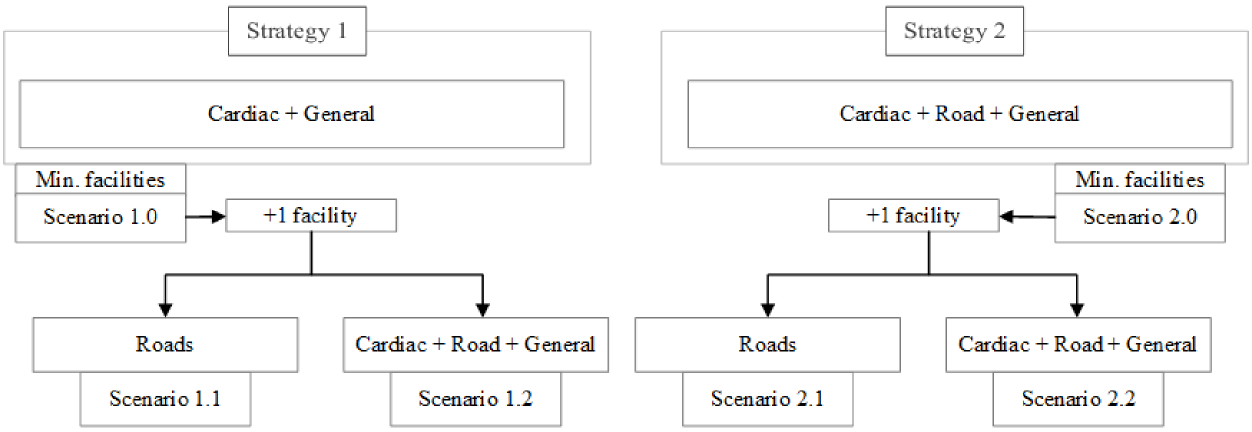


Figure 6. Strategies and Scenarios Scheme.

To run the optimization model, $DDMS_{type}$, we transform Porto into a nodal network. This network is composed of 87 representing the sub-census zones of the city. Each node is a possible station location as per Figure 4. The demand is allocated to each node using a radial distance clustering process. This means that each emergency call will be allocated to the closest node, where the distance metric is calculated by SQRT($\Delta x^2 + \Delta y^2$).

The time matrices, $R^r_{s,j}$, were calculated using Google Maps' API. The OD (Origin – Destination) travel times were computed for the morning peak hour. Each O/D pair was equal to 1 if the calculated travel time was inferior to the time limit and 0 otherwise. For the time limit r_1, we assume twelve minutes, which is the usual planning time in EMS; for r_3, we assume a time of eight minutes, which is the limit of the survival function for cardiac arrest (Erkut et al., 2008); and finally for r_2, we assume a time of ten minutes, which acts as a middle point between our higher and lower time limit values.

The optimization model was solved using the software package CPLEX with a dynamic search algorithm to solve the MIP problem. As a first analysis, the model was run with different maximum response times to assess how this indicator can influence the total number of stations required (Figure 7). Next, the scenarios solutions were computed, and the results were compiled in Figure 8 and Table 7, validating the applicability of the model.

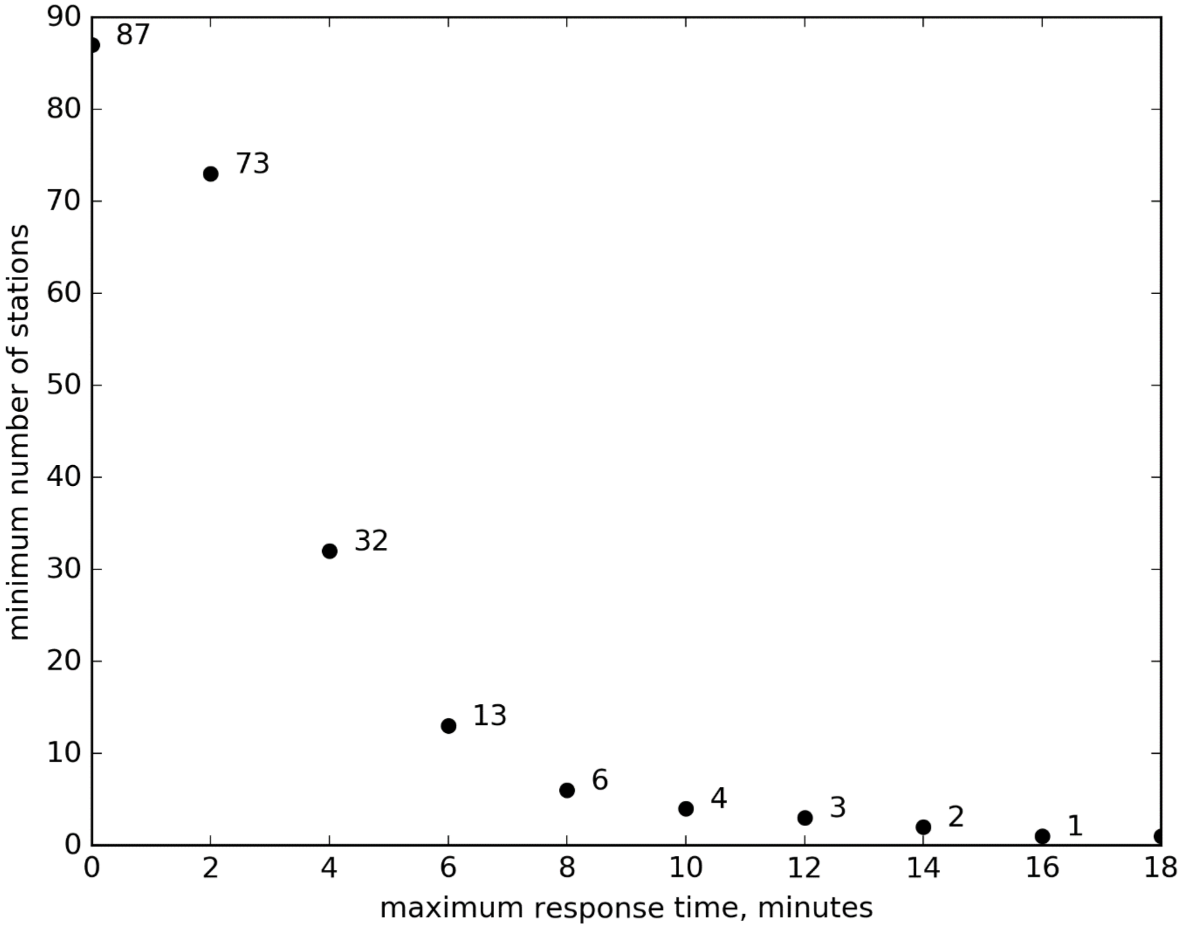


Figure 7. Minimum number of stations required for different maximum response times.

In Figure 7, it is evident that as we make the maximum response time stricter, the minimum number of required stations grows exponentially. For the settled limit of twelve minutes, a total of three stations are required, and to cover all demand points within the cardiac arrest response limit, the station requirements would double. It is also interesting to note that for more than sixteen minutes, the network requires only one station to cover the entire city.

Although the present example embodies a small urban area that requires a minimum of three stations for a maximum response time of twelve minutes, the model produces different solutions for the different scenarios, Figure 8 and Table 7; thus, the decision of locating an ambulance station is not obvious and depends highly on the addressed objectives.

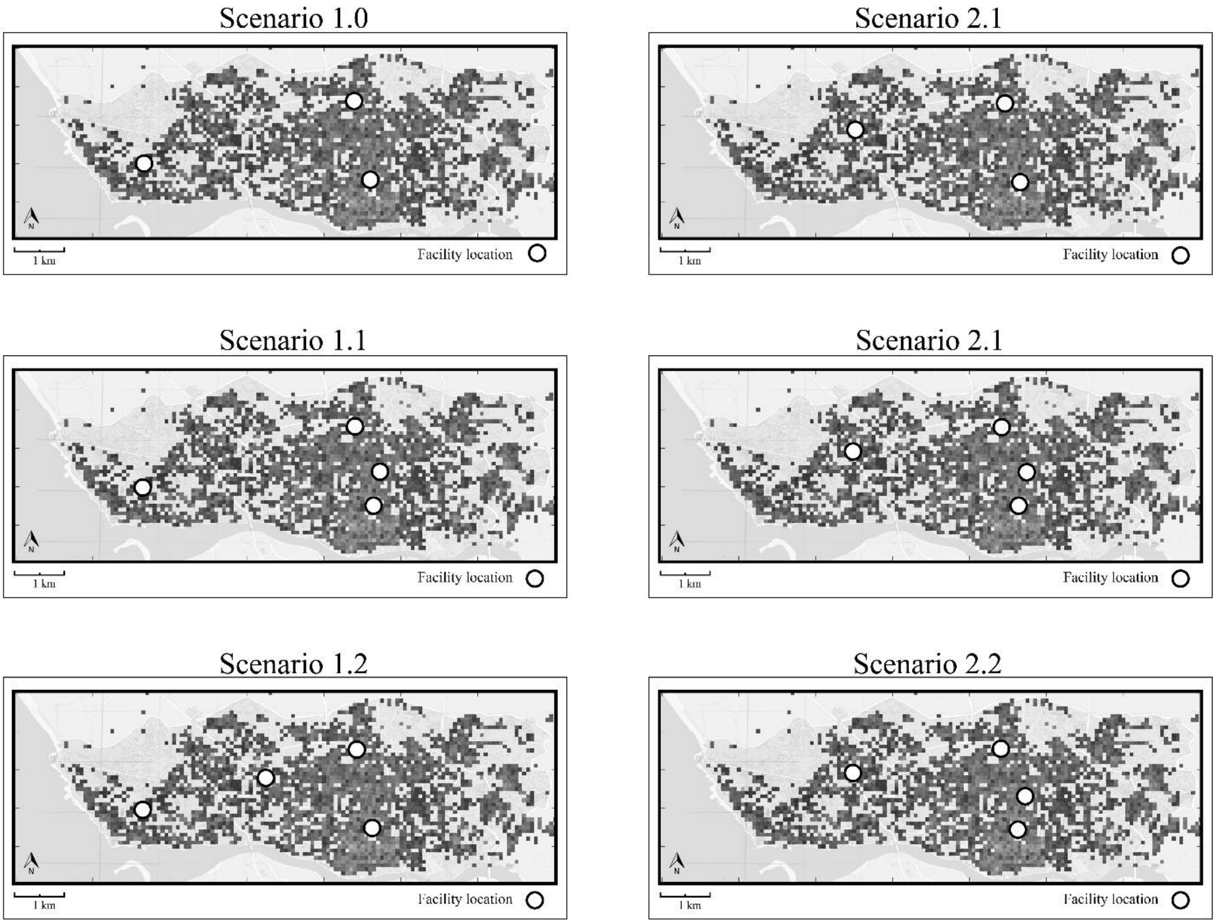


Figure 8. Optimization model solutions for each scenario.

Table 7. Optimization model results for each scenario and objective function results

	Total number of served calls within the three objectives					
	Scenario 1.0	Scenario 1.1	Scenario 1.2	Scenario 2.0	Scenario 2.1	Scenario 2.2
Station Locations name	20,45,62	20,45,62,76	20,45,62,26	45,62,81	45,62,81,76	45,62,81,76
Cardiac arrest	208	219*	229	195	206*	206
Road crash	600*	854	854	636	867	867
Others	6409	6409*	6409	6409	6409*	6409
Total	7217	7482	7492	7240	7482	7482

Note: *Collateral Gain – optimization is not trying to maximize this value

First, scenario 1.0 (strategy 1 baseline station location) produces a wider solution, whereas in scenario 2.0 (strategy 2 baseline station location), the stations are more concentrated in the central area. Specifically, for scenario 1.0, the left-most station is located close to the west border of the city, whereas in scenario 2.0, this station occupies a central place in the west side of the city. When upgrading each base strategy, different solutions are achieved depending on what is accounted for in the upgrade. For both strategies, if the upgrade is made accounting only for the double coverage of road crashes, the additional station will be allocated between the two east-most stations. This clearly shows that this is the area more prone to road crashes; thus, the model tries to double cover the road crashes there. It is important to mention that this area has four major fast roads. When the upgrade accounts for the three objectives in strategy 2 (scenario 2.2), there is no change in the solution, but when the same upgrade is made in strategy 1 (scenario 1.2) the additional station is placed in the geometric center of the base stations. Due to the location of one of the stations in the west border of the city, in scenario 1.2, the model tries to capture the missing overlap of the base stations so that it can capture more cardiac arrests and better double cover road crashes.

The maximum possible score of the objective function, corresponding to a station in each possible station location, is 7634, which corresponds to a maximum of 263 served cardiac arrests within eight minutes, 948 double covered road crashes and 6419 high priority emergencies covered in ten minutes. All scenarios cover the same number of high priority emergencies, 6409 cases, which represent 99.8% of all possible high priority emergencies. When the initial strategy addresses road crashes (scenario 2.0), in contrast to when it does not (scenario 1.0), a 6% increase in double road crash coverage is visible, whereas there is a 6% decrease in the eightieth minute coverage of cardiac arrests. When upgrading each base, strategies using only the road crash score, scenario 1.1, increase the road crash score by 42% and cardiac score by 5%, whereas scenario 2.1 increases the road score by 36% and cardiac score by 6%. Nevertheless, in absolute values, scenario 2.1 is ahead of scenario 1.1. If instead, the upgrade uses all three scores (scenario 1.2 and 2.2), only scenario 1.2 changes, covering 10% more cardiac arrests compared to the base scenario. Everything else remains the same.

3.6. DISCUSSION AND CONCLUSIONS

This paper proposes a methodology to plan the uEMS in the long term (annual planning), taking into account the specificities of road crashes in the funding of emergency medical service by road safety authorities. The methodology embodies two steps: a demand indicator assessment model and an ambulance station location model. For the demand indicator model, we developed a negative binomial model for different medical emergency types that was used as the input for the proposed station location model, DDMS_{type}.

The data from Porto adjusted well to the proposed demand indicator assessment models. The hypothesis we proposed revealed to be plausible because the various models as well as their predictors showed high significance. It is worth noting that highly dense urban areas have a higher number of events, but as the population density grows this number drops. Populations with no economic activity (elderly and population with no jobs) are the most correlated demographics in term of health emergency needs. As expected, rapid transit areas as well as areas with high volumes of traffic attract a higher number of road crash emergencies.

It is also worth noting that the more we try to detail the demand indicator, the less robust the model becomes (Likelihood Ratio Chi-Square). This might indicate a higher randomness in the explanatory parameters or a high degree of homogeneity of the demographics and land use indicators of the city of Porto.

The $DDMS_{type}$ model brings a new way of planning uEMS and allows a discrete prioritization of different types of events and the use of double coverage and different standards of coverage. In this sense, it is not required to have an individual uEMS system to respond to road crashes, and any possible road safety investment may also benefit other types of medical emergencies.

For this study, the initial planning strategy is demonstrated to not be a restriction on the future upgrade of uEMS to address road crashes, which indicates that deeper effort must be set to study micro planning strategies such as mid-term management with the dynamic allocation of ambulances or real-time uEMS management. Moreover, the optimization results do not show significant differences when planning with or without intention to improve the response to road crashes. This might indicate two possibilities: one, the degree of homogeneity of the demographics and land use of the tested city is too high and two, the layer of intervention when planning to improve the road crash response needs to go into micro-management, such as vehicle management, vehicle deployment and vehicle routing. Nevertheless, as a rule of thumb to better respond to road crashes, uEMS should have ambulance stations close to fast roads, double coverage when possible, and rapid response to areas with a large population with no economic activity and low population density.

The methodology underlying the hypotheses has been proven plausible through its application on a real city; therefore, it is prone to be applied worldwide. Although tested in only one city, we can claim that for better serve road crashes, ambulance stations should be allocated not only close to high traffic volume areas but also closer to fast roads both to ensure a quick response and because they are positively correlated with the probability of road crash emergencies.

For future research and for implementation by practitioners, case studies should be explored by integrating cities with different morphologies and different realities, and the collection of data for 3 years should be used to estimate each model parameters with greater detail.

Deeper investigations must occur in terms of the dynamic ambulance location and dispatching by assessing how different times of the day or different days of the week lead to different uEMS demand patterns. Finally, effort should be made to address ambulance dispatching and routing in the case of road crash situations, which alone lead to singular traffic conditions.

3.7. ACKNOWLEDGMENTS

This work was supported by FCT (Portuguese national funding agency for science, research and technology) under the grant PD/BD/52355/2013. We also show gratitude to INEM (Portuguese Institute of Medical Emergency) for the information and data that they provided us with.

3.8. REFERENCES

Aguero-Valverde, J. & Jovanis, P. P. 2006. Spatial analysis of fatal and injury crashes in pennsylvania. Accident Analysis & Prevention, 38, 618-625.

Aldrich, C. A., Hisserich, J. C. & Lave, L. B. 1971. An analysis of the demand for emergency ambulance service in an urban area. American journal of public health, 61, 1156-1169.

Bonneson, J. A. 2010. Highway safety manual. Washington, DC: American Association of State Highway and Transportation Officials.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. & Zhao, L. 2005. Statistical analysis of a telephone call center. Journal of the American Statistical Association, 100, 36-50.

Cadigan, R. T. & Bugarin, C. E. 1989. Predicting demand for emergency ambulance service. Annals of Emergency Medicine, 18, 618-621.

Channouf, N., L'ecuyer, P., Ingolfsson, A. & Avramidis, A. 2007. The application of forecasting techniques to modeling emergency medical system calls in calgary, alberta. Health Care Management Science, 10, 25-45.

Church, R. & Velle, C. R. 1974. The maximal covering location problem. Papers in Regional Science, 32, 101-118.

Couto, A. & Ferreira, S. 2011. A note on modeling road accident frequency: A flexible elasticity model. Accident Analysis & Prevention, 43, 2104-2111.

Culley, L. L., Henwood, D. K., Clark, J. J., Eisenberg, M. S. & Horton, C. 1994. Increasing the efficiency of emergency medical services by using criteria based dispatch. Annals of Emergency Medicine, 24, 867-872.

Daskin, M. S. 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. Transportation Science, 17, 48-70.

Daskin, M. S. & Stern, E. H. 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. Transportation Science, 15, 137-152.

De Maio, V. J., Stiell, I. G., Wells, G. A. & Spaite, D. W. 2003. Optimal defibrillation response intervals for maximum out-of-hospital cardiac arrest survival rates. Annals of Emergency Medicine, 42, 242-250.

Erkut, E., Ingolfsson, A. & Erdogan, G. 2008. Ambulance location for maximum survival. Naval Research Logistics, 55, 42-58.

Evans, J. D. 1996. Straightforward statistics for the behavioral sciences, Brooks/Cole.

Ferreira, S. 2010. A segurança rodoviária no processo de planeamento de redes de transporte em meio urbano. Universidade do Porto.

Ferreira, S. & Couto, A. 2013a. Hot-spot identification: A categorical binary model approach. Transportation Research Record: Journal of the Transportation Research Board, 2386, 1-6.

Ferreira, S. & Couto, A. 2013b. Traffic flow-accidents relationship for urban intersections on the basis of the translog function. Safety Science, 60, 115-122.

Geedipally, S. & Lord, D. 2010. Identifying hot spots by modeling single-vehicle and multivehicle crashes separately. Transportation Research Record: Journal of the Transportation Research Board, 2147, 97-104.

Gendreau, M., Laporte, G. & Semet, F. 1997. Solving an ambulance location model by tabu search. Location Science, 5, 75-88.

Gunes, E. & Szechtman, R. A simulation model of a helicopter ambulance service. Simulation Conference, 2005 Proceedings of the Winter, 4-4 Dec. 2005 2005. 7 pp.

Hauer, E. 1997. Observational before-after studies in road safety--estimating the effect of highway and traffic engineering measures on road safety.

Hauer, E. & Bamfo, J. Two tools for finding what function links the dependent variable to the explanatory variables. Proceedings of the ICTCT 1997 Conference, Lund, Sweden, 1997.

Henderson, S. G. 2005. Should we model dependence and nonstationarity, and if so how? Proceedings of the 37th conference on Winter simulation. Orlando, Florida: Winter Simulation Conference.

Hogan, K. & Revelle, C. 1986. Concepts and applications of backup coverage. Management Science, 32, 1434-1444.

Hogg, J. M. 1968. The siting of fire stations. *Journal of the Operational Research Society*, 19, 275-287.

Kepaptsoglou, K., Karlaftis, M. & Mintsis, G. 2012. Model for planning emergency response services in road safety. Journal of Urban Planning and Development ASCE, 138, 18-25.

Knight, V. A., Harper, P. R. & Smith, L. 2012. Ambulance allocation for maximal survival with heterogeneous outcome measures. Omega, 40, 918-926.

Kvålseth, T. O. & Deems, J. M. 1979. Statistical models of the demand for emergency medical services in an urban area. American Journal of Public Health, 69, 250-255.

Lan, B. & Persaud, B. 2011. Fully bayesian approach to investigate and evaluate ranking criteria for black spot identification. Transportation Research Record: Journal of the Transportation Research Board, 2237, 117-125.

Mayorga, M. E., Bandara, D. & Mclay, L. A. 2013. Districting and dispatching policies for emergency medical service systems to improve patient survival. IIE Transactions on Healthcare Systems Engineering, 3, 39-56.

Mcconnel, C. E. & Wilson, R. W. 1998. The demand for prehospital emergency services in an aging society. Social Science & Medicine, 46, 1027-1031.

McCormack, R. & Coates, G. 2015. 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, 294-309.

Meirich, S. 2008. Mapping guide for a european urban atlas. GSE Land Consortium.

Newgard, C. D., Schmicker, R. H., Hedges, J. R., Trickett, J. P., Davis, D. P., Bulger, E. M., Aufderheide, T. P., Minei, J. P., Hata, J. S., Gubler, K. D., Brown, T. B., Yelle, J.-D., Bardarson, B. & Nichol, G. 2010. Emergency medical services intervals and survival in trauma: Assessment of the “golden hour” in a north american prospective cohort. Annals of Emergency Medicine, 55, 235-246.e4.

Pons, P. T. & Markovchick, V. J. 2002. Eight minutes or less: Does the ambulance response time guideline impact trauma patient outcome?1. The Journal of Emergency Medicine, 23, 43-48.

Revelle, C. & Hogan, K. 1989. The maximum availability location problem. Transportation Science, 23, 192-200.

Sánchez-Mangas, R., García-Ferrrer, A., De Juan, A. & Arroyo, A. M. 2010. The probability of death in road traffic accidents. How important is a quick medical response? Accident Analysis & Prevention, 42, 1048-1056.

Savas, E. S. 1969. Simulation and cost-effectiveness analysis of new york's emergency ambulance service. Management Science, 15, B-608-B-627.

Setzler, H., Saydam, C. & Park, S. 2009. Ems call volume predictions: A comparative study. Computers & Operations Research, 36, 1843-1851.

Siler, K. F. 1975. Predicting demand for publicly dispatched ambulances in a metropolitan area. Health Services Research, 10, 254-263.

Toregas, C., Swain, R., Revelle, C. & Bergman, L. 1971. The location of emergency service facilities. Operations Research, 19, 1363-1373.

Valinsky, D. 1955. Symposium on applications of operations research to urban services—a determination of the optimum location of fire-fighting units in new york city. Journal of the Operations Research Society of America, 3, 494-512.

Who 2011. Global plan for the decade of action for road safety 2011-2020. World Health Organization.

Zhu, Z., Mcknew, M. A. & Lee, J. 1992. Effects of time-varied arrival rates: An investigation in emergency ambulance service systems. Proceedings of the 24th conference on Winter simulation. Arlington, Virginia, USA: ACM.

4. How do Traffic and Demand Daily Changes define Urban Emergency Medical Service (uEMS) Strategic Decisions? A multi-period survival approach[10]

Marco Amorim[11], Sara Ferreira[11], Antonio Couto[11]

Abstract

This paper presents a methodology to locate vehicle base stations using a scenario based optimisation to address daily traffic and demand changes, which are due to what we define as city dynamics. The model allows us to understand better how these daily changes affect an urban emergency medical service (uEMS) response system.

The methodology incorporates two steps. The first step uses scenario-based optimisation and survival function theory to locate vehicle base stations, whereas the second step uses agent-based simulation to assess the solution performance and compare it with average-period and non-survival prone solutions. The proposed models are tested for different situations using real data from the city of Porto.

The results of the sensitivity analyses show the importance of understanding the dynamics of cities and how they impact uEMS response systems. Useful insights regarding the number of stations and the average response time are addressed together with the minimum number of stations required for different maximum response time limits and different survival coefficients.

Finally, we conclude that a multi-period solution improves response time because it accounts for city dynamics and that a heterogeneous survival-based approach benefits victims' by properly measuring the system response concerning the victims' outcome.

Keywords: emergency medical service; scenario-based optimisation; simulation; city dynamics; survival functions; multi-period approach

---

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

Final version at: https://doi.org/10.1016/j.jth.2018.12.001

[11] CITTA, University of Porto – Faculty of Engineering, Porto, Portugal

4.1. INTRODUCTION

4.1.1. BACKGROUND, MOTIVATION, AND CONTRIBUTION

The study of road crashes, their implications and how to minimise their impact, is of high interest within the transport research community. This is particularly true for those that focus on road safety by studying the post-crash response of the emergency medical service (EMS).

Moreover, the World Health Organization presented in 2011 a global plan of action for road safety for the decade 2011-2020 (WHO, 2011). This plan indicates that researchers should focus on the post-crash response through activity 7 of pillar 5 of the document, where it reads: "Encourage research and development into improving post-crash response". One way of improving post-crash response is by studying the EMS response and assess where improvements can be made.

Some researchers have worked to create models for planning EMS solely to assist road crashes in a city (Kepaptsoglou et al., 2012) or in a specific road network (Zhu et al., 2012). However, emergency medical services usually respond to all types of medical emergencies, and no separate service may exist to assist just one type of medical emergency. One can argue that there are moral issues when resources are available and cannot be used in an active emergency because they are exclusive to another type of emergency. Moreover, Amorim et al. (2017) show that general planning of the emergency medical service generates a similar road crash response performance when compared to an EMS planned to prioritise road crashes. Therefore, from a transportation research perspective, it makes sense to study the emergency medical service as a whole.

In recent works, the focus of EMS response research has been on dynamic EMS, where vehicles are dynamically allocated, dispatched or routed to be prepared for the upcoming hours (Vasić et al., 2014, Zhang, 2012, Panahi and Delavar, 2009), and on the fact that emergency medical calls are heterogeneous - i.e. response time affects victims' survival differently (McCormack and Coates, 2015, Erkut et al., 2008, Blackwell and Kaufman, 2002).

This work aims to study the importance of city dynamics when planning an urban emergency medical service (uEMS) response system. An urban emergency medical service is defined as a service that responds to 'habitual emergencies' thus can be solved by a single organisation. Hence, disaster services and specific hazard emergencies are out of this work scope (Simpson and Hancock, 2009). Further, the work focuses on medical emergencies that take place in high-density urban areas; therefore, the service is subject to city dynamics. City dynamics is defined as an urban area where dynamism exists, and dynamism is described as a force that stimulates changes in short periods of time, such as hours or days (Silva et al., 2014). In sum, this work studies how daily traffic and population changes affect the uEMS strategic planning.

More specifically, we claim that the location of people and traffic, through the day, is not static in an urban environment (Lam et al., 2015, Vasić et al., 2014), and these two variables (people and traffic) are the most relevant ones when designing an urban EMS strategic plan – i.e. people in constant movement represent a possible dynamic demand (Krishnan et al., 2016, Wang et al., 2015), whereas traffic represents the network load because it constrains how quickly an emergency vehicle can reach a medical emergency (Erkut et al., 2009, Kim, 2016, Ingolfsson et al., 2008, Budge et al., 2010, Westgate et al., 2013) and, it correlates with road crashes and injuries (Ferreira and Couto, 2013, Amorim et al., 2017).

To assess how the urban behaviour interferes with uEMS, we propose a scenario-based optimisation model to locate uEMS vehicle stations according to victims' heterogeneity and city dynamics. Subsequently, we compare it with less robust solutions using numerical simulation and different performance metrics. In short, this framework analyses the performance of the uEMS response under different station configurations and contributes to the literature in the following ways:

- Formalizes a methodology to plan a strategic EMS response solution prepared for a dynamic environment;
- Uses the concept of urban dynamics and victims' survival, thereby implementing a scenario-based survival optimisation model;
- Uses a numerical application of the proposed methodology and models;
- Assesses the impact of city dynamics using several performance metrics calculated through simulation.
- Compare the proposed solution with static or non-survival models, showing the importance of these two concepts and their applicability.

4.1.2. EMS Response Models

The first emergency service location models date back to the year 1955 with the fire station location problem by Valinsky (1955) and Hogg (1968), and with the EMS station location problem by Savas (1969). Nevertheless, it was the work of Toregas et al. (1971) and Church and Velle (1974) that brought the emergency station location problem to the operation research community.

Toregas et al. (1971) present a solution that ensures all demand is covered by a maximum time or distance threshold which was named Location Set Covering Problem (LSCP). Church and Velle (1974) improve Toregas et al. (1971) work by using the concept of maximal coverage to implement the resources limitation neglected by Toregas et al. (1971). The addition of resources limitation resulted in the Maximal Coverage Location Problem (MCLP). These classic location problems were soon surpassed by stochastic models that try to deal with uncertainty in an attempt to come closer to the practitioners' needs. The most important ones are the Maximum Expected Covering Location Problem (MEXCLP) by Daskin and Stern

(1981), (1983) and the Maximum Availability Location Problem (MALP) by Hogan and ReVelle (1986), (1989). The authors implement facility reliability and busyness probability to solve the fact that once a facility is called for service the demand under its coverage is no longer covered.

More recently, Maxwell et al. (2009) classified research on dynamic allocation problems into three categories depending on the following: when real-time solution is required to make redeployment decisions (Brotcorne et al., 2003, Kolesar and Walker, 1974, Gendreau et al., 2001, Nair and Miller-Hooks, 2006); when solving the model involves computing optimal vehicle positions for every number of available vehicles via an integer programming formulation in an offline preparatory phase (Ingolfsson, 2006, Gendreau et al., 2005); or when one intends to incorporate system randomness into the model by using Markov decision processes (Berman, 1981c, Berman, 1981b, Berman, 1981a, Zhang et al., 2008, Alanis et al., 2013, Berman and Odoni, 1982, Jarvis, 1981) or make decisions under particular system configurations (Andersson and Varbrand, 2006, Andersson, 2005).

The bibliography shows that multi-period location models, where time is discrete, are a better practical solution for dynamic location problem than average-period models because in the latter time is continuous. This is proven by Miller et al. (2007) and supported by Boloori Arabani and Farahani (2012).

The concept of scenario-based approaches is also used when uncertainty is present. Serra and Marianov (1998) solved the p-median problem (PMP) under scenario-based demand uncertainty. When the number of facilities, or vehicles, is uncertain, Current et al. (1998) propose a scenario-based approach and solve the problem with a general-purpose mixed integer programming (MIP) solver. A detailed literature review focusing on the different EMS logistical problems can be read in the work of Reuter-Oppermann et al. (2017)

Moreover, with the advance of computer power and the availability of powerful personal computers, simulation models have become a useful tool for researchers wanting to formulate more realistic and complex problems, be it to assess solutions or to support optimised solutions (Restrepo et al., 2008, Maxwell et al., 2010, Yue et al., 2012, McCormack and Coates, 2015, Iannoni et al., 2009, Su and Shih, 2003).

Nevertheless, in urban Emergency Medical Services (uEMS), contrary to non-emergency facility location problems, underestimated or overestimated solutions have not only a monetary impact but carry a social impact. A wrong decision leads to higher response times to life-threatening medical emergencies, which impacts victims' survivability. To better understand the full range of the EMS system and how to plan it, the reader is pointed to the literature review made by Aringhieri et al. (2017) where the authors made a detailed analysis of the vehicle location, and relocation problem, and described dispatching and routing policies. The authors also study the interplay between the EMS system and other health services, forecast techniques and resource management.

When talking about victims' survivability, Sánchez-Mangas et al. (2010) studied the impact of medical response in road crashes and concluded that reducing the EMS response by 10 minutes may reduce road crash fatalities by 30%. Although this number can vary depending on many factors, it is obvious that a quicker medical response will result in improved medical assistance (Blackwell and Kaufman, 2002, Pons et al., 2005). In conformity, Erkut et al. (2008) note that survival probability and victims' heterogeneity models should prevail over coverage concepts when dealing with medical emergencies. These types of models have already been used in recent works (Knight et al., 2012, McCormack and Coates, 2015).

McCormack and Coates (2015) showed that without additional resources it is possible to increase cardiac arrest victims' survival; however, the proposed model only divides medical emergencies into two types: cardiac arrests and non-cardiac arrests. Another drawback in comparison to what we propose is the fact that the authors simplify the simulation by using approximated distances and average speeds when calculating travel times; adopting the same traffic conditions for inbound and outbound directions. This not only leads to synthetic travel conditions but also eliminates the possibility to account for the commuting impact in the network - i.e. higher travel times for inbound routes during the morning versus higher travel times for outbound routes during the afternoon. With cities becoming smart due to the introduction of intelligent systems and easy access to real-time information, such as Urban Traffic Control systems, access to traffic information can be a reality to uEMS and was proven to be beneficial on the tactical level (Amorim et al., 2018).

In a different approach, Kepaptsoglou et al. (2012) assume a uEMS model to solely respond to road crashes, disregarding other types of medical emergencies. Knight, Harper, and Smith (2012) address the heterogeneity of medical emergencies more directly. They propose a Maximal Expected Survival Location Model for Heterogeneous Patients using an exponential survival function for cardiac arrests, step functions for other types of medical emergencies, and a weight variable to prioritise emergencies.

Amorim et al. (2017) investigate uEMS station location for long-term planning periods and identify differences in the station configurations depending on how they assess victims' heterogeneity. However, by using an average-period approach, they are unable to detect the influence of city dynamics in the system response, and the solution might fail for specific periods of time according to the different traffic and demand characteristics. In the other hand, Dibene et al. (2017) implement robust scenario-based solutions for the classic Location Set Covering Model (LSCM), the Maximal Covering Location Problem (MCLP) and the Double Standard Model (DSM). They consider several factors such as if it is a work or off-day, the time of the day, geographical organisation and call priority, but not directly applying survival functions when measuring the system performance. They prove that the current solution in Tijuana, Mexico, could be improved regarding response time and demand coverage but could not show evidence regarding victims' survival.

Moreover, they only account for dynamics related to demand, thus not accounting for traffic changes. Krishnan et al. (2016) apply risk-based metrics to the vehicle location problem using Conditional-Value-at-Risk (CVaR). However, they tackle the problem under the view of system failure which assesses the number of calls not served; thus, they do not consider the victims' heterogeneity or victims' survival.

Recent work by Zaffar et al. (2016) compares performance metrics used in emergency vehicle location models with a focus on coverage, response time and survivability, and conclude that survivability models perform better in both survival and coverage metrics using a simulation-optimisation model. The authors show that demand varies in time and space along the day and week. Nevertheless, the proposed model disregards traffic changes and simplifies the travel time factor by assuming Manhattan distances and using an average speed for the emergency vehicles. Moreover, the study does not account for victims' heterogeneity, and victims' survivability is assessed by a linear function simplified from McLay and Mayorga (2010) which focuses on cardiac arrest.

There is a gap in the study and performance assessment of EMS strategic decisions such as vehicle or station locations. The literature review shows the progress made in performance metrics and robust solutions, but there is yet no significant scientific input in the use of multi-period survival-based solutions to assess the impact of dynamic urban factors such as traffic and demand. Our work tries to fill this gap by providing a data-driven scenario optimisation solution that accounts for demand and traffic fluctuations and presents a performance comparison between the multi-period approach and average-periods solutions by analysing, through simulation, different performance metrics and using real data.

This work will not focus on the problem of the number and allocation of vehicles to stations at the tactical level (planning for short periods). However, it is essential to provide the reader with literature that can sufficiently fill this gap. van Essen et al. (2013) study EMS planning at both strategic and tactical level, discussing the problem of sub-optimal solutions when tackling the two problems separated. They propose a combined solution for the two problems. Another critical tactical decision in EMS, particularly in urban environments, is the dispatching and possible reallocation of the vehicles. Schmid (2012) studies EMS at this level and formulates the allocation and relocation problem for the medical emergency. The formulation presented can also be adapted to determinate the adequate number of vehicles at each station.

4.2. METHODOLOGY

4.2.1. FRAMEWORK

To understand the impact of urban dynamics in the strategic planning of urban EMS the use of a multi-period approach is favoured. We propose an optimisation model with the use of scenarios, where each one translates the state of the city at a specific moment. This model will locate vehicle stations in order to maximise the victims’ survival according to the changing characteristics of the city, such as demand and traffic. An initial sensitivity analysis of the model parameters is made to assess the relation between the number of stations with the average response time and the maximum response time. A study to measure how the weights for cardiac arrests and road crashes affect the stations' location is conducted to analyse the heterogeneity parameters of the survival function.

Finally, a multi-period prone solution is proposed and compared in a simulated environment with average-period and non-survival solutions. These solutions are obtained by averaging the parameters of each scenario or by discarding the use of survival functions. The simulation model uses an agent-based approach where a road network and the uEMS demand represent the city dynamism according to data obtained from Porto city. Finally, we compare the robustness of the different solutions through a sensitive analysis by varying the emergencies location and the error in the travel time estimation.

4.2.2. Optimization Model

A system optimisation requires a performance measure. In a uEMS response network, the most common measures of performance found in the literature are coverage and system reliability. However, different types of emergencies have different requirements and priorities. Thus, the concept of maximum survival, first presented by Erkut et al. (2008), can measure the system's performance in the perspective of the victim.

The performance P_i of a uEMS response to an event i of type k can be defined by a survival function that depends on the time between the event start and the arrival of the assistance team, r_i, as per equation (4.1). The sum of all response performances is a straightforward performance metric for the emergency system. Nevertheless, other possible metrics are the average, minimum and mode of all the EMS responses. For simplicity, we chose the sum of all uEMS response performances to assess the system's overall performance.

$$P_i = f^k\left(r_i\right) \tag{4.1}$$

The introduction section discussed that a city behaves as a dynamic entity where traffic load and people location vary with time but repeats in cycles (e.g. daily, weekly, monthly). Accordingly, a multi-period approach with a scenario-based optimisation is preferred to yield a solution that performs and adapts as best as possible through the system's life. A scenario-based optimisation is typically used to deal with stochastic problems were uncertainty exists. This is usually the case when part of the model's inputs is unknown; thus the system designer predicts possible scenarios where a positive performance of the system is mandatory.

Here the goal is slightly different; real data is used to define the model inputs, and each scenario is a representation of a period from a defined cycle. The model's goal is to provide a solution that will perform as well as possible throughout the defined cycle using period instances as scenarios.

This method allows a station location solution for a cycle, C, of length $T = t_p - t_0$. C is an infinite set of instants t_i, with inputs $f(t_i)$, where $t_0 \leq t_i \leq t_p$.

However, for short periods a static behaviour is assumed. Thus, C has a finite number of periods ($\#S_i = S$) so that $C = [S_0, S_1, \ldots, S_s]$ is a cycle C with periods S_0, $S_1, \ldots,$ and S_s, where S_0 is the period between 0 and n, S_1 is the period between n and m, and S_s is the period between l and p with $l > m$, thus the model inputs become:

$f(t_0) = f(t_1) = f(t_2) = \ldots = f(t_n) \neq f(t_{n+1}) = f(t_{n+2}) = \ldots = f(t_m) \neq \ldots \neq f(t_{l+1}) = f(t_{l+2}) = \ldots = f(t_p)$, and $f(t_{p+a}) = f(t_0)$ with a as an infinitesimal.

The proposed model maximises, in a cycle, victims' survival by deciding where to locate vehicle stations, e, and allocating them to demand cluster nodes, p, in a nodal network for different scenarios s:

$$\text{maximize} \sum_k \sum_s \sum_l \sum_p y_{s,l,p} \times e_{s,p} \times f^k\left(r_{s,l,p}\right) \tag{4.2}$$

subject to

$$\sum_l y_{s,l,p} \times r_{s,l,p} \leq M_r \quad \forall\, p \text{ in } A, \forall\, s \text{ in } A \tag{4.3}$$

$$\sum_l y_{s,l,p} = 1 \quad \forall\, p \text{ in } A, \forall\, s \text{ in } A \tag{4.4}$$

$$y_{s,l,p} \leq x_l \quad \forall\, p \text{ in } A, \forall\, s \text{ in } A, \forall\, l \text{ in } L \tag{4.5}$$

$$\sum_l x_l \leq M_l \; . \tag{4.6}$$

$$x_l \in \{0,1\}, \ \forall l \in L \tag{4.7}$$

$$y_{s,l,p} \in \{0,1\}, \ \forall (s,l,p) \in A \tag{4.8}$$

where

S is the set of periods s[period 1, period 2, … , period s] and $S = C$,

L is the set of possible vehicle stations location [station 1, station 2, … , station l],

P is the set of demand cluster nodes p [node 1, node 2, … , node p],

A is the set of availability tuples such that *a [tuple (s,l,p)* | *if* $r_{slp} \leq M_r$ *]*,

$e_{s,p}$ is the number of events in demand cluster node p for period s,

x_l = 1 if a vehicle station is located at l and 0 otherwise,

y_{slp} = 1 if during period s node p is served by a station located at l, and 0 otherwise,

r_{slp} is the travel time required for a vehicle located at l to arrive at p during s,

M_r is the maximum allowed response time, and

M_l is the maximum number of stations.

Function (4.2) maximises the sum of the survival for each event type occurring at each scenario of the defined cycle. Inequality (4.3), equation (4.4) and inequality (4.5) control the model properties. Inequality (4.3) defines an upper bound for the response time. Decision variable y_{slp} and equation (4.4) are added to ensure that for every node in each scenario, only one station is allocated. This constraint ensures that the maximisation model will take into consideration the closest station when measuring survival. It is, however, important to note that this happens because the model does not take into consideration the number of available vehicles; thus the closest station is always available to respond. In the case the reader wants to address station or vehicles availability, the model can be adapted by using, for example, the concept of vehicle busy fraction (Daskin, 1983) or the expected response time model (Berg et al., 2016). Inequality (4.5) ensures that if node p is served by a station at l during s, then a station must be located at l.

Furthermore, there is the problem of deciding how many stations should be placed. This can be addressed if one can assess the worth of the performance gain by adding a new station. This relation is not yet defined; thus we opt to define inequality (4.6) to limit the number of stations (which can be assessed by the total available funding) and allow the model to run for different upper bounds.

The use of the availability set A allows a reduction in the number of decision variables, (4.8), as well as the number of constraints and the summation parts of the objective function, thus reducing the size of the problem.

Although the influence of the response time on cardiac arrest's outcome is very well defined in the literature, other types of medical emergencies do not have such survival function studies. To overcome this issue, when applying the model to a real case it is assumed that every type of survival function follows an exponential law, represented by a survival coefficient m^k and constant C^k, similar to the survival functions found on McCormack and Coates (2015), Erkut et al. (2008) and Knight et al. (2012). This transforms (4.2) into (4.9):

$$\text{maximize} \sum_k \sum_s \sum_l \sum_p y_{l,p} \times e_{s,p} \times [1 + \exp(C^k + m^k \times r_{s,l,p})]^{-1} \tag{4.9}$$

where K is the set of type of events k {cardiac arrest, car crash, …, others}

4.2.3. A SIMULATION MODEL FOR PERFORMANCE ASSESSMENT

To calculate the significance of city dynamics on uEMS systems and their impact on strategic station locations, a first scenario-based optimisation model test is made for the different maximum number of stations. This will give a first impression on how the number of stations might affect the maximum response time that can be theoretically offered. Further, it is assessed how different response time thresholds lead to a different number of minimum stations. Finally, the sensitivity of the survival coefficients used in the different survival functions is analysed to assess the dynamics of victims' heterogeneity, i.e. cardiac arrest versus road crashes.

After this first analysis, a case study is presented, and the different solutions for comparison are computed. This will result in different station configurations, where multi-period and survival approaches can be compared with classical approaches.

To assess the performance of these configurations and predict the impact of city dynamics several numerical simulations are run using real data, and different performance indicators are calculated, i.e. average response time, maximum response time, victims' survival.

The solutions performance is computed using an agent-based model, where an authority agent, the *city agent*, controls lower level agents: the event agents, *road network agent*, *vehicle agents*, and *node agents*. These agents coexist in an environment that simulates a spatial area defined by nodes, key locations, and a set of arcs connecting those nodes (Algorithm 1), and is similar to simulation models found in the literature (Haghani and Yang, 2007, McCormack and Coates, 2015, Su and Shih, 2003).

Algorithm 1 General simulation algorithm

<u>Definitions:</u>

T = *simulation period*

t = *timestamp*

j = *step*

j = *60 s*

t = *0*

<u>While *t* < *T*</u>

1. Update city

- Sets the environment conditions, s, from possible status $S = \{s_1, s_2, \ldots, s_n\}$, where $s = f(time)$
- Move events from events waiting list $E^w = \{e_1, e_2, \ldots, e_m\}$ to events active list E^a if the timestamp of event $e_m(t) < time$, and generate the assisting time required, e_n^{atime}

---

2. For all vehicles in the network:
- Vehicle time to destination, a_d^t, is updated → $a_d^t = a_d^t - j$
- If $a_d^t = 0$ → transfer vehicle to destination

3. For all active events $e_n^a \in E^a$:
- if no vehicle is allocated → run Vehicle dispatching algorithm – closest vehicle rule
- If the vehicle is at the occurrence location → Update assisting timer, $e_n^{atime} = e_n^{atime} - j$
- If $e_n^{atime} \leq 0$, assisting time ended → run Vehicle to hospital routing algorithm – closest hospital rule

4. Update nodes of type Hospital
- If the vehicle arrived → Transfer event to the hospital
- Ask network to return vehicle to its station → set new a_d^t

5. Update results dictionary, $R_{\{i\}\{j\}}$, with i = t and j = a^g
- For all vehicles in the network → if not in original station, a^g, $R_{\{i\}\{j\}} = R_{\{i\}\{j\}} + 1$, with $i = t$ *and* $j = a^g$

6. $\boldsymbol{t} = \boldsymbol{t} +$ ***step***
- If $t < T$ go back to 1

The *city agent* is responsible for dispatching a *vehicle agent* when required and listens to the *event agent* at the time of its activation, as an EMS entity. The *city agent* is also accountable for storing all other agents and requesting update orders from them.

The *event agent* is responsible for communicating with the *city agent* when it is activated and to keep the *city agent* informed of its current state by communicating through the allocated *vehicle agent*. When being assisted, the *event agent* is responsible for generating its assisting time, and when this time terminates, it will request the assisting vehicle to be transported to the closest hospital. The *vehicle agent* inquires the *network agent* for the closest hospital.

The closest dispatch rule (Haghani and Yang, 2007, Jagtenberg et al., 2017, Yang et al., 2005) and the closest hospital rule goes through a list of stations sorted by travel time and chooses the first one with an idle vehicle.

The *network agent* is responsible for routing all *vehicle agents* and for computing the fastest real-time OD route. It simulates traffic conditions and EMS vehicle movements by using nodes and arcs as an abstraction of the reality, and pre-computed travel times for different daily conditions using Google's Directions API.

The *vehicle agent* keeps track of its position in the *network agent* and informs the *city agent* when it arrives at each goal so that the *city agent* can record the performance indicators in a *data agent*. It travels to the node where the event occurs, assists the event, brings the event to the closest hospital and returns to its base. It is entirely dependent on orders placed by other agents.

The *node agent* has three types: node, hospital and station. This agent assists the *network and city agents* by storing vehicles and events.

Finally, a *data agent* keeps track of the individual performance of each EMS response for final calculations.

4.3. MODEL APPLICATION

4.3.1. DATA-DRIVEN TEST CASE

The proposed methodology was tested with real data from Porto, comprised of observations between May 2012 and May 2013. Porto city (approximately 238 000 inhabitants) is the centre of the Metropolitan Area of Porto (approximately 1 759 000 inhabitants). Daily, there are around 216 000 commuting trips to the city of which 63% of the trips have their origin outside the city (110 000 trips to workplaces and 35 000 trips to education places). This makes Porto an excellent experimental case because it captures city dynamics, especially those associated with commuting. Nevertheless, the small size of the city might limit our conclusions when compared with larger population or density European cities. However, we claim that the proportion of commuting trips from outside the city, and the city land use configuration are the factors that most influence the city dynamism.

According to the available EMS calls and travel times data, the daily uEMS response network operation was divided into three periods of equal length: The morning period (6:00 am to 2:00 pm), the afternoon period (2:00 pm to 10:00 pm), and the night period (10:00 pm to 6:00 am). These periods are eight hours long, which is the usual daily working time across many countries. Further, when analysing demand and traffic variation, Figure 9, it is possible to see how the travel time varies during the day and fits the chosen division. As an example,Figure 9.a shows how for a radial road link of the city, Southeast inbound and outbound, the travel times are higher for the inbound direction during the morning, but during the afternoon the higher travel times correspond to the outbound direction. Regarding demand, particularly the priority calls, there is a clear higher number of calls during the daytime than during the night. During the morning demand rate increases and then during the afternoon period it starts to decrease. The network operation also differentiates weekdays (Monday through Friday) from weekend days (Saturday and Sunday). Accordingly, a total of 5 periods are formed: Period 1 Weekday 6 am to 2 pm, Period 2 Weekday 2 pm to 10 pm, Period 3 Weekday 10 pm to 6 am, Period 4 Weekend 6 am to 10 pm and Period 5 Weekend 10 pm to 6 am. The weekend morning and afternoon periods were joined together due to their similarities concerning traffic conditions.

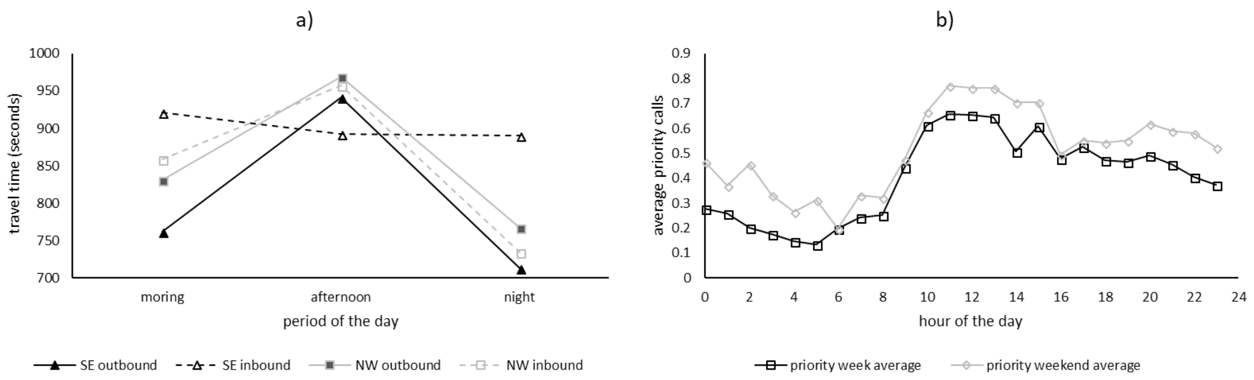


Figure 9. Dynamism in travel time and demand during the day

For the maximum response time, it is known that without any intervention, the survival rate of a cardiac arrest victims drops, linearly, to zero after 10 minutes (Eisenberg et al., 1990). Moreover, Valenzuela et al. (1997) indicate that the time interval needed for EMTs or paramedics to attach a defibrillator and clear the patient for defibrillation is estimated to be 2 minutes past EMT arrival or 1 minute past the time of initiation of CPR. This leads to a threshold of 8 minutes for the medical team to arrive at the event scene if we consider a 1-minute average for the dispatching time.

The Porto's EMS data was collected from the INEM (National Institute of Medical Emergency of Portugal) database and contains information on the type of emergency, timestamp and address of the occurrence. There are a total of 33 736 events in a one-year period. The addresses were geolocated using a python script that connects with the Google Maps API for geocoding.

The care-assisting time on the crash scene for each event is unknown; nevertheless, there are no negative times thus it is assumed a gamma distribution with a 10-minute average and standard deviation of 5 minutes according to Pons and Markovchick (2002).

For the optimisation model, the city network was converted into a nodal network where each node is the centroid of the city census subzones, for a total of 87 nodes. Each event was allocated to the closest node using a radial-distance based cluster algorithm.

The vehicle station location was assumed possible in any of the 87 nodes. Afterwards, a python script was created to use the Google Directions API and calculate the OD matrix of time travels for the different periods. This script asks Google Directions API for the fastest travel time, by car, between two coordinates for the morning peak hour (8 am), the afternoon peak hour (6 pm), the weekend peak hour (3 pm), and the uncongested travel time. The latter was allocated to the night periods.

All the data was processed and stored in an SQL Database using Python, SQLite3 and DB Browser for SQL. Later, the data was prepared to be used by the optimisation and the simulation models, the time travel matrix and the availability set were compiled into python raw files to reduce the number of calls to the SQL database and the data processing time when running the models. The two models were also programmed in Python. For the optimisation model, the Gurobi Optimizer python library, a state-of-the-art math programming solver, was used.

The optimisation model was run for this study case, followed by the simulation model. A sensitivity analysis was conducted by changing the relevant optimisation model parameters to understand their implications on the scope of this work.

4.3.2. MODEL PREPARATION AND COMPUTING RESOURCES

To support our claims, we propose a thorough sensitivity analysis regarding the spatial and temporal dynamics that may influence how the EMS system is planned.

With the optimisation model, we tested the impact of the maximum number of stations, M_l, the maximum response time, M_r, and the victims' heterogeneity with an emergency type weight, α^k. With the simulation model, we tested the impact of different uEMS network configurations from the optimisation model.

Each model's run was computed on a machine with an Intel quad-core processor at 1.73GHz and 8GB of memory RAM in a WIN10 64bits operative system. The models were implemented using Python v.2.7.8 and Gurobi v.6.5.2, both in 64 bits.

We assessed the computing time of the optimisation model for different problem sizes. The optimisation model was run for crescent integer values of M_l (from 1 till M_{inimum}) until the model returned a solution (thus identifying the minimum number of stations M_{inimum}), and then we recorded the total running time and the optimisation time for the identified M_{inimum}. The results are presented in Figure 10. It is evident that as the problem size grows, the total runtime grows exponentially due to the growth of the minimum number of required stations. However, when the model runs for one single M_l the exponential growth is much slower and reaches a maximum of 0.7 seconds for a problem size of 80 nodes.

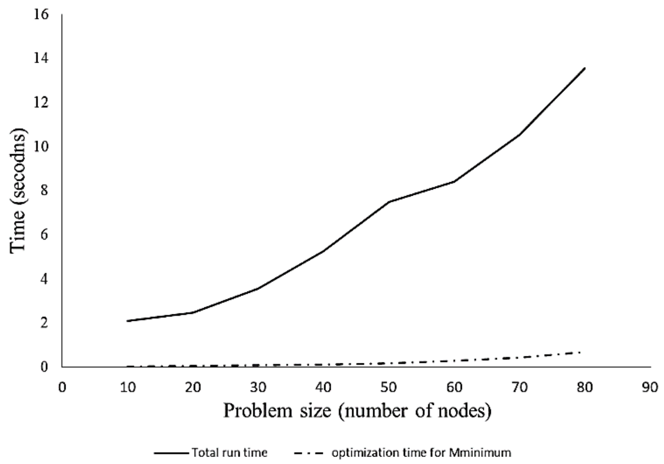


Figure 10. Optimization model performance

4.4. RESULTS AND DISCUSSION

4.4.1. STATION LOCATIONS OPTIMISATION AND ITS INFLUENCE

In this analysis, we test different values of M_l and M_r, and from the produced results we assess the impact of the number of stations on the average response time and the station network requirements for different thresholds of the maximum response time. Furthermore, we propose a base case that will serve as an overall solution for the presented optimisation problem. We also assess solutions regarding, and not regarding, the explained concepts of dynamism (scenario-based solution), and victims' heterogeneity (through survival functions). This will indicate the impact of city dynamics on a uEMS response system. We also test the sensibility of the victims' heterogeneity by weighting cardiac arrest and road crashes victims differently, thus assessing the importance of considering the heterogeneity of medical emergencies and its spatial impact on the optimisation solution.

4.4.1.1. INFLUENCE OF THE NUMBER OF STATIONS IN THE AVERAGE RESPONSE TIME

The optimisation model was run for different thresholds of uEMS vehicle stations, M_l. Figure 11 shows these outcomes, where the objective function result was converted into the average travel time.

As the number of stations increases, the average response time quickly drops in the first few additional stations and then slows down as the number of stations approaches the number of nodes. It is important to remember that events were clustered into nodes; thus a station implemented in a particular node will respond to the events of that node instantly. It is also important to understand that the response time is only the driving time; it does not account for the time the emergency call is being processed and the time for the paramedic team to prepare the victim for any necessary intervention.

Moreover, Figure 11 shows an apparent correlation between the average response time and the number of stations implemented. To the naked eye, there seems to be a hyperbolic or exponential relationship between both variables. Nevertheless, it is important to note that as the number of stations increases, theoretically, the average response time will never reach zero. Besides, when the average time grows, towards infinite, the number of stations required will lower but will never be null.

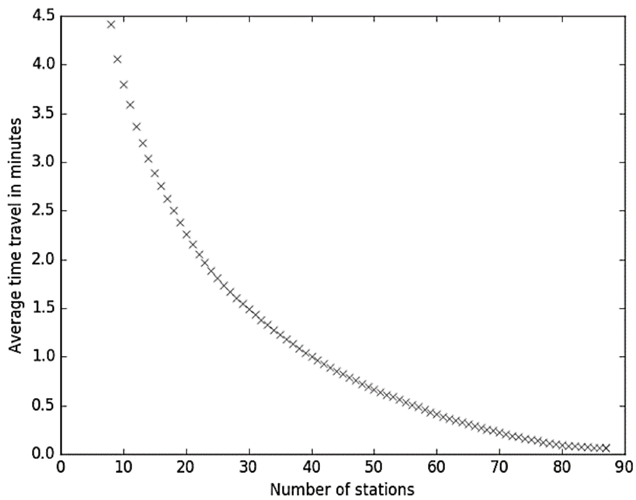


Figure 11. Average time travel for the different number of implemented stations

In a further analysis, we tested several types of fitting curves and several variable transformations to assess a possible law between the number of stations and the average travel time. Figure 12 groups the best-found relations.

The analysis leads to the identification of two different correlations. One occurs in the first seven observations (sample 1), Figure 12.a and Figure 12.c, and the other occurs in the remaining observations (sample 2), Figure 12.b and Figure 12.d. Undoubtedly, a power law explains the sample 1 correlation, Figure 12.a, whereas an exponential law better describes the correlation in sample 2, Figure 12.b, or, if we transform $x \rightarrow 1/(x + 10)$ by a linear law, Figure 12.d.

There is a disruption at the 7^{th} observation, corresponding to 7 stations implanted in the network. When adhering to the *x* transformation, the samples behave differently. The sample 1 average time drops more than 30% faster than sample 2 when *1/(n+10)* decreases (number of stations increase), pointing to differences in the network behaviour at the macro scale (few stations try to support the whole network) and microscale (many stations exist in the network which allows for each of them to focus on specific city areas) due to possible dynamic effects.

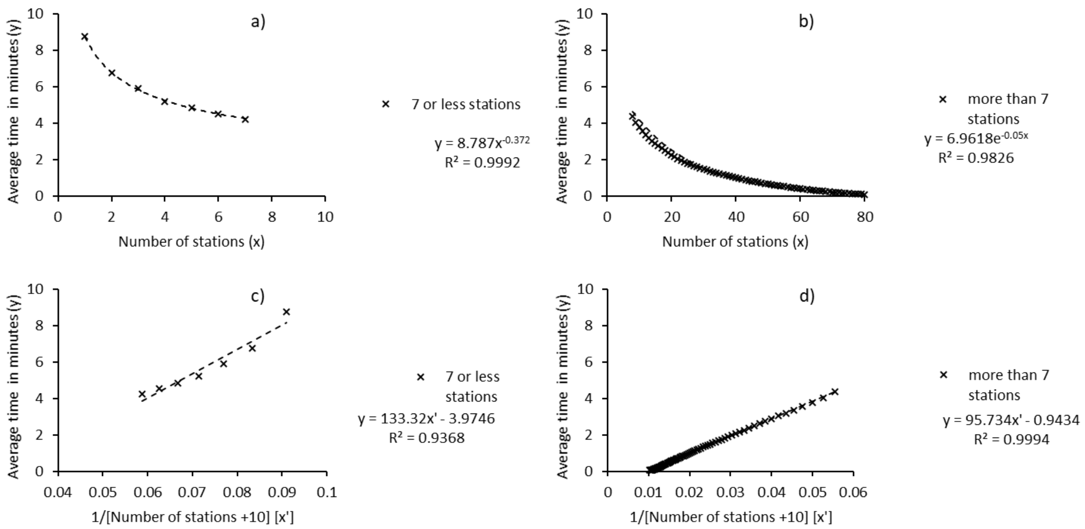


Figure 12. Correlation analysis between average time and number of stations

4.4.1.2. MINIMUM NUMBER OF STATIONS REQUIRED FOR THE MAXIMUM RESPONSE TIME

The maximum response time is one of the critical parameters in an EMS optimisation system. The response time defines the quality of an EMS system; nevertheless, a shorter response time requires more stations.

Figure 13 shows the decrease in required stations when the maximum response time is increased. For 5 minutes of the maximum response time, 24 stations are required, but as soon as this limit is extended by a half minute, the requirements drop to 18 stations. When increasing the time by one-third (from 5 minutes to approximately 8 minutes), the required stations drop to one-third (from 24 to approximately 8). After the 8th minute, the number of required stations drops at a lower rate. With an increase of 5 minutes (total of 13 minutes), the number of required stations drops from 8 to 3 stations. The maximum critical times are 6.5 minutes and 9 minutes. These seem to be the boundaries of a quick but costly response system (<6.5 minutes response time and >13 stations required), a standard response system (between 6.5 minutes and 9 minutes, and between 13 and 6 stations), and a slow but cheap response system (>9 minutes response time and <6 stations required).

These results show that a maximum response time of approximately 7 to 8 minutes can better equilibrate both the number of stations (10 to 8 stations) and the quality of the uEMS service. In fact, from 10 to 11 minutes, the number of required stations is the same as when the limit is set to 9.5 minutes

Nevertheless, this value is tightly connected to the road network configuration and land use.

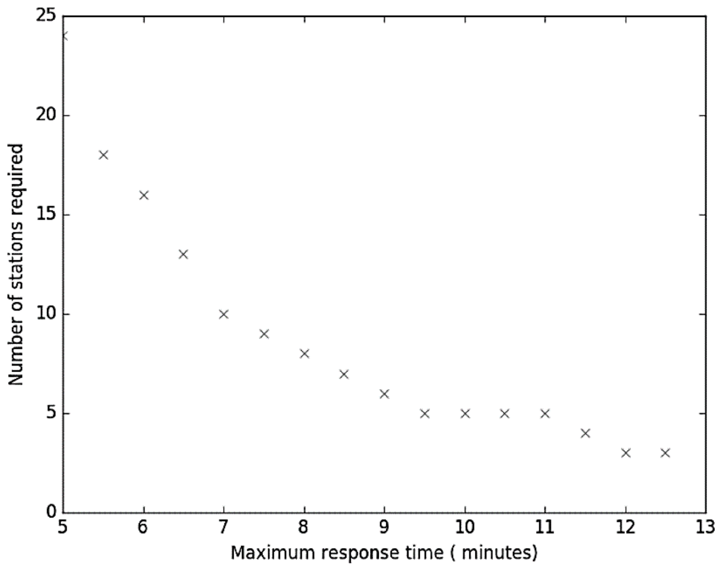


Figure 13. Number of minimum stations required for different maximum response times

4.4.1.3. Influence of emergencies heterogeneity in stations location: road crash vs cardiac arrest

To analyse the influence of victims' heterogeneity, we focus on cardiac arrests and road crashes. This choice is twofold. First, cardiac arrests are considered the event in which survival is most influenced by the response time. Second, road crash victims, although not always in a life-threatening situation, impose a road network impact. This means that while a road crash event is active, it is locally causing traffic jams that can quickly propagate through the city. These traffic jams or delays will influence the uEMS response to other events.

A batch of test cases was computed varying the weight, α^k, of each victims' emergency type by 2^n with $n = \{0, 1, 2, 3, 4, 5, 6\}$, function (4.10).

$$\sum_k \sum_s \sum_l \sum_p y_{s,l,p} \times e_{s,p} \times \alpha^k \times r_{s,l,p} \tag{4.10}$$

The idea underneath it is to understand how victims' heterogeneity behaves regarding spatial occupation. For each test case, a centroid is calculated by averaging the position of the optimal station location. The centroids for each tested case are presented in Figure 14.

The coordinate (0, 0) shows the centroid when both emergency types have the same weight. We increase the cardiac arrest weight and observe that the solution centroid moves towards the city centre -

from northwest to southeast. In the opposite situation (the road crash weight is increased relative to the cardiac arrest weight), we observe an opposite movement of the solution centroid - from southeast to northwest, and towards the outer bound of the city.

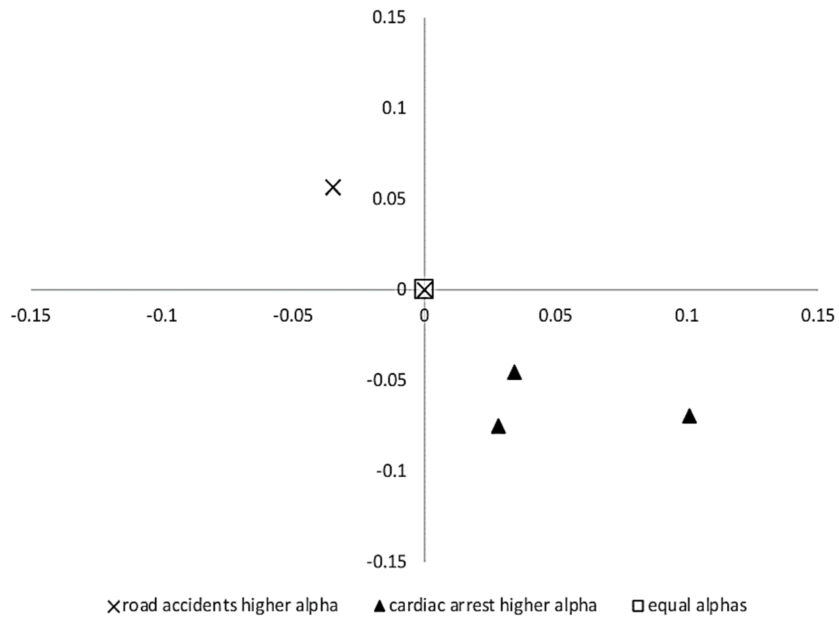


Figure 14. Offset of each centroid from the city centre in percentage (for different combinations of alpha).

4.4.1.4. Scenario-based and average-period approach solutions with and without survival implementation

We define our solutions for comparison by setting the number of stations to 7. For cardiac arrest victims, we assumed an exponential survival function with parameters m^k = 0.262 and C^k = 0.679 according to one of the survival functions presented by Erkut et al. (2008). For other life-threatening emergencies, no previous work computes explicitly such functions or parameters. Because of that, it was assumed a survival function with parameters m^k = 0.200 and C^k = 0. This translates to a softer survival decay where it is assumed a threshold of 12 minutes for a similar survival rate when compared with the cardiac arrest example, that is achieved at the 8-minute threshold (Pons and Markovchick, 2002).

Five solutions are computed with a maximum response time of 15 minutes:

- A "Robust Survival" solution, which uses the proposed scenario-based optimisation model that maximises survival,
- A "Robust" solution which uses the scenario-based optimisation but without maximising survivality;
- A "Static Survival" solution which uses average-period optimisation and maximises survival;

- A “Static” solution which uses average-period optimisation but without maximising survivality, and
- A "Response Coverage" solution which minimises the response time to each network node.

The station locations for each solution are presented in Figure 15. We notice that the robust solutions and the Response Coverage solution have the stations more dispersed across the urban area, while it seems that the static solutions concentrate stations in the areas with the higher number of yearly EMS calls. Noticeably, for this urban area, when we account for demand and traffic dynamism during the day and week, the location of the stations adapts from a position centred in the higher demand nodes slightly towards nodes with lower yearly demand. This means that the model is trying to compensate for the demand movements, weighting each period and node accordingly. It is clear evidence of the importance of considering city dynamics when planning for strategic decisions.

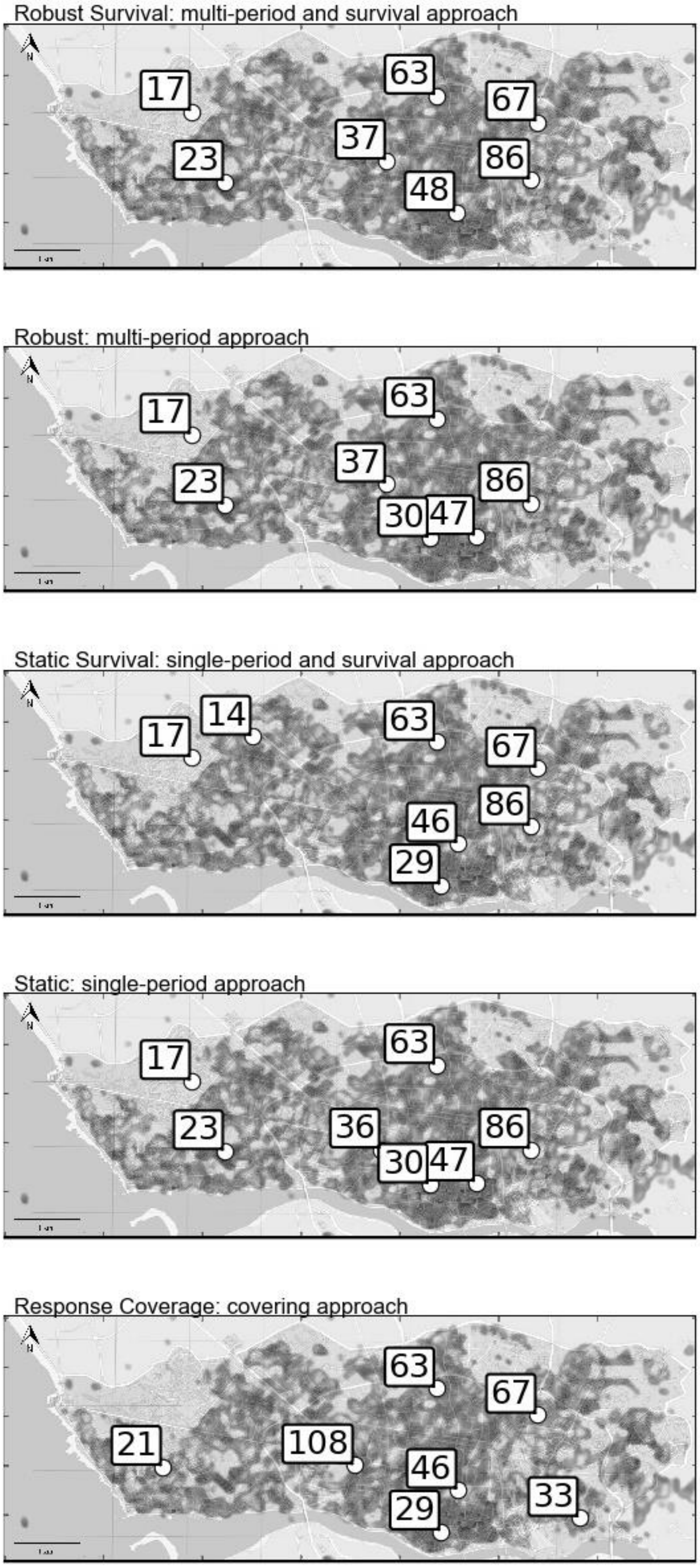


Figure 15. Station locations for the different solutions.

4.4.2. Solutions performance in a real environment through simulation

4.4.2.1. Background

To assess how the different proposed solutions perform in a realistic environment we use the simulation model presented in the methodology section. Contrary to the optimisation model that computes over multi-scenarios (or an average scenario in the Static solutions), simulation runs through a whole year, minute by minute and events occur according to a real-life case. The simulation model allows for a continuous evaluation of each solution using different indicators and tries to link theory and practice. As mentioned in the subsection 4.4.1.4, the solutions obtained are restricted to a 7-station network.

First, the solutions are tested with real data in a twofold hypothesis:

- resources are unrestricted, i.e. the closest station has always an idle vehicle, and
- resources are restricted, i.e. each station has two vehicles at the start of the simulation in a total of 14 vehicles.

Secondly, the solutions are tested in hypothetical scenarios where demand location or traffic conditions can vary from the observed ones. The variation of the demand location follows the tendency observed in the real data, i.e. the probability of an event occur in location i at period s, and is controlled by probability P(change location). An error term ε' controls the traffic variations and is drawn from a normal distribution with mean zero and standard deviation σ. The error adjusts the observed travel time t to $t' = t + t \times \varepsilon'$, this is, during strategic decisions the observed travel times can have a maximum error of $\sigma/2$ with a confidence of 95%. Or in other words, the real travel times can vary up to $\sigma/2$.

4.4.2.2. Performance on the real data

Table 8 and Table 9 show a summary of the real data simulation results. The classic metrics average response time and percentage of emergencies covered by time thresholds of 8, 12, 15 and 20 minutes, are used for the first assessment.

The results show that the solutions that account for city dynamics have better performance due to a repositioning of the stations to accommodate demand and traffic changes. Static solutions focus in quickly responding to nodes with higher demand result of an average-period approach. When resources are scarce, the service performance degrades mostly due to demand peaks. A local shortage of resources leads to a drop in the system performance - the closest station is unavailable; thus, a vehicle further away needs to be allocated to the emergency. While a global shortage of resources leads to response delays, i.e. there are not enough vehicles to serve all the active emergencies. These delays seem to be more frequent when the solutions are computed using survival functions. Nevertheless, the use of survival functions does not significantly affect the average response time; however, there is a noticeable impact in the time threshold metric.

When focusing on the victims' survival, the analysis is done only for life-threatening emergencies. In the introduction section, we emphasised that accounting for victims' heterogeneity is essential. The exponential survival functions parameters m^k and C^k were set equal to those used during the optimisation. Non-life-threatening emergencies have a survival rate of 1, therefore can be discarded from the analysis. For this analysis, we add an extra performance metric based on a survival threshold. This metric is calculated using a binary evaluation which equals to 1 if the EMS response takes less than 8 minutes to cardiac arrests and 12 minutes to other life-threatening emergencies. Otherwise the response is scored with a zero.

The unrestricted resources results were plotted in Figure 16 and Figure 17, and the limited resources results in Figure 18 and Figure 19.

Figure 16 and Figure 18 show the survival gain of each solution when compared to the most straightforward solution - the Response Coverage. Figure 17 and Figure 19 show how many life-threatening emergencies took longer than the defined thresholds to be responded.

The results are clear about which solution provides a better service to the EMS victims. As expected, when we account for victims' heterogeneity and survival and use a robust solution that considers city dynamics we can provide a solution that is much more adequate. For the tested cases, The Robust Survival solution reached almost the double gain than the other solutions (Robust, Static Survival and Static solutions) during unrestricted resources. When resources are restricted, the robust solutions gain is not so substantial, but the use of survival functions outperforms their counterparts. It is evident that using survival as a metric will benefit the solutions that had survival into account. Nevertheless, although these solutions slightly underperform when assessed by the classic metrics in an unrestricted resources scenario, when resources are scarce the difference shown by the classic performance metrics is negligible. When assessed by survival thresholds, Figure 17 and Figure 19, the survival based solutions outperform the other solutions in both resource unrestricted and limited resources hypothesis.

It is interesting to notice how the Robust and Static Survival solutions perform similarly when resources are unrestricted, but the performance of the robust solution relatively drops when resources are limited.

One final remark worth mentioning is the fact that during the analysed year there seems to exist two timeframes where all solutions' performance degrades. First, at the start of the simulation, in May 2012, and second a more steep performance fall happens around the 25 000th events, during February 2013. Although the Robust Survival solution is relatively less affected, it is worth note that there might exist dynamism that influences the uEMS response in monthly basis.

Table 8. Simulation results summary for the unrestricted resources test of each solution.

unrestricted resources		time in minutes Robust Survival	Robust	Static Survival	Static	Response Coverage
Response time	Average	4.24	4.22	4.41	4.23	4.40
	Max	12.50	13.68	12.50	13.68	12.50
	Std	2.20	2.42	2.51	2.42	2.05
	>8 min	0.024	0.042	0.068	0.039	0.035
	>12 min	0.001	0.001	0.001	0.001	0.001
	>15 min	0.000	0.000	0.000	0.000	0.000
	>20 min	0.000	0.000	0.000	0.000	0.000
delay	total	0	0	0	0	0
	Average	0.00	0.00	0.00	0.00	0.00
	Max	0.00	0.00	0.00	0.00	0.00

Table 9. Simulation results summary for the two vehicles per station test of each solution.

Two vehicles per station		time in minutes Robust Survival	Robust	Static Survival	Static	Response Coverage
Response time	Average	4.77	4.77	4.88	4.83	4.90
	Max	24.07	23.53	25.67	23.87	24.02
	Std	2.66	2.84	2.80	2.89	2.52
	>8 min	0.088	0.098	0.110	0.104	0.087
	>12 min	0.010	0.013	0.011	0.015	0.014
	>15 min	0.003	0.004	0.003	0.005	0.002
	>20 min	0.000	0.001	0.000	0.000	0.000
delay	total	139	130	140	130	119
	Average	28.32	18.51	27.31	23.52	23.32
	Max	66.00	43.00	71.00	52.00	51.00

Figure 16. Accumulated survival gain of each solution compared with the Response Coverage solution when resources are unlimited, and a total of 7 stations exist.

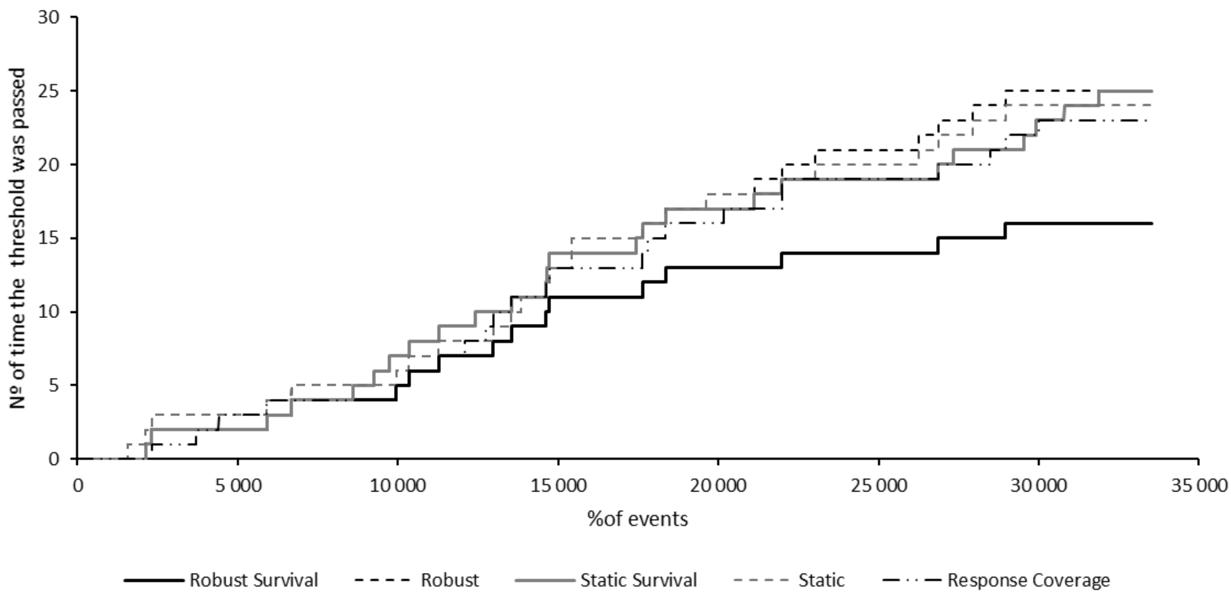


Figure 17. The number of times the EMS response crossed a fixed threshold when resources are unlimited, and a total of 7 stations exist. The threshold is 8 minutes for cardiac arrest emergencies and 12 minutes for other life-threatening emergencies.

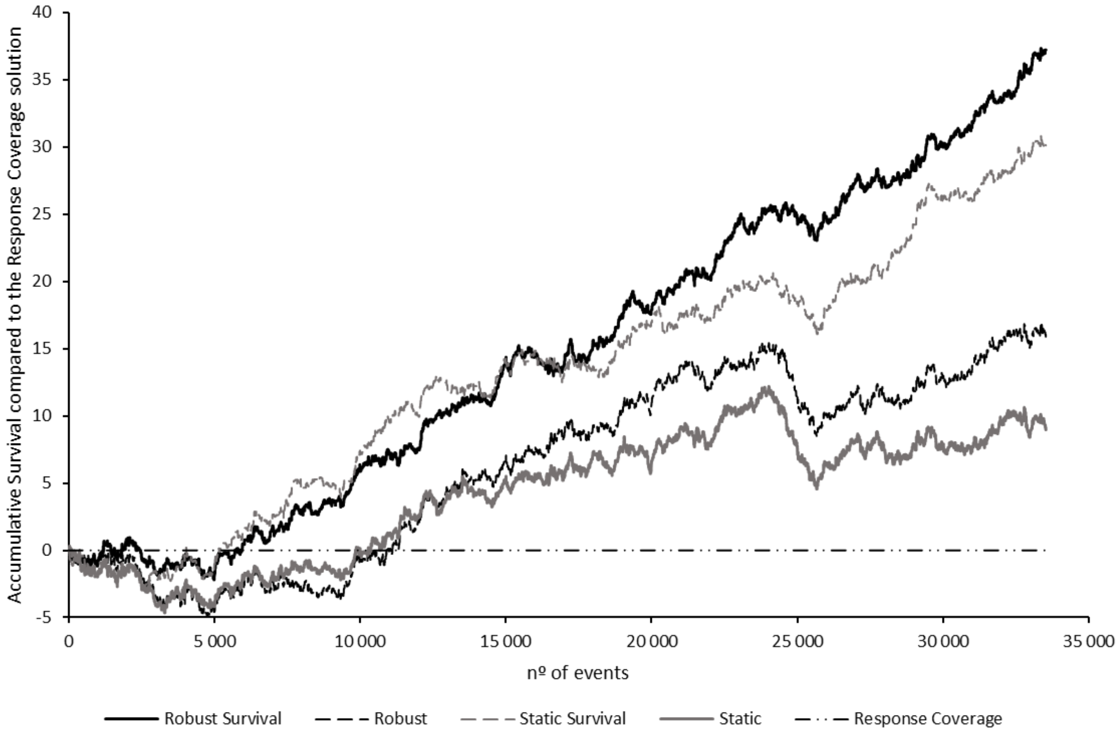


Figure 18. Accumulated survival gain of each solution compared with the Response Coverage solution when resources are limited to 2 vehicles per station and a total of 7 stations exist.

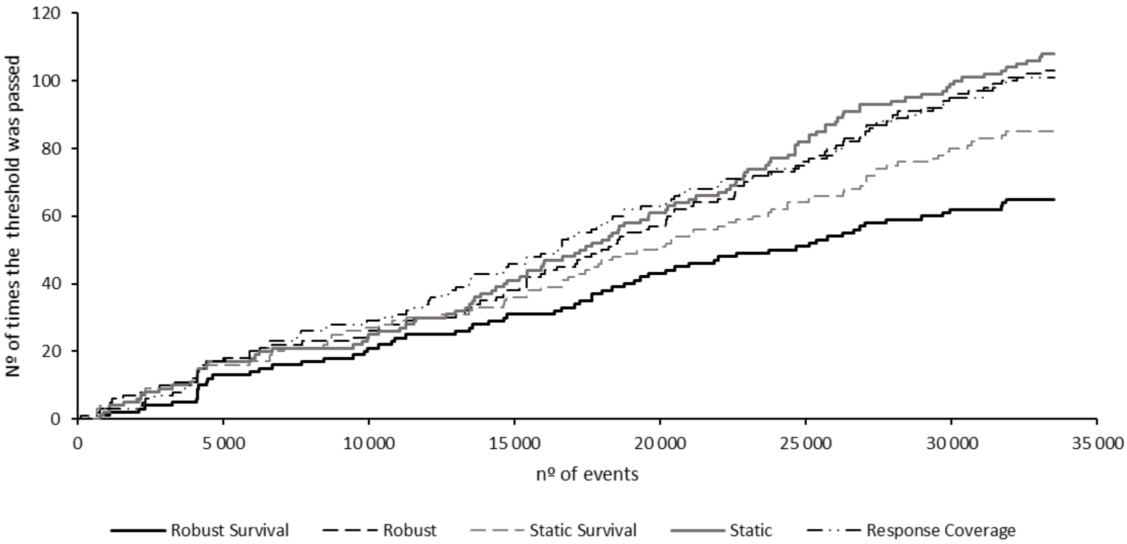


Figure 19. The number of times the EMS response crossed a fixed threshold when resources are limited to 2 vehicles per station and a total of 7 stations exist. The threshold is 8 minutes for cardiac arrest emergencies and 12 minutes for other life-threatening emergencies.

4.4.2.3. PERFORMANCE ON SYNTHETIC DATA

We changed the real data to test how the models will adapt to the uncertain future. These changes were applied to the location of the emergencies and to the travel times as discussed in section 4.4.2.1.

First, the simulation was run for each solution using a parameter *Pd* that controls the probability of an emergency occurring in a location different from the observed one. Figure 20 shows the average survival of each solution for different *Pd* values - ranging from a 0.0 to a 1.0 probability. The results show that the robust solutions have the higher performance when *Pd* ranges up to 0.75. This observation supports the applicability of our scenario-based model and the importance of considering city dynamism when making strategic decisions.

Second, the simulation was run for each solution using a maximum error term ε' for the estimated travel time, ranging from 0.0 to 1.0. Let us remember that an error of ε' means that there is a confidence of 95% that the estimated travel time $r\varepsilon \in [r - r\varepsilon', r + r\varepsilon']$, where r is the real travel time. The obtained results showed that during strategic decisions the use of estimated travel times (with an error up to ε') does not produce differences in the relative performance of the solutions. However, it is worth to note that each solution performance drops once the estimated travel times experienced an error ε' of 0.5, and the drop accentuates when the error crosses the 0.75 value.

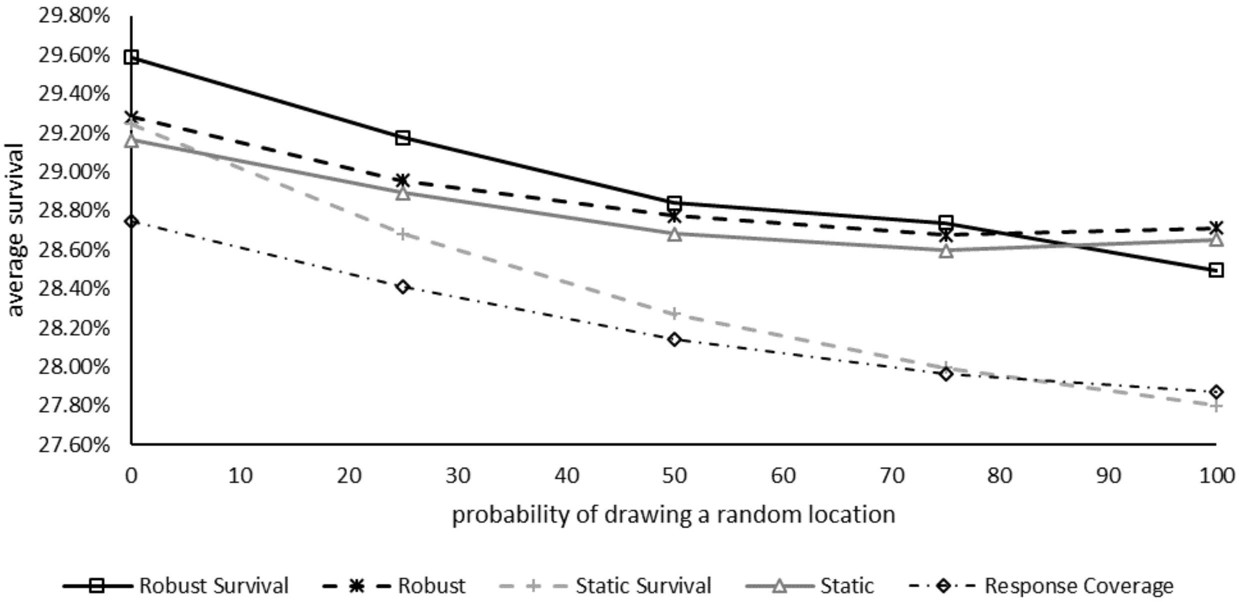


Figure 20 Solutions performance in synthetic data with different probabilities of changing the emergencies location

4.5. Conclusions

This work opens doors to the study of city dynamics and its influences in the strategic planning of an uEMS response system.

We defined a performance metric for the uEMS response by summing the survival score of each rescued victim. Afterwards, we proposed a scenario-based optimisation model where the scenarios capture day periods to infer city dynamics.

An agent-based model simulation is offered to assess uEMS performance and stresses the importance of a dynamic system.

The optimisation model was validated and, after minor simplifications, performed quickly, allowing for several cases to be tested within a reasonable time. The validation and sensitivity analyses were performed using real data from Porto city, collected during one year in a total of 33 736 events occurring between 10th May 2012 and 9th May 2013. This analysis confirm the importance of multi-period approaches to correctly inform the decision makers on how to better locate uEMS station.

The division of the timeline in periods is a simple and efficient way to deal with city dynamics and is proven to be relevant in the positioning of uEMS stations. The city is dynamic with people and traffic concentrating in distinct parts of its network throughout the day as proven by the dynamic versus non-dynamic solutions analysis. Moreover, road crashes and cardiac arrests were proven to have different time and space behaviours, supporting our assumptions and showing the relevance of victims' heterogeneity.

Regarding the availability of vehicles, it was shown that proper management of resources is fundamental to avoid response delays which can quickly propagate to future calls. Using the proposed concepts of city dynamism is an essential point for future investigation into tactical uEMS decisions such as vehicle allocation.

The results revealed that the use of survival functions and their parameters impact the location of stations, which stresses the need of further research in the survival topic, particularly for road crashes and other types of meaningful (survival or system related) emergency events.

Different period sizes should be tested as exposed in the results discussion where we show that different months might have different dynamic behaviours. Moreover, it is important to note that the simulation model should be relaxed to allow vehicles to be reallocated to different stations and to allow vehicles that are returning from a hospital to be allocated to an active event without the need to return to their base first and be allocated after.

While this work focuses on strategic decisions, it is essential to progress with a similar investigation for tactical decisions and possibly an integrated solution that locates stations and allocates vehicles.

An important step that also needs further discussion is the implementation of Urban Traffic Control (UTC) in the EMS planning. These techniques can benefit both strategic and tactical planning of the service, first by allowing more accurate traffic information in the design of the existing scenarios, and second by allowing tactical decision to be made upon real-time information. Both these points can further improve the proposed methodology by providing a more accurate and broader database of travel times to be used in the scenarios for the optimisation, and also to provide the simulation model with higher precision.

Finally, we suggest studying the disruption that a road crash causes in the road network and how it might interfere with the EMS response.

4.6. ACKNOWLEDGEMENTS

We acknowledge the support of FCT (Portuguese national funding agency for science, research and technology) under the grant PD/BD/52355/2013 during the development of this work. We also express gratitude to INEM (Portuguese Institute of Medical Emergency) for providing us with relevant information and data.

4.7. REFERENCES

Alanis, R., Ingolfsson, A. & Kolfal, B. 2013. A markov chain model for an ems system with repositioning. Production and Operations Management, 22, 216-231.

Amorim, M., Ferreira, S. & Couto, A. 2017. Road safety and the urban emergency medical service (uems): Strategy station location. Journal of Transport & Health, 6, 60-72.

Amorim, M., Ferreira, S. & Couto, A. 2018. Emergency medical service response: Analyzing vehicle dispatching rules. Transportation Research Record, 0, 0361198118781645.

Andersson, T. 2005. Decision support tools for dynamic fleet management. Doktorsavhandling, Linköpings Universitet, Sverige.

Andersson, T. & Varbrand, P. 2006. Decision support tools for ambulance dispatch and relocation. J Oper Res Soc, 58, 195-201.

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

Berg, P. L. V. D., Essen, J. T. V. & Harderwijk, E. J. Comparison of static ambulance location models. 2016 3rd International Conference on Logistics Operations Management (GOL), 23-25 May 2016 2016. 1-10.

Berman, O. 1981a. Dynamic repositioning of indistinguishable service units on transportation networks. Transportation Science, 15, 115-136.

Berman, O. 1981b. Repositioning of distinguishable urban service units on networks. Computers & Operations Research, 8, 105-118.

Berman, O. 1981c. Repositioning of two distinguishable service vehicles on networks., IEEE Transactions on Systems, Man and Cybernetics, 11, 187-193.

Berman, O. & Odoni, A. R. 1982. Locating mobile servers on a network with markovian properties. Networks, 12, 73-86.

Blackwell, T. H. & Kaufman, J. S. 2002. Response time effectiveness:Comparison of response time and survival in an urban emergency medical services system. Academic Emergency Medicine, 9, 288-295.

Boloori Arabani, A. & Farahani, R. Z. 2012. Facility location dynamics: An overview of classifications and applications. Computers & Industrial Engineering, 62, 408-420.

Brotcorne, L., Laporte, G. & Semet, F. 2003. Ambulance location and relocation models. European Journal of Operational Research, 147, 451-463.

Budge, S., Ingolfsson, A. & Zerom, D. 2010. Empirical analysis of ambulance travel times: The case of calgary emergency medical services. Management Science, 56, 716-723.

Church, R. & Velle, C. R. 1974. The maximal covering location problem. Papers in Regional Science, 32, 101-118.

Current, J., Ratick, S. & Revelle, C. 1998. Dynamic facility location when the total number of facilities is uncertain: A decision analysis approach. European Journal of Operational Research, 110, 597-609.

Daskin, M. S. 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. Transportation Science, 17, 48-70.

Daskin, M. S. & Stern, E. H. 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. Transportation Science, 15, 137-152.

Dibene, J. C., Maldonado, Y., Vera, C., Trujillo, L., De Oliveira, M. & Schütze, O. 2017. The ambulance location problem in tijuana, mexico. In: Schütze, O., Trujillo, L., Legrand, P. & Maldonado, Y. (eds.) Neo 2015: Results of the numerical and evolutionary optimization workshop neo 2015 held at september 23-25 2015 in tijuana, mexico. Cham: Springer International Publishing.

Eisenberg, M. S., Horwood, B. T., Cummins, R. O., Reynolds-Haertle, R. & Hearne, T. R. 1990. Cardiac arrest and resuscitation: A tale of 29 cities. Annals of Emergency Medicine, 19, 179-186.

Erkut, E., Ingolfsson, A. & Erdogan, G. 2008. Ambulance location for maximum survival. Naval Research Logistics, 55, 42-58.

Erkut, E., Ingolfsson, A., Sim, T. & Erdoğan, G. 2009. Computational comparison of five maximal covering models for locating ambulances. Geographical Analysis, 41, 43-65.

Ferreira, S. & Couto, A. 2013. Hot-spot identification: A categorical binary model approach. Transportation Research Record: Journal of the Transportation Research Board, 2386, 1-6.

Gendreau, M., Laporte, G. & Semet, F. 2001. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. Parallel Computing, 27, 1641-1653.

Gendreau, M., Laporte, G. & Semet, F. 2005. The maximal expected coverage relocation problem for emergency vehicles. J Oper Res Soc, 57, 22-28.

Haghani, A. & Yang, S. 2007. Real-time emergency response fleet deployment: Concepts, systems, simulation & case studies. In: Zeimpekis, V., Tarantilis, C. D., Giaglis, G. M. & Minis, I. (eds.) Dynamic fleet management: Concepts, systems, algorithms & case studies. Boston, MA: Springer US.

Hogan, K. & Revelle, C. 1986. Concepts and applications of backup coverage. Management Science, 32, 1434-1444.

Hogg, J. M. 1968. The siting of fire stations. J Oper Res Soc, 19, 275-287.

Iannoni, A. P., Morabito, R. & Saydam, C. 2009. An optimization approach for ambulance location and the districting of the response segments on highways. European Journal of Operational Research, 195, 528-542.

Ingolfsson, A. The impact of ambulance system status management. Presentation at 2006 INFORMS Conference, 2006.

Ingolfsson, A., Budge, S. & Erkut, E. 2008. Optimal ambulance location with random delays and travel times. Health Care Management Science, 11, 262-274.

Jagtenberg, C. J., Van Den Berg, P. L. & Van Der Mei, R. D. 2017. Benchmarking online dispatch algorithms for emergency medical services. European Journal of Operational Research, 258, 715-725.

Jarvis, J. P. 1981. Optimal assignments in a markovian queueing system. Computers & Operations Research, 8, 17-23.

Kepaptsoglou, K., Karlaftis, M. & Mintsis, G. 2012. Model for planning emergency response services in road safety. Journal of Urban Planning and Development ASCE, 138, 18-25.

Kim, B. 2016. Exploring emergency areas for medical service using microscopic traffic simulation model. Spatial Information Research, 24, 75-84.

Knight, V. A., Harper, P. R. & Smith, L. 2012. Ambulance allocation for maximal survival with heterogeneous outcome measures. Omega, 40, 918-926.

Kolesar, P. & Walker, W. E. 1974. An algorithm for the dynamic relocation of fire companies. Operations Research, 22, 249-274.

Krishnan, K., Marla, L. & Yue, Y. Robust ambulance allocation using risk-based metrics. 2016 8th International Conference on Communication Systems and Networks (COMSNETS), 5-10 Jan. 2016 2016. 1-6.

Lam, S. S. W., Zhang, J., Zhang, Z. C., Oh, H. C., Overton, J., Ng, Y. Y. & Ong, M. E. H. 2015. Dynamic ambulance reallocation for the reduction of ambulance response times using system status management. The American Journal of Emergency Medicine, 33, 159-166.

Maxwell, M. S., Henderson, S. G. & Topaloglu, H. 2009. Ambulance redeployment: An approximate dynamic programming approach. Winter Simulation Conference. Austin, Texas: Winter Simulation Conference.

Maxwell, M. S., Restrepo, M., Henderson, S. G. & Topaloglu, H. 2010. Approximate dynamic programming for ambulance redeployment. INFORMS Journal on Computing, 22, 266-281.

Mccormack, R. & Coates, G. 2015. 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, 294-309.

Mclay, L. A. & Mayorga, M. E. 2010. Evaluating emergency medical service performance measures. Health Care Management Science, 13, 124-136.

Miller, T., Friesz, T., Tobin, R. & Kwon, C. 2007. Reaction function based dynamic location modeling in stackelberg–nash–cournot competition. Networks and Spatial Economics, 7, 77-97.

Nair, R. & Miller-Hooks, E. A case study of ambulance location and relocation. Presentation at 2006 INFORMS Conference, 2006.

Panahi, S. & Delavar, M. 2009. Dynamic shortest path in ambulance routing based on gis. International journal of Geoinformatics, 5.

Pons, P. T., Haukoos, J. S., Bludworth, W., Cribley, T., Pons, K. A. & Markovchick, V. J. 2005. Paramedic response time: Does it affect patient survival? Academic Emergency Medicine, 12, 594-600.

Pons, P. T. & Markovchick, V. J. 2002. Eight minutes or less: Does the ambulance response time guideline impact trauma patient outcome? The Journal of Emergency Medicine, 23, 43-48.

Restrepo, M., Henderson, S. G. & Topaloglu, H. 2008. Erlang loss models for the static deployment of ambulances. Health Care Management Science, 12, 67-79.

Reuter-Oppermann, M., Van Den Berg, P. L. & Vile, J. L. 2017. Logistics for emergency medical service systems. Health Systems, 6, 187-208.

Revelle, C. & Hogan, K. 1989. The maximum availability location problem. Transportation Science, 23, 192-200.

Sánchez-Mangas, R., García-Ferrrer, A., De Juan, A. & Arroyo, A. M. 2010. The probability of death in road traffic accidents. How important is a quick medical response? Accident Analysis & Prevention, 42, 1048-1056.

Savas, E. S. 1969. Simulation and cost-effectiveness analysis of new york's emergency ambulance service. Management Science, 15, B-608-B-627.

Schmid, V. 2012. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. European Journal of Operational Research, 219, 611-621.

Serra, D. & Marianov, V. 1998. The p-median problem in a changing network: The case of barcelona. Location Science, 6, 383-394.

Silva, T. H., Melo, P. O. S. V. D., Almeida, J. M. & Loureiro, A. a. F. 2014. Large-scale study of city dynamics and urban social behavior using participatory sensing. IEEE Wireless Communications, 21, 42-51.

Simpson, N. C. & Hancock, P. G. 2009. Fifty years of operational research and emergency response. J Oper Res Soc, 60, S126-S139.

Su, S. & Shih, C.-L. 2003. Modeling an emergency medical services system using computer simulation. International Journal of Medical Informatics, 72, 57-72.

Toregas, C., Swain, R., Revelle, C. & Bergman, L. 1971. The location of emergency service facilities. Operations Research, 19, 1363-1373.

Valenzuela, T. D., Roe, D. J., Cretin, S., Spaite, D. W. & Larsen, M. P. 1997. Estimating effectiveness of cardiac arrest interventions a logistic regression survival model. Circulation, 96, 3308-3313.

Valinsky, D. 1955. Symposium on applications of operations research to urban services—a determination of the optimum location of fire-fighting units in new york city. Journal of the Operations Research Society of America, 3, 494-512.

Van Essen, J. T., Hurink, J. L., Nickel, S. & Reuter, M. 2013. Models for ambulance planning on the strategic and the tactical level, Netherlands, TU Eindhoven, Research School for Operations Management and Logistics (BETA).

Vasić, Č., Predić, B., Rančić, D., Spalević, P. & Avdić, D. 2014. Dynamic relocation of emergency ambulance vehicles using the avl component of the gps/gprs tracking system. Acta Polytechnica Hungarica, 11.

Wang, T., Wang, F., Yin, X., Han, N., Zhang, P., Kou, Y. & Jiang, B. 2015. Changes and trends of pre-hospital emergency disease spectrum in beijing in the past decade (from 2003 to 2012). Journal of Local and Global Health Science, 2015, 34.

Westgate, B. S., Woodard, D. B., Matteson, D. S. & Henderson, S. G. 2013. Travel time estimation for ambulances using bayesian data augmentation. 1139-1161.

Who 2011. Global plan for the decade of action for road safety 2011-2020. World Health Organization.

Yang, S., Hamedi, M. & Haghani, A. 2005. Online dispatching and routing model for emergency vehicles with area coverage constraints. Transportation Research Record: Journal of the Transportation Research Board, 1923, 1-8.

Yue, Y., Marla, L. & Krishnan, R. An efficient simulation-based approach to ambulance fleet allocation and dynamic redeployment. AAAI, 2012.

Zaffar, M. A., Rajagopalan, H. K., Saydam, C., Mayorga, M. & Sharer, E. 2016. 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.

Zhang, L. 2012. Simulation optimisation and markov models for dynamic ambulance redeployment. ResearchSpace@ Auckland.

Zhang, O., Mason, A. & Philpott, A. Simulation and optimisation for ambulance logistics and relocation. Presentation at the INFORMS 2008 Conference, 2008.

Zhu, S., Kim, W. & Chang, G.-L. 2012. Design and benefit-cost analysis of deploying freeway incident response units: Case study for capital beltway in maryland. Transportation Research Record: Journal of the Transportation Research Board, 104-114.

5. AN ACTIVE LEARNING METAMODELING APPROACH FOR POLICY ANALYSIS: APPLICATION TO AN EMERGENCY MEDICAL SERVICE SIMULATOR[12]

Francisco Antunes[13], Marco Amorim[14], Bernardete Ribeiro[13], Francisco Pereira[15]

Abstract

Simulation approaches constitute a well-established tool to model, understand and predict the behavior of transportation systems, and ultimately to assess the performance the transportation policies. Due to their multidimensionality and stochastic natures, such systems are not often approachable through conventional analytic methods, making simulation modeling the only reliable tool of study. Nevertheless, despite its clear advantages, simulation models can turn out to be computationally expensive when embedded with enough detail. An immediate answer to this shortcoming is the use of simulation metamodels that are designed8 to approximate the simulators' results. In this work, the authors propose a metamodeling approach based on active learning that seeks to improve the exploration of the simulation input space and the associated output behavior. A Gaussian Process (GP) is used as a metamodel to approximate the simulation results. The GPs are able to nicely handle the uncertainty associated with their predictions, which eventually can be improved with active learning through simulation requests. This provides a practical way to analyze the simulator's behavior and therefore to assess the performance of policies regarding the underlying real-world system in study, while allowing, at the same time, to bypass exhausting experimental exploration. The authors illustrate their methodology using an Emergency Medical Service (EMS) simulator. Two outputs are analyzed and compared, namely, the survival rate and average response time. The medical emergency response time recommendation of eight minutes is explored as well its relation with the survival rate. The results show that this methodology is able to identify regions in the simulation input space that might affect the application performance of medical policies with regards to emergency vehicles services.

Keywords: Active Learning, Simulation Metamodeling, Gaussian Processes, Emergency, Medical Service

---

[12] Submitting to Simulation Modelling Practice and Theory – review received and revision done on 17th May

[13] University of Coimbra – Faculty of Sciences and Technology, Coimbra, Portugal

[14] CITTA, University of Porto – Faculty of Engineering, Porto, Portugal

[15] Technical University of Denmark, Bygningstorvet, Denmark

5.1. INTRODUCTION

Real-world transportation systems are characterized by their overwhelming complexity, due to the multitude of variables involved and corresponding relationships. In order to model, understand, and then predict the behavior of such systems and to assess their performances, simulation modeling is often the only reliable tool available. Especially due to their high dynamism and dimensionality, as well as their intrinsic stochastic nature, these systems cannot be evaluated studied analytically (Law et al., 2007).

Simulation models are virtual representations of the reality, often in a sufficiently simplified form, that are considered as experimental virtual environments to test different system designs, and therefore to understand the impact of certain policies and interventions (Marengo, 2014). Due do their exploratory nature, simulation tools prove to be highly advantageous for policy analysis (Bankes, 1992), (Bankes, 1993). However, simulation models can become computationally expensive to run, exhibiting prohibitive runtimes along with great workloads. To address this shortcoming, simulation metamodels are usually employed to approximate the simulation model itself and thus the function inherently defined by it.

Along with the consideration of simulation metamodels with the objective to decrease the burden of conducting expensive computer experiments, a machine learning paradigm, called active learning (Settles, 2010), can be also taken into account. Active learning is particularly useful in situations in which data is difficult to obtain, as it aims to enhance the models' predictive performance with fewer training data points. Therefore, both active learning and simulation metamodeling can be jointly used, on the one hand, to minimize the need for simulation runs, and on the other hand to achieve and maintain a reasonably good approximation of the simulation model. The Gaussian Process (GP) framework (Rasmussen and Williams, 2005), is a well-known modeling tool, widely applied in numerous research and application fields. Due to its Bayesian formalism and highly non-linear properties, it constitutes an excellent option for designing active learning strategies based on simulation metamodeling settings.

This work presents an active learning metamodeling methodology to address the problem of policy analysis within the context of computationally expensive simulation models. A GP is considered to approximate the simulator's behavior and then used to explore the simulation input space. The fitness of the GP is then iteratively improved with active learning via simulation requests, by decreasing the associated variance of the given predictions over the simulation input region in study. This provides an alternative way to perform policy analysis, while avoiding, at the same time, a potential large number of simulation runs.

The presented methodology is tested using an Emergency Medical Service (EMS) simulator. Two simulation outputs are studied, namely, the average survival rate and the average response time. Then, motivated by (Pons and Markovchick, 2002), the medical emergency response time recommendation of 8 minutes (480s) is analyzed as well as its relation with the survival rate. The results, as proof of concept, show that this methodology is able to identify regions within the simulation input space that directly influence the successful application of medical policies with regards to emergency vehicle dispatching services.

5.2. LITERATURE REVIEW

The development and application of simulation metamodels (Kleijnen,1975, Kleijnen, 1979, Kleijnen, 1987, Friedman, 2012) can be traced back to the 70’s (Barton, 1998). Their main purpose is to serve as parsimonious approximations for simulation models, so that expensive simulation runs can be avoided. Specific features such as mathematical simplifications, speed and interpretability are usually attributed to metamodels. Consequently, the use of metamodels within simulation analysis provides an additional level of understanding of the underlying system, as well as of the relationships between the system input and output variables, while maintaining a computationally simple and economic approach to the problem (Friedman and Pressman, 1988). Simulation metamodels are often described by computationally fast and easy-to-implement functions that approximate the true but unknown function intrinsically defined by the simulation model itself. It is common that many of these inputs are shared with those of the simulation model, although it is not entirely necessary, as Kleijnen and Sargent (2000) points out.

The earliest applications of metamodeling involved simple queuing simulation systems (Kleijnen, 1975) and used multi-linear regression metamodels. At the time, as the computational resources were evidently low and scarce, when compared to the current days, the use of metamodels emerged as a practical tool to overcome the difficulties posed by even the simplest simulation systems. Although the technology evolved ever since, essentially providing more computational power, it also leveraged the demand and the opportunities for modeling increasingly complex system models. Both computational power and system models’ complexity increased in the same direction and with a similar intensity, therefore explaining the need for the use of metamodels nowadays. This shows that the application of metamodeling techniques is not only exclusively related with the available computing power but also to the demand of highly detailed models, which are eventually conditioned by it. Notice that with more realistic models usually comes large or even prohibitive.

Simulation running times, which cannot be used for real-time applications or practical simulation behavior analysis. In these situations, the use of metamodels as an auxiliary tool is always preferred (Kleijnen and Sargent, 2000) and can eventually be used to explore the simulation behavior in a less expensive manner.

The GP framework is a well-known modeling approach widely used as a simulation metamodel (Boukouvalas, 2010, Chen et al., 2011, Chen, Hadinoto, Yan, and Ma, Conti and O’Hagan, 2010, Kleijnen, 2009). Initially, GP-based metamodels had only been used in deterministic simulations. Nevertheless, (Van Beers and Kleijnen, 2003) started using GPs for random simulations, showing that they had great potential in this kind of applications too. Since then, several similar approaches arise (Kleijnen and Van Beers, 2005, Boukouvalas et al., 2009, Boukouvalas, Cornford, and Singer, Ankenman et al., 2010, Ankenman, Nelson, and Staum). Jones and Johnson (2009) generally discusses the use of GP as a reliable metamodel for the design and analysis of computer experiments. Boukouvalas (2010) constructs two GP representations and develops an experimental design to extend the metamodel framework to account for heteroscedasticity in the simulation output. In Kleijnen and Van Beers (2004), the authors developed a method based on application-driven sequential designs using a GP metamodel. Wang et al. (2005) shows that the GP, as a metamodeling technique, is able to accurately approximate the complex functions often associated with non-linear and non-monotonic probabilistic design space. In Bastos and O'Hagan (2009) the authors introduce several diagnosis metrics to validate the GP framework as a simulation metamodel.

In the context where simulation data can be difficult to obtain (e.g. computationally expensive simulation experiments) in a systematic way, active learning can prove to be a powerful learning paradigm to enhance the application of simulation metamodels. As a subfield of supervised machine learning, active learning is essentially an iterative sampling strategy that allows any algorithm designed upon it to actively select the data points from which it learns. Thus, instead of selecting a large number of random points, the algorithm searches for the most informative ones, in a sequential way, so that both the model training efficiency and its prediction performance are improved with as few training data points as possible (Settles, 2010).

According to Wang and Zhai, (2016), an arbitrary active learning strategy encompasses five essential elements enclosed in the following quintuple ***(L, U , M, O, Q)***. First, ***L*** is the labeled training date set. Then, the set of the unlabeled data points is represented by ***U***. Generally, **#*U* >> #*L***, i.e., the number of unlabeled data points is much higher than that of the labeled ones. The machine learning model is represented by ***M***. Depending on the nature of the problem being modeled, it can be a classification or regression model, which in turn affects the nature of the labels in ***L*** of being discrete or continuous, respectively. ***O*** denotes the oracle whose role is to provide labeled instances from the underlying process in study. Finally, ***Q*** is the query function that encodes the strategies and criteria for finding and selecting the most informative instances of ***U*** to be added to ***L***. There have been many query strategies formulations which essentially translate into different perspectives to approach the problem in question. As mentioned by Settles (2010), depending on both the nature of the problem and the model being used, several query frameworks can be adopted (Settles, 2010), such as uncertainty sampling (Lewis and Gale, 1994), query-by-committee (Seung et al., 1992), expected model change (Settles et al.,2008) and error reduction (Roy and McCallum, 2001), variance reduction (Geman et al., 1992) or density-weighted methods (Settles and Craven, 2008).

Closely related to the concept of the query function is the definition of a stopping rule. Due to its iterative nature, techniques based on active learning must be stopped at a certain time, either manually by the user or automatically by a stopping criterion. In both cases, the stopping rule must take into account the trade-off between the overall prediction performance and generalization capacity of the machine learning model and the associated costs of acquiring new labeled data. Such costs could be, for example, the computational workload or running time associated to any simulation model.

Active learning has been used in a variety of research fields. However, this paper focus on those applications involving simulation metamodels, particularly the GPs, where the simulation model plays the role of oracle. As mentioned earlier, by providing a fully non-linear Bayesian approach, the GPs allows for an intuitive way to develop active learning algorithms due to their ability to explicitly model the uncertainty present in the data. As high variance can be associated with high degree of uncertainty, both the posterior mean and variance, provided by the GPs for any given data point, can be directly used to explore the simulation input space and thereby guide the search for the most informative data points. Such active learning schemes involving GPs are usually associated to exploration-exploitation problems, often studied in Bayesian Optimization contexts (Ling et al., 2016).

According to Schulz et al. (2017), while a pure exploration approach aims to learn an unknown function of interest as accurately and fast as possible, in an exploration-exploitation setting the goal is to find the input that maximizes the output of an unknown function, in an equally fast manner. Under an active learning strategy, Bayesian optimization uses a reward function to more efficiently select the next unlabeled data point, according to an automatic trade-off between the domain regions where the objective function is very uncertain (exploration) and those of where the same function is expected to be attain high value (exploitation), as seen in Brochu et al. (2010).

Within the transportation literature the application of simulation metamodels is still rare and relatively recent. The available research can be broadly categorized into traffic prediction and optimization of networks, as mentioned by Song et al. (2017). Some of such works include, for example, metamodeling for mesoscopic simulation (Ciuffo et al., 2013, Ciuffo, Chen et al., 2015) and for travel behavior and dynamic traffic optimization (Zhang et al., 2014). In Antunes et al. (2018) the authors proposed a restricted batch-mode active learning strategy to address the problem of efficiency of the metamodeling process using a simple traffic simulation and a Demand-Responsive Transportation (DRT) system simulator.

5.3. METHODOLOGICAL APPROACH

In this work, a straightforward pool-based active learning strategy is adopted. The experimental design is depicted in Figure 21. Here the unlabeled data set ***U*** is entirely available for querying and represents the simulation input region in which we aim to explore the simulator's behavior. The pool of labeled instances ***L*** is comprised of simulation results, i.e., input-output tuples. The machine learning model ***M*** is a GP, whereas the query function ***Q*** is based on the analysis of the predictive variance provided by the latter at each point in ***U***. The general idea within this experimental design is to assume that the functional relationship between the simulation input vector $\boldsymbol{x}$ and the output $\boldsymbol{y}$ is described by a GP. After the GP is fitted to $\boldsymbol{L}$, the provided conditional distribution for is used to predict the output values for $\boldsymbol{U}$. This makes it possible to bypass simulation runs and to approximate the simulator behavioral structure, making the exploration process more efficient. The predictive variance is used as a measure of fitness and it should decrease as the iterative process evolves. Finally, this trained GP model is used as a simulation metamodel to explore the behavior of the simulator and then to conduct policy analysis and assessment. More details regarding the GP framework, as well as the adopted methodology are presented in the following. Regarding the implementation of the GP framework, a freely available Matlab package from Rasmussen and Williams, 2005) was used. For the GP mean covariance function, the widely known and applied Squared Exponential, or Radial Basis Function, given by $\boldsymbol{k_f(x, x') = \sigma^2 exp(-(||x-x'||^2) / (2l^2))}$, was selected. Here, $\boldsymbol{\sigma^2}$ and $\boldsymbol{l^2}$ are respectively the signal variance and the characteristic length-scale. On the other hand, for the GP mean, a constant function dependent on the average of output values of the training sets was used.

5.3.1. GAUSSIAN PROCESSES

Despite being quiet and old topic in the field of probability and statistics, the application of Gaussian Processes within machine learning tasks has emerged in the past decade. As seen in Rasmussen and Williams (2005), a Gaussian Process (GP) is a stochastic process from which each finite set of variables follows a multivariate Gaussian distribution. It is usually denoted as $\mathcal{GP}(m_f(\mathbf{x}), k_f(\mathbf{x}, \mathbf{x}^i)$, where $m_f(\mathbf{x})$ and $k_f(\mathbf{x}, \mathbf{x}^i)$ are respectively a mean and a covariance function, with $\mathbf{x}$ and $\mathbf{x}^i$ being two different input data points. Thus, a GP is sufficiently characterized by these two functions.

From a regression point-of-view, and as an intrinsic Bayesian approach, the GP modeling assumes a prior over functions, i.e., in the functional dependency established by $\boldsymbol{y} = \boldsymbol{f}(\mathbf{x})$, where $E \sim \mathrm{N}(0, \sigma^2)$, it follows that $\boldsymbol{f}(\mathbf{x}) \sim \mathcal{GP}(m_f(\mathbf{x}), k_f(\mathbf{x}, \mathbf{x}^i))$. Most of the common mean and covariance functions typically have several free parameters, also called hyper-parameters of the GP, which can be optimized by marginal likelihood maximization subjected to the training data. The conditional distribution of a new unlabeled data point $\boldsymbol{x}$ is then given by $\boldsymbol{f} \,|X, \mathbf{y}, \mathbf{x} \sim \mathrm{N}(k^{\mathrm{T}} [K\ ]^{-1}\mathbf{y}, k - k^{\mathrm{T}}[K_y]^{-1}k_{f*})$, with $k_{f*} = k_f(X, \mathbf{x}_*)$, $k_{f**} = k_f(\mathbf{x}_*, \mathbf{x}_*)$, K_y the covariance matrix, X the design matrix and $\mathbf{y}$ the set of training target values. Note that the prediction provided by the GP is fully-defined Gaussian distribution, rather than a single point-wise estimate. This Bayesian property allows the GP to encapsulate the uncertainty not only of its owns predictions, but also that of the underlying signal being modeled. In turn, this uncertainty can be intuitively used to as an information criterion to design active learning strategies.

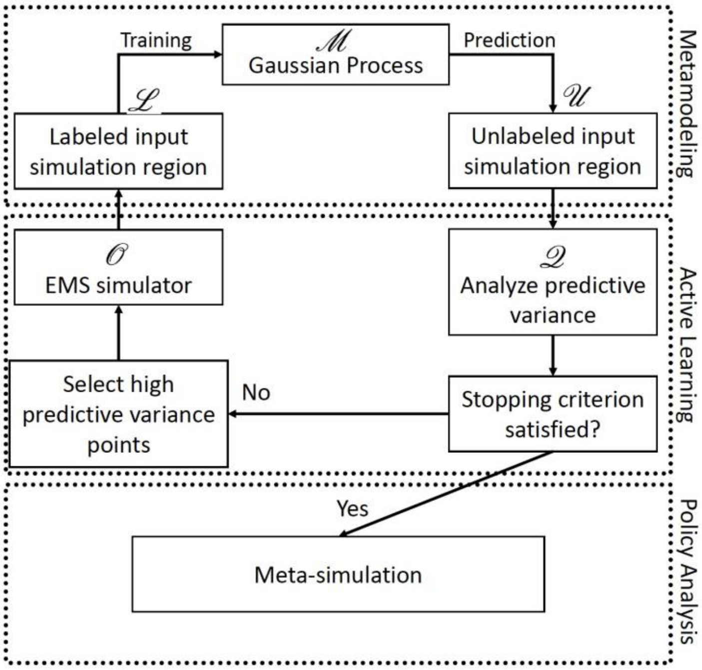


Figure 21. Active learning metamodelling experimental design. It is divided into three main blocks or steps. First, the simulation metamodeling approximates the simulation in question using a GP. Then, an active learning strategy is used to iteratively increase the fitting quality of the GP, in this case, by decreasing its total predictive variance across the unlabeled input simulation region. Finally, when the stopping criterion is verified, policy analysis by means of the provided meta-simulator is conducted.

5.3.2. METAMODELLING STRATEGY

This work follows a modeling strategy similar to the one presented by Antunes et al. (2018). The strategy used in this work is presented in Figure 22. Active learning metamodeling strategy.Figure 22 in the form of pseudo-algorithm. The algorithm starts with the initialization data set which is comprised of the first simulation input-output results. A GP is trained in this data set ***(L)***, via likelihood maximization, and both the predictive mean and variance are obtained for each point in the input simulation region of interest ***(U)***. Then, the algorithm selects the top five data points with the highest predictive variance values, $\boldsymbol{top_5}$. This is meant to force an active selection of points, rather than selecting them at random, as the hypotheses is that the points with highest variance contain more information regarding the underlying process than the ones with lower values of variance. After this selection, the algorithm requests the simulator for the real output values (labels) corresponding to $\boldsymbol{top_5}$. This is also called a batch-mode active learning. It is particularly useful as it allows for the query processes to be parallelized. In this case, the points are being selected in batches of five. Eventually, these points are added to ***L***. The process repeats until the stopping criterion is satisfied.

The stopping criterion is defined by the relation between the Initial Total Variance (ITV) and the Current Total Variance (CTV). Whereas the former is computed and saved computed in the first iteration, the latter represents the current amount of variance summed over ***U*** and updated in each of the following iterations.

As ***α*** lies in ***(0, 1)***, the process stops in CTV reaches a fraction of ITV. In this work, ***α = 0.1***, which means that the stopping rule is satisfied when CTV is less than ***(1 − α)%***, or ***90%*** of ITV .

Inputs: $\alpha \in (0, 1)$, L, U.

While $CTV \geq ITV$ **do**

1: Train the GP in L and predict the output values for each point in U, $k^{\mathrm{T}}\,[K_y]^{-1}\mathbf{y}$, and obtain their corresponding

predictive variances, $k_f \quad - k^{\mathrm{T}}\,[K\;]^{-1}k$.

2: Determine the top 5 highest predictive variance points in U, top_5

3: By simulation requests, obtain the true values for top , L^{top} as the new labeled set.

4: Expand the labeled pool: $\mathrm{L} = \mathrm{L} \cup \mathrm{L}^{top_5}$.

5: Update CTV.

end

Output: GP trained in L and associated predictions over U.

Figure 22. Active learning metamodeling strategy.

Note that in the first iteration, CTV = ITV holds. However, has the process evolves, CTV has a tendency to decrease since the GP will inevitably acquire more training data points, and thus greater knowledge about the simulation function. It is also worth noting that number of iterations required to satisfy this criterion is directly dependent on the size of the initial training set as well as on the value of α itself. If the initial set is both sized up and representative enough, then the GP may be able to generalize well for the unobserved points present in ***U***. On the other hand, if ***α*** is near ***1***, then the criterion will be easily satisfied in a few iterations. Therefore, the right trade-off between these two elements must be carefully taken into account.

5.3.3. EMS SIMULATOR

In the emergency medical service planning, evaluating strategic and tactical decision in the real system, be either policies or operational solutions, is usually physically unfeasible. Optimization is then the preferable tool to mimic the system real conditions and produce insightful assessments of decisions planned to be implemented. This is a synthetic process to provide empirical evidences of how a certain choice or change might impact a certain targeted performance metric.

A simulation algorithm to numerically compute EMS solutions performance in terms of average response time and average survival is adapted from the work of Amorim et al. (2018), and is resumed in Algorithm 1 depicted in Figure 23. This optimization model is used to assess the performance of station and vehicle locations in an urban area. It also provides a platform to measure the performance of dispatching policies and other tactical decisions. This works focuses on strategic and tactical decisions and their performance using two metrics: the classical average response time, and a survival metric that tries to capture victims' outcomes.

The simulation model can be described as an agent-based model controlled by a city agent that takes the role of the emergency medical service, and allocates and dispatches vehicles using a closest dispatching rule (Haghani and Yang, 2007, Jagtenberg et al., 2017, Yang et al., 1923). A network agent simulates traffic stochastic conditions and EMS vehicle movements by using nodes and arcs as an abstraction of the reality, pre-computing travel times for different periods of the day and the week using Google's Directions API. Events are agents that abstract EMS calls and are triggered according to a historical database and a stochastic location change, denoted by w_1. At trigger time, the city allocates them a vehicle agent for assistance.

The network agent informs the vehicle agent of the traffic conditions (translated to travel times) as it moves from node to node to assist and transport event agents. The assisting time is a stochastic variable that assumes a gamma distribution with a mean of 10 minutes and a standard deviation of 5 minutes, according to observations from Pons and Markovchick (2002).

The stochastic location change captures a possible randomness in the demand by re-allocating a medical emergency event e_p at node p to node p', with a probability of w_1, according to $e_{p'} = p(g(E_s), e_p, w_1, w_2)$. Here, $g(E_s)$ is a function that chooses a location from a set of possible locations weighted by the observed demand at period s, and $p(e_1, e_2, w_1, w_2)$ is a function that picks either e_1 or e_2 with probability w_1 and w_2, respectively, where $w_1 + w_2 = 1$ must hold.

Algorithm 1 Simulation algorithm

Definitions:
$\boldsymbol{N}$ = set of nodes $\boldsymbol{n}$
$\boldsymbol{n}$ = node, where $\boldsymbol{s}$ = node of type station and $\boldsymbol{h}$ = node of type hospital
$\boldsymbol{V}$ = set of vehicles $\boldsymbol{v_s}$
$\boldsymbol{v_s}$ = vehicle in station $\boldsymbol{s}$
$\boldsymbol{S}$ = set of stations $\boldsymbol{s}$
$\boldsymbol{H}$ = set of hospitals $\boldsymbol{h}$
$\boldsymbol{E}$ = set of events $\boldsymbol{e^t_n}$
$\boldsymbol{e^t_n}$ = emergency event occurring at node $\boldsymbol{n}$ during $\boldsymbol{t}$
$\boldsymbol{M}$ = set of matrices $\boldsymbol{M^t}$
$\boldsymbol{M^p}$ = matrix of real travel times for period $\boldsymbol{p}$
$\boldsymbol{T}$ = total simulation time
$\boldsymbol{t}$ = time
step = temporal resolution
f( ) = programming function

While $\boldsymbol{t < T}$:

1. **Update city( )** "set $\boldsymbol{t}$ and activate $\boldsymbol{e^t_n}$."
2. **Update network( )** "interact through every $\boldsymbol{v_s}$ to travel one ***step*** and transfers it to destination nodes"
3. **Update events( )** "activates $\boldsymbol{e^t_n}$ and the vehicle dispatching algorithm"
   - -Network calculates time travel from all stations
   - -Network returns the shortest one
   - -Vehicle dispatching algorithm runs
4. **Update vehicles job( )** "updates $\boldsymbol{v_s}$ status"
   - -If $\boldsymbol{v_s}$ arrived to $\boldsymbol{e^t_n}$, activate assisting timer
   - -If assisting timer ends, request network to be processed to $\boldsymbol{h}$
   - -If $\boldsymbol{v_s}$ arrived to $\boldsymbol{h}$, transfers $\boldsymbol{v_s}$ to $\boldsymbol{s}$.
5. **Update results( )** "calculates the EMS performance at the current ***step***"
6. $\boldsymbol{t = t +}$ ***step***

Figure 23. Simulation algorithm for the used Medical Emergency Service.

On the other hand, traffic stochasticity is controlled by parameter $\boldsymbol{\varepsilon}$. It introduces an error term in any observed travel time $\boldsymbol{r_{s,l,p}}$, which measures the time it takes to travel from point $\boldsymbol{i}$ to point $\boldsymbol{p}$ during $\boldsymbol{s}$, according to $\boldsymbol{r^l = r_{s,l,p} + r_{s,l,p} \times N(\mu, {}^{\varepsilon})}$. When the mean is set to zero and the standard deviation to $\boldsymbol{\varepsilon/2}$, a confidence of 95% exists that the travel time has at most an error of $\boldsymbol{\varepsilon \in [0, 1]}$.

5.4. RESULTS

In this section, the presented methodology is applied to the Emergency Medial Service (EMS) simulator developed by Amorim et al. (2018). As previously seen, this simulator is designed in terms of three kinds of input dimensions, namely, the location change probability, traffic error and vehicle station locations. In terms of value ranges, the first two lie in the interval [0, 1], whereas the locations assume discrete positive values, [0, 1, 2, 3, ...), representing the number of vehicles allocated to each position. Figure 24 shows the location of these stations, 90 in total, scattered in the city of Porto, Portugal. Note that not all the depicted labels follow a sequential order. The last four locations are respectively labeled with 96, 108, 109 and 112. The considered simulation outputs are the average survival rate and the average response time.

In Table 10 a sample of the used data is presented. The dimension, or feature, denoted $\boldsymbol{x_1}$ is the location change probability, whereas $\boldsymbol{x_2}$ is the traffic error. The locations and the associated number of vehicles are coded in features $\boldsymbol{x_3}$ to $\boldsymbol{x_{92}}$. With respect to the simulation outputs, $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ are the survival rate and response time, respectively.

Each output was modeled independently, i.e., a different GP was used to approximate the simulation results. Figures 25-28 present the obtained results. Notice that Figures 25(a)-(b) and 27(a)-(b) do not depict any kind of time series, but rather a straightforward way to simply represent the prediction originated from high- dimensional data and their corresponding confidence intervals. Therefore, particularly in these representations there is no notion of neighborhood between observations, nor the sequential order of each observation is of any relevance.

In both cases, the active learning metamodeling process started with the GP training using 100 random simulation points. Then, the prediction was conducted over 906 points scattered along the simulation input space. For $\boldsymbol{y_1}$, Figure 25(a) shows the first GP approximation along with a confidence interval of 95%. As expected for this first iteration, the uncertainty of the model, which is encoded into the variance of each prediction, is relatively high. However, after 75 iterations, this variance decreases, as depicted in Figure 25 (b). For $\boldsymbol{y_2}$, similar conclusions can be drawn. However, for the same stopping criterion with $\boldsymbol{\alpha = 0.1}$, the active learning procedure only required 59 to achieve a reduction of 90% of the total initial variance (see Figure 25). In Pons and Markovchick (2002), an empirical study involving 3576 patients transported to a single Level I trauma center was conducted in order to assess the 8 min guideline for ambulance response. The authors concluded that there was no significant difference in the survival rate, due to traumatic injuries, between the patients who were assisted within and above the established response time policy, respectively. In the same work, it is mentioned that the mortality odds ration is of 0.81 for response times greater than 480s. Obviously, the odds for the patients' survival were of 0.19. Taking these two elements into account, a meta-simulation analysis was conducted in terms of the two outputs provided by the studied EMS simulator. Figure 26(a) shows that the obtained survival rates averages were conditioned by both the traffic error and location change probability inputs. From traffic errors in the order of 60%, it is visible that emergency cases associated to lower rates of survival start to emerge. On the other hand, as the probability of location change increases, the rate of survival also decreases. Most of these low survival observations are concentrated in the upper right corner, slightly within [0.5, 1] × [0.6, 1]. Similar conclusions can also be derived from the observation of Figure 26 (b) and (c). Again, and this time with respect to the number of vehicles in location 1, which corresponds to variable x3, most of the mortal occurrences are associated with higher values of location change probability and traffic error. This is particularly more evident for the latter simulation input.

Regarding the traffic error and taking into account the earlier presented details of the use EMS simulator, the obtained results meet the initial expectations. This simulation input represents the error associated with the traffic prediction in comparison with the real traffic, immediately prior to the vehicles dispatching. Therefore, it makes sense that higher traffic errors may lead to inadequate operating decisions, resulting in higher response times and consequently in more fatal occurrences. With respect to the location change probability the results also confirm the intuition. Increasing this probability will lead to higher variability in the location of the life-threading events or emergency calls, in comparison with historical data.

For the average response time, similar results were attained. In Figure 28(a), it can be concluded that the observations associated with response times greater than or equal to 480 seconds are concentrated in [0, 1] × [0, 6]. Contrary to what one could expect, not so many of these observations had led to survival rates below 0.19. This matches with the conclusions made by Pons and Markovchick (2002), i.e., that average response times greater than 8 minutes do not necessary lead to higher mortality rates. Nevertheless, high response times combined with high probability of call location change, seem to lead to higher values of mortality rate. In the end, it is true that when the traffic error increases, the delays in the vehicle arrivals increase accordingly. It is important to take into account this kind of variation encoded into these two simulations inputs. Note that the traffic error does not have a direct implication on the vehicles' travel times. Instead, what is being measured by this EMS simulator is to which extent traffic errors lead to bad choices of dispatching the correct vehicles, both in terms of number and station locations. Given a certain emergency call, and ideally assuming no traffic congestion, the obvious vehicle to dispatch would be the closest one. However, when traffic congestion exists and, besides that, the real traffic information is only available with error, the dispatching solution is not so obvious. Moreover, the distance is no longer measured in space units, but rather in time ones, due to existence of traffic congestion, which in turns makes the problem more challenging. Take, for instance, the following example where station 1 is 10 minutes from the emergency (E) event location and station 2 is 20 minutes way. Given an error of ±20% means that emergency service operator may assume that stations 1 and 2 are, respectively, 12 and 16 minutes from E. In practice, especially in this kind of life-threatening situations, this represents a great error with potentially dramatic consequences. For this case, the decision result will be the same, i.e., a vehicle will be dispatched from location 1, which turns out to be the best decision. However, consider an alternate configuration, where both station 1 and 2 are, respectively, 10 and 11 minutes from E. Here, the same traffic error could lead the operator to take a poor decision based on the assumption that station 1 is distanced in 12 minutes and station 2 in 9 minutes, which, in reality, is not true. In such situation, this poor decision would be, for example, to send a vehicle from location 2, which is, in fact, 1 minute further than the other. Therefore, the same error can lead to different decisions and different outcomes.

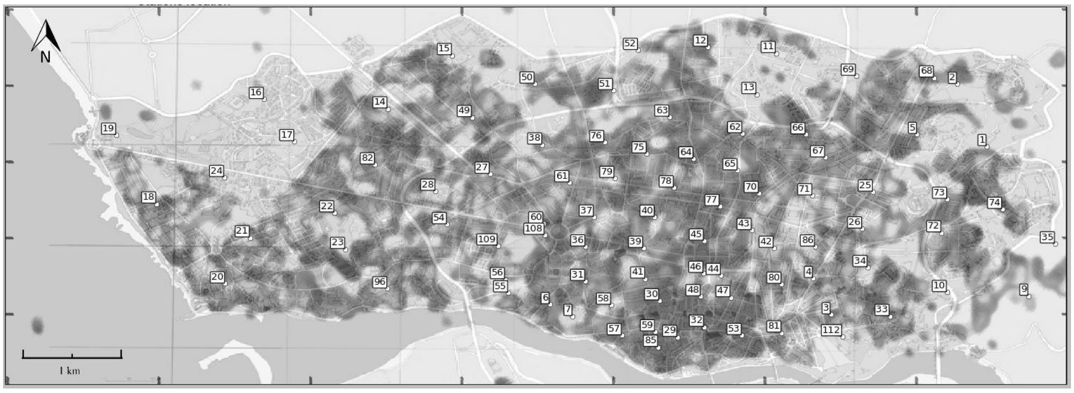


Figure 24. Locations of the 90 emergency vehicle stations in Porto, Portugal.

Table 10. A data sample showing the dimensional structure with 92 features. X_1, X_2, Y_1 lie in the interval (0, 1), $X_1 - X_{92}$ assume discrete values in {0, 1, 2, ...} and Y_2 is a real-valued variable.

a	x_2	x_3	x_4	…	x_{90}	x_{91}	x_{92}	y_1	y_2
0.0445	0.2921	0	0	…	0	0	0	0.3547	379.027
0.4923	0.7756	0	0	…	2	1	0	0.1953	521.336
0.0708	0.3740	1	0	…	0	0	0	0.3800	331.063
0.2199	0.6818	2	1	…	0	1	0	0.2252	515.080

On the other hand, the probability change location is especially important to induce a stochastic behavior to the emergency calls' locations, so that the conclusions are not drawn exclusively from historical data. The simulation model must take into account the inherent dynamics of emergency events and should be able to respond accordingly.

In order to validate the generalization capacity of the obtained GP approximations, for each simulation output, 30 random runs using 200 points were conducted in a 10-fold cross-validation scheme. These points were randomly sampled from the simulation input space. To evaluate the predicted performance of these GPs, five well-known metrics were used, namely, the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Root Squared Error (RRSE) and the Pearson's correlation coefficient (Corr.). The results are summarized in Table 11. These metrics were computed by comparing the predicted values given by the GP (from the last iteration of the active learning procedure) against the real known output values. The GP yields slightly better performance for the survival rate output, despite requiring more iterations to satisfy the active learning stopping criterion (see Figure 25(c) and Figure 28(c)). Nevertheless, in both cases the GP presents a good generalization and prediction performances.

Table 11. Average results obtained from 30 random computer runs using 200 test points and a 10-fold cross- validation scheme, for the two studied simulation outputs, average survival rate (y_1) and average response time (y_2).

Output	**RMSE**	**MAE**	**RAE**	**RRSE**	**Corr.**
y_1	0.0299	0.0238	0.4941	0.5263	0.8511
y_2	38.8339	30.9991	0.5014	0.5345	0.8459

(a) (b) (c)

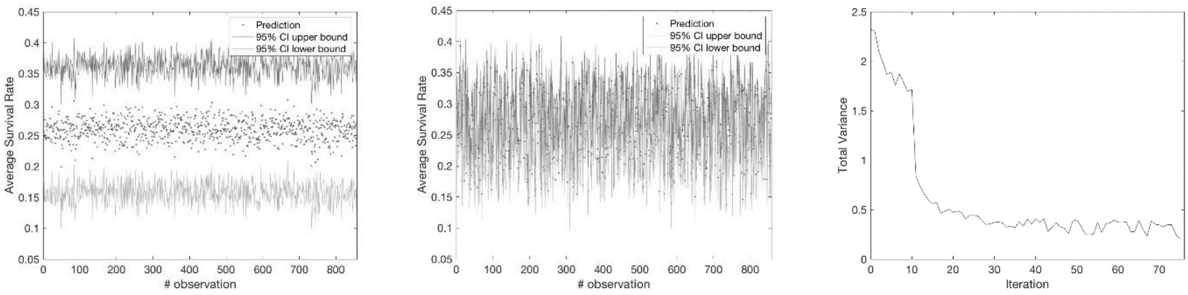


Figure 25. Results for the average survival rate output: (a) depicts the initial GP predictions over 906 simulation points, along with the corresponding 95% bounds, whereas (b) the final GP approximation. 100 random input points were used as the initial training set. Panel (c) shows the required number of iterations to satisfy the stopping criterion with α = 0.1. In each iteration, the top five predictive variance points were selected to be added to the training set.

(a) (b) (c)

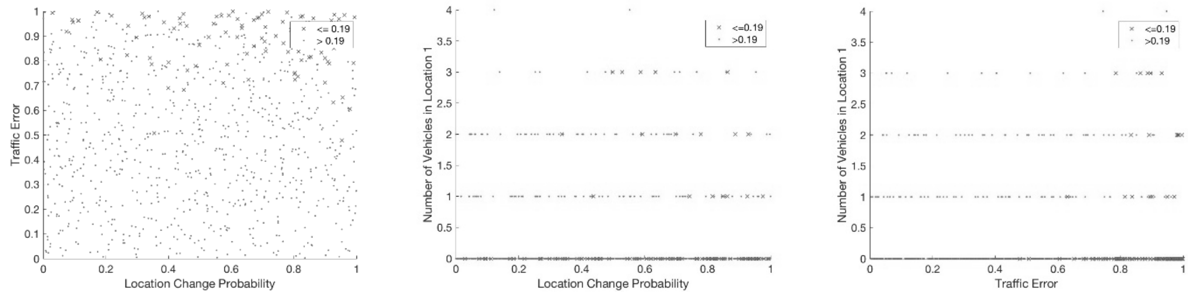


Figure 26. Comparison of the obtained predicted values for different input dimension with an average survival rate threshold of 0.19: (a) Location Change Probability versus Traffic Error, (b) Location Change Probability versus Location 1 and (c) Traffic Error versus Location 1.

(a) (b) (c)

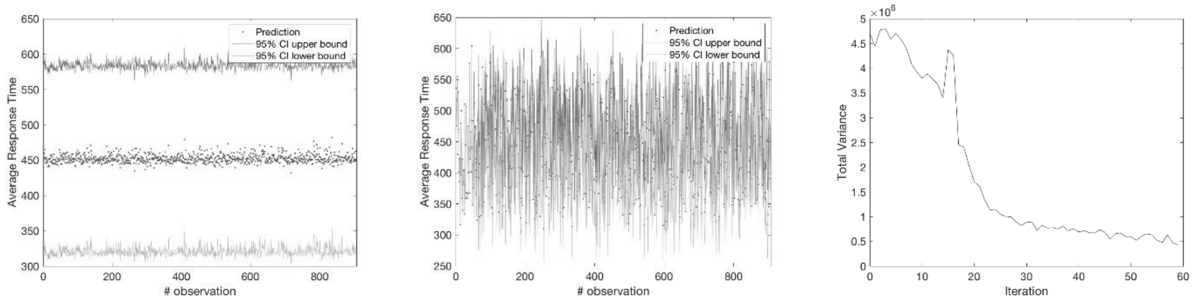


Figure 27. Results for the average survival rate output: (a) depicts the initial GP predictions over 906 simulation points, along with the corresponding 95% bounds, whereas (b) the final GP approximation. 100 random input points were used as the initial training set. Panel (c) shows the required number of iterations to satisfy the stopping criterion with α = 0.1. In each iteration, the top five predictive variance points were selected to be added to the training set.

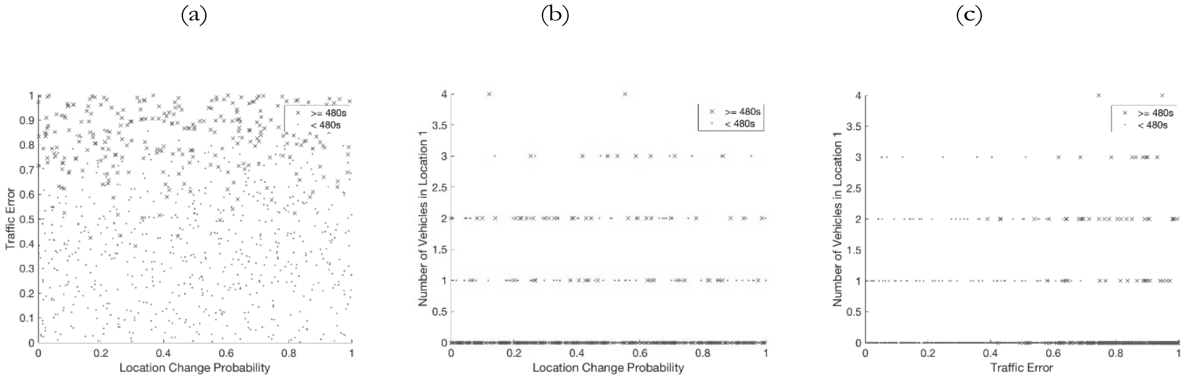


Figure 28. Comparison of the obtained predicted values for different input dimension with an average response time threshold of 480 seconds (8 minutes): (a) Location Change Probability versus Traffic Error, (b) Location Change Probability versus Location 1 and (c) Traffic Error versus Location 1

5.5. CONCLUSION & FUTURE WORK

This paper presents a methodology based on active learning metamodeling to address the problem of exploring the behavior of simulators developed in the context of transportation simulation and policy analysis. Particularly, an Emergency Medical Service (EMS) is used for illustration and two important output thresholds were considered: 0.19 for the average survival rate and 8 minutes (480s) for the average response time.

The presented work provides proof of concept that shows that proposed methodology is able to help in the identification of important regions of the simulation input space that have a direct impact in the performance or implementation of certain policies, while avoiding several simulations runs at the same time. Moreover, if the simulation input space proves to be sufficiently high-dimensional or if each simulation run shows prohibitive computational workload and runtimes, an exhaustive exploration process is virtually impossible. Therefore, the joint use of active learning strategies and simulation metamodels designed to identity policy-relevant regions, has great potential in practice, especially for decision making processes.

The results show that this work can be improved in several research directions. Firstly, the presented methodology can improve in terms of graphical representation. Working with high-dimensional data constitutes a great challenge. The graphical results (policy-relevant regions) should be provided in such a way that they make it easier for policies to be analyzed and conclusions obtained. Moreover, the presented methodology can also be improved by combination of clustering techniques applied the results so that such important input regions are more easily identifiable.

Secondly, this work did not take into account any specific design for computer experiments, such as the well-known the Latin hypercube scheme. These designs provide sampling strategies that lead to a good understanding of the underlying properties of simulation models. This can improve the statistical significance as well as the prediction performance of simulation metamodels.

Thirdly, in order to further validate its potential, this works must be replicated not only using other kinds of transport simulation models, but also considering more complex policy analysis, possibly taking into account combination of policies.

Fourthly, it would be interesting to provide a certain degree of interactivity with the user. For instance, instead of letting the active learning independently decide which data points to request the simulator to label, the user could be able to actively engage in this process too. This could represent an interesting feature for decision making practitioners, as it would directly involve them and their expertise in the modeling process itself.

Lastly, but not least, another important issue is the possible correlation between the simulation outputs, which was not considered in this work, i.e., each output was modeled in an independent way. Multi-output regression, such as the multi-output Gaussian Process framework, is able to improve its prediction performance by considering relationships across outputs. The integration of such models in the proposed methodology would also be an interesting line for future research.

5.6. ACKNOWLEDGMENTS

The authors greatly acknowledge the support of FCT (Portuguese national funding agency for science, research and technology) under the grants PD/BD/128047/2016 and PD/BD/52355/2013 during the development of this work.

5.7. REFERENCES

Amorim, M., S. Ferreira, and A. Couto, Emergency Medical Service Response: Analyzing Vehicle Dispatching Rules. *Transportation Research Record: Jour- nal of the Transportation Research Board*, 2018, p. 0361198118781645.

Ankenman, B., B. L. Nelson, and J. Staum, Stochastic kriging for simulation metamodeling. *Operations Research*, Vol. 58, No. 2, 2010, pp. 371– 382.

Antunes, F., B. Ribeiro, F. Pereira, and R. Gomes, Efficient Transport Simulation With Restricted Batch-Mode Active Learning. Transactions on Intelligent Transportation Systems, Vol. PP, 2018, pp. 1–10.

Bankes, S. C., *Exploratory modeling and the use of simulation for policy analysis*. RAND CORP SANTA MONICA CA, 1992.

Bankes, S., Exploratory modeling for policy analysis. *Operations Research*, Vol. 41, No. 3, 1993, pp. 435–449.

Bastos, L. S. and A. O'Hagan, Diagnostics for Gaussian process emulators. *Tech- nometrics*, Vol. 51, No. 4, 2009, pp. 425–438.

Boukouvalas, A., D. Cornford, and A. Singer, Managing uncertainty in complex stochastic models: Design and emulation of a rabies model, 2009.

Boukouvalas, A., *Emulation of random output simulators*. Ph.D. thesis, Aston University, 2010.

Brochu, E., V. M. Cora, and N. De Freitas, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.

Chen, T., K. Hadinoto, W. Yan, and Y. Ma, Efficient meta-modelling of complex process simulations with time–space-dependent outputs. *Computers & chemical engineering*, Vol. 35, No. 3, 2011, pp. 502–509.

Chen, X., Z. Zhu, X. He, and L. Zhang, Surrogate-based opti- mization for solving a mixed integer network design problem. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2497, No. 2497, 2015, pp. 124–134.

Ciuffo, B., J. Casas, M. Montanino, J. Perarnau, V. Punzo, Gaussian process metamodels for sensitivity analysis of traffic simulation models: Case study of aimsun mesoscopic model, Transportation Research Record: Journal of the Transportation Research Board 2390 (2390) (2013) 87–98 (2013).

Conti, S. and A. O'Hagan, Bayesian emulation of complex multi-output and dynamic computer models. *Journal of Statistical Planning and Inference*, Vol. 140, No. 3, 2010, pp. 640–651.

E. del Castillo, Statistical metamodeling of dynamic network loading. *Transportation Research Procedia*, Vol. 23, 2017, pp. 263–282.

Friedman, L. W. and I. Pressman, The metamodel in simulation analysis: Can it be trusted? *Journal of the Operational Research Society*, Vol. 39, No. 10, 1988, pp. 939–948.

Friedman, L. W., *The simulation metamodel*. Springer Science & Business Media, 2012. [Barton(1998)] Barton, R. R., Simulation metamodels. In *Simulation Conference Proceedings, 1998. Winter*, IEEE, 1998, Vol. 1, pp. 167–174.

Geman, S., E. Bienenstock, and R. Doursat, Neural networks and the bias/variance dilemma. *Neural computation*, Vol. 4, No. 1, 1992, pp. 1–58.

Haghani, A. and S. Yang, *Real-Time Emergency Response Fleet Deployment: Concepts, Systems, Simulation & Case Studies*, Springer US, Boston, MA, pp. 133–162, 2007.

J. Perarnau, and V. Punzo, Gaussian process metamodels for sensitivity analysis of traffic simulation models: Case study of AIMSUN mesoscopic model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2390, No. 2390, 2013, pp. 87–98.

Jagtenberg, C., P. van den Berg, and R. van der Mei, Benchmarking online dispatch algorithms for Emergency Medical Services. *European Journal of Operational Research*, Vol. 258, No. 2, 2017, pp. 715 – 725.

Jones, B. and R. T. Johnson, Design and analysis for the Gaussian process model.*Quality and Reliability Engineering International*, Vol. 25, No. 5, 2009, pp. 515–524.

Kleijnen, J. P. and R. G. Sargent, A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, Vol. 120, No. 1, 2000, pp. 14–29.

Kleijnen, J. P. and W. C. Van Beers, Application-driven sequential designs for simulation experiments: Kriging metamodelling. *Journal of the Operational Research Society*, Vol. 55, No. 8, 2004, pp. 876–883.

Kleijnen, J. P. and W. C. Van Beers, Robustness of Kriging when inter- polating in random simulation with heterogeneous variances: Some experiments. *European Journal of Operational Research*, Vol. 165, No. 3, 2005, pp. 826–834.

Kleijnen, J. P., A comment on Blanning's Metamodel for sensitivity analysis: the regression metamodel in simulation. *Interfaces*, Vol. 5, No. 3, 1975, pp. 21–23.

Kleijnen, J. P., Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, Vol. 192, No. 3, 2009, pp. 707–716.

Kleijnen, J. P., Regression metamodels for generalizing simulation results. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, 1979, pp. 93–96.

Kleijnen, J., Model behaviour: Regression metamodel summarization. *Encyclopedia of Sys- tems and Control*, Vol. 5, 1987, pp. 3024–3030.

Law, A. M., W. D. Kelton, and W. D. Kelton, *Simulation modeling and analysis*, Vol. 2. McGraw-Hill New York, 2007.

Lewis, D. D. and W. A. Gale, A sequential algorithm for training text classifiers. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and Development in Information Retrieval*, Springer-Verlag New York, Inc., 1994, pp. 3–12.

Ling, C. K., K. H. Low, and P. Jaillet, Gaussian Process Plan- ning with Lipschitz Continuous Reward Functions: Towards Unifying Bayesian Optimization, Active Learning, and Beyond. In *AAAI*, 2016, pp. 1860–1866.

Marengo, M. C., Urban Simulation Models: Contributions as Analysis-Methodology in a Project of Urban Renewal. *Current Urban Studies*, Vol. 2, No. 03, 2014, p. 298.

Pons, P. T. and V. J. Markovchick, Eight minutes or less: does the ambulance response time guideline impact trauma patient outcome? *The Journal of Emergency Medicine*, Vol. 23, No. 1, 2002, pp. 43 – 48.

Rasmussen, C. E. and C. Williams, *Gaussian processes for machine learning (Adaptive computation and machine learning)*. The MIT Press, 2005.

Roy, N. and A. McCallum, Toward optimal active learning through monte carlo estimation of error reduction. *ICML, Williamstown*, 2001, pp. 441–448.

Schulz, E., M. Speekenbrink, and A. Krause, A tuto- rial on Gaussian process regression with a focus on exploration-exploitation scenarios. *bioRxiv*, 2017, p. 095190.

Settles, B. and M. Craven, An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the conference on empirical methods in natural language processing*, Association for Computational Linguistics, 2008, pp. 1070–1079.

Settles, B., *Active Learning Literature Survey*. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2010.

Settles, B., M. Craven, and S. Ray, Multiple-instance active learning. In *Advances in neural information processing systems*, 2008, pp. 1289–1296.

Seung, H. S., M. Opper, and H. Sompolinsky, Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, 1992, pp. 287–294.

Song, W., K. Han, Y. Wang, T. Friesz, and

Van Beers, W. C. and J. P. C. Kleijnen, Kriging for interpolation in random simulation. *Journal of the Operational Research Society*, Vol. 54, No. 3, 2003, pp. 255–262.

Wang, L., D. Beeson, S. Akkaram, and G. Wiggs, Gaussian process meta-models for efficient probabilistic design in complex engineering design spaces. *ASME Paper No. DETC2005-85406*, 2005.

Wang, X. and J. Zhai, *Learning from Uncertainty*. CRC Press, 2016.

Yang, S., M. Hamedi, and A. Haghani, Online Dispatching and Routing Model for Emergency Vehicles with Area Coverage Constraints. *Transportation Research Record: Journal of the Transportation Research Board*, 1923, pp. 1–8.

Zhang, L., X. He, C. Xiong, Z. Zhu, et al., Bayesian stochastic Kriging metamodel for active traffic management of corridors. In *IIE Annual Conference. Proceedings*, Institute of Industrial and Systems Engineers (IISE), 2014, p. 1790.

6. An Integrated Approach for Strategic and Tactical Decisions for the Emergency Medical Service: Exploring Optimization and Metamodel-Based Simulation for Vehicles Location[16]

Marco Amorim[17], Francisco Antunes[18], Sara Ferreira[17], Antonio Couto[17]

Abstract

Choosing locations for emergency medical service stations and vehicles has been thoroughly investigated. However, the formulations presented to solve this question are not always done in a way that can be applied in practice, because they are based on oversimplified mathematical functions, which makes them unrealistic. The problem persists as integrated strategic and tactical approaches require an analytical complexity that often invalidates the exact solution.

This work proposes an integrated strategic and tactical planning decision methodology that complements an optimization model with a local search using a metamodel as a proxy of the real system. This allows empirical evidence to be inferred for vehicle location solutions that improve performance in the real system.

The methodology is applied to the city of Porto, and two dilemmas are tested to show proof of application. First, the debate between the integrated versus non-integrated approach is analysed. Second, an assessment of the advantages of adding a vehicle or a station to the planning budget is analysed.

The conclusions of this research support the advantages of an integrated approach indicated by other studies. The results also show that adding a new vehicle to the system is more advantageous than adding a new station when it comes to victims' survival. These two application examples provide proof of the methodology applicability and open doors for future research on the subject.

Keywords: Emergency Medical Service; Ambulance Location Problem; Optimization; Gaussian Process; Simulation Metamodel

---

[16] An Integrated Approach for Strategic and Tactical Decisions for the Emergency Medical Service: Exploring Optimization and Metamodel-Based Simulation for Vehicles Location. Submitted to Computers & Industrial Engineering. Review received and revision to be submitted by June 2019

[17] CITTA, University of Porto – Faculty of Engineering, Porto, Portugal

[18] University of Coimbra – Faculty of Sciences and Technology, Coimbra, Portugal

6.1. INTRODUCTION

6.1.1. MOTIVATION

Stations, distribution centers and other facilities that have a radius of action, and need to meet certain demands within that radius, are often in operation for many years, and therefore subject to physical and temporal changes in the environment in which they operate. Classic facility and vehicle location problems are usually faced with highly uncertain costs, variable demand and ambiguous travel times, as well as other measures that are hard to correlate. Consequently, these types of problems require decision-making tools to deal with the underlying uncertainties and avoid the risk of underestimating or overestimating their design, which in return results in a lower system performance or more expensive solution.

Emergency Medical Service (EMS) strategic and tactical decisions are two of the classic facilities and vehicle location problems. In general, designing an EMS response system from the transportation perspective covers the location of vehicle facilities, allocation of vehicles to facilities and response policies to outline rules that help to decide which vehicle is to be dispatched to a certain medical emergency. The design of the EMS response plan can be divided into two planning stages: strategic and tactical decisions.

At the strategic level, long-term decisions are usually made concerning the location of emergency vehicle stations, and a long-lasting infrastructure is established for the EMS response (Tufuor et al., 2018, He et al., 2018, Amorim et al., 2017). These decisions result in building or renting warehouse structures, which should last for many years to come and must accommodate the required resources, e.g. response vehicles. Tactical decisions, in contrast, define mid or short-term decisions (Amorim et al., 2018, van Essen et al., 2013), such as the allocation of response vehicles to EMS stations. Hence, there is no permanent commitment to the chosen solution.

These two planning stages have been studied in depth, especially focusing on the mathematical problem – the algorithm itself and problem solving techniques, mainly heuristics. Nevertheless, mathematical models are an abstraction of the reality and often require simplifications that lead to one important question: what is being improved? What looks optimal on paper might not correctly translate into practice. Recently, new research directions have pointed towards practice ready decision tools, such as simulation or scenario-based optimization. These techniques are able to integrate the system characteristics successfully and can elaborate and solve more complex problems having a practical focus.

The present study follows the aforementioned new direction and studies a methodology for integrated strategic and tactical planning decisions. As stated before, strategic decisions have a permanent commitment and a wrong decision can affect the performance of the forthcoming tactical decisions. We believe that separating these two decisions can lead to sub-optimal solutions. This fact has been shown by van Essen et al. (2013). However, they point out that a combined approach leads to an increase in the problem size and extra simplifications might be necessary so that a solution can be found.

A two-stage methodology is proposed for solution exploration with empirical evidence. In the first stage, a scenario-based optimization model provides integrated strategic and tactical planning with the necessary simplifications to deliver a solution within a reasonable time. The second stage focuses on exploring the neighbourhood of the optimized solution for alternative solutions that might perform better using empirical evidence. A simulation model is an alternative for empirical proof inference because experimenting in the real system is prohibitive. However, simulation models require excessive computing resources and are time-consuming. Evaluating a unique solution might take several minutes or even hours. For this reason, a metamodel-based simulation is preferred because it allows a much quicker evaluation of alternative solutions with a minor accuracy loss.

This work contributes to the state of the art by:

- providing a methodology to integrate strategic and tactical EMS planning decisions with a practical focus,
- supporting the claim that separate strategic and tactical decisions lead to sub-optimal solutions,
- using a metamodel-based simulation to complement mathematical optimisation derived solutions with empirical evidence and applications to a case study.

6.1.2. EMS STATION AND VEHICLE LOCATION

Emergency medical service location problems are one of the classic problems in operational research (OR) and mathematical programming. The two most important studies in the field of EMS and station location are Toregas et al. (1971) and Church and Velle (1974). The former one presents a solution to solve the location set covering problem (LSCP) making sure all demands are covered within a maximum time or distance radius. While the latter focuses on maximizing the coverage.

Nonetheless, full coverage is hard to reach especially when resources are limited, which is the case of practical applications. Li et al. (2011) reviewed covering models for EMS and highlighted many models that relax some of the assumptions made by Toregas et al. (1971).

The follow up model formulations tried to take into consideration the problem of facility or vehicle availability. From the hierarchical approach proposed by Daskin and Stern (1981), which firstly integrated multiple coverage standards, Gendreau et al. (1997) formulated the famous Double Standard Model (DSM). This model has been widely used as a solution to account for vehicle or facility unavailability, ensuring that an alternative facility or vehicle is available within a second standard distance. ReVelle and Hogan (1989) tackled the problem with a different approach and formulated a probabilistic version of the LSCP with the requirement that all the demand points must be covered with a reliability level α.

Another important point to highlight is that demand and travel times are not evenly distributed temporally and spatially, and thus the busy fraction of a vehicle varies from facility to facility. This problem was investigated and a maximal expected coverage location model with time variation (TIMEXCLP), where varying demands are incorporated over time, is proposed by Repede and Bernardo (1994).

Once more, the review carried out by Li et al. (2011) indicates several other extensions of the maximal expected coverage location model (MEXCLP). Fujiwara et al. (1987) and (1988) applied simulations to make a further analysis on the optimality of an EMS location problem in Bangkok using MEXCLP. While most recently, Schmid and Doerner (2010) developed a multi-period version of the DSM which takes into account time-varying coverage to optimize coverage at various points in time simultaneously. The work concluded that it is vital for EMS location problems to consider time-dependent variations in travel times and coverage respectively. As a follow up, Dibene et al. (2017) uses a similar approach in the city of Tijuana and concludes that demand coverage and response times in Tijuana can be enhanced by relocating the current stations without needing additional resources.

However, the above-mentioned authors have focused their models on operational performance metrics such as response time or coverage, these are the common metrics used both in practice and research (Bélanger et al., 2016, Ünlüyurt and Tunçer, 2016). One of the greatest impacts of planning EMS is the medical response time and how it can change patients' survival. Accordingly, Erkut et al. (2008) developed the Maximum Survival Location Problem (MSLP) which incorporates a survival function into the covering model to maximize patients' survival.

A second issue of the former models is the fact that using analytic formulations is still oversimplifying the real system. Notwithstanding, using robust solutions such as the scenario-based approach, which already captures part of the stochasticity existing in daily emergency medical services, is still a high simplification of the reality. Simulation models come as a response to this problem because they attempt to provide a tool that can deliver an empirical platform that translates the real system.

Using simulation models allow researchers to formulate more realistic and complex problems, usually to assess solutions or to support optimization models (Restrepo et al., 2008, Maxwell et al., 2010, Yue et al., 2012, McCormack and Coates, 2015, Iannoni et al., 2009, Su and Shih, 2003, Bélanger et al., 2016, Aboueljinane et al., 2013, Ünlüyurt and Tunçer, 2016). McCormack and Coates (2015) investigate how simulation can enhance the level of realism in EMS models, making it applicable to complex real-life systems, if proper data exist. Yet, a more detailed and complex model comes with the cost of higher computing power and time. For a deeper review on simulation applied to EMS problems the reader is forwarded to the review made by Aboueljinane et al. (2013).

One can assess a short set of solutions by using simulations. However, when it comes to a local search, where thousands or millions of alternative solutions might exist, using simulations becomes infeasible. When the search for the solution requires extensive experimentation, simulating each instance becomes time consuming, thus simpler estimations are preferred; these are models of the model, which have been introduced as metamodels by Blanning (1974) and further statistically developed by Kleijnen (1975). In the transport research field, metamodels have only recently been applied, e.g. in traffic predictions (Antunes et al., 2018, Song et al., 2017), network optimizations (Song et al., 2017) or in Demand-Responsive Transportation (Antunes et al., 2018). Furthermore, Barton and Meckesheimer (2006) discuss and show the usability of metamodels in optimization problems to explore local solutions that can better fit the real system characteristics.

The idea of complementing optimization with a metamodel to evaluate local solutions is shown to be valid, and as far as the authors know, there are no relevant studies in this regard within the scope of EMS stations and vehicle locations.

6.2. METHODOLOGY

6.2.1. FRAMEWORK

A two-stage methodology is proposed to integrate EMS response strategic and tactical decisions in a unique plan design. In the first stage, an optimization model looks for a conceptual solution that performs optimally on paper, and in the second stage, a metamodel-based simulation is used to explore alternative solutions for empirical evidence and to chose the best one.

The core of the methodology is a multi-period scenario-based optimization model that mathematically finds a solution for the EMS location problem. Afterwards, a metamodel trained to use a Gaussian Process is applied to assess alternative local solutions that are generated by a combinatorial algorithm. This process leads to possible alternatives that will empirically perform better in the real system.

The two fundamental stages are achieved using three tools: an optimization model; a local search supported by a metamodel-based simulation; and a simulation model to train the metamodel. A diagram of these stages and the necessary modelling processes are systematized in Figure 29.

The optimization model locates medical emergency vehicles and stations while maximizing patients' survival. This model considers the station's "busy-fraction" to define the number of vehicles, it uses "discrete time" to define scenarios, and has a set of parameters that allow for tailored solution pending operational constraints. The main goal is to provide an optimized station and vehicle locations.

A Gaussian Process based metamodel is trained using data points obtained from a simulation model to serve as empirical evidence of the solution performance within a stochastic environment that mimics the real system. A metamodel-based simulation is preferred because, with a minor loss in accuracy, it overcomes the existing simulation models because they use many resources (computational and human) and time, which would lead to a drastic reduction in the number of solutions that could be tested. The only time-consuming tasks are the metamodel parameter calibration and generation, via simulation, of the associated training.

The next sections will further detail the optimization model, the simulation model to train the metamodel, and the metamodel formulation with the respective algorithm for the local search.

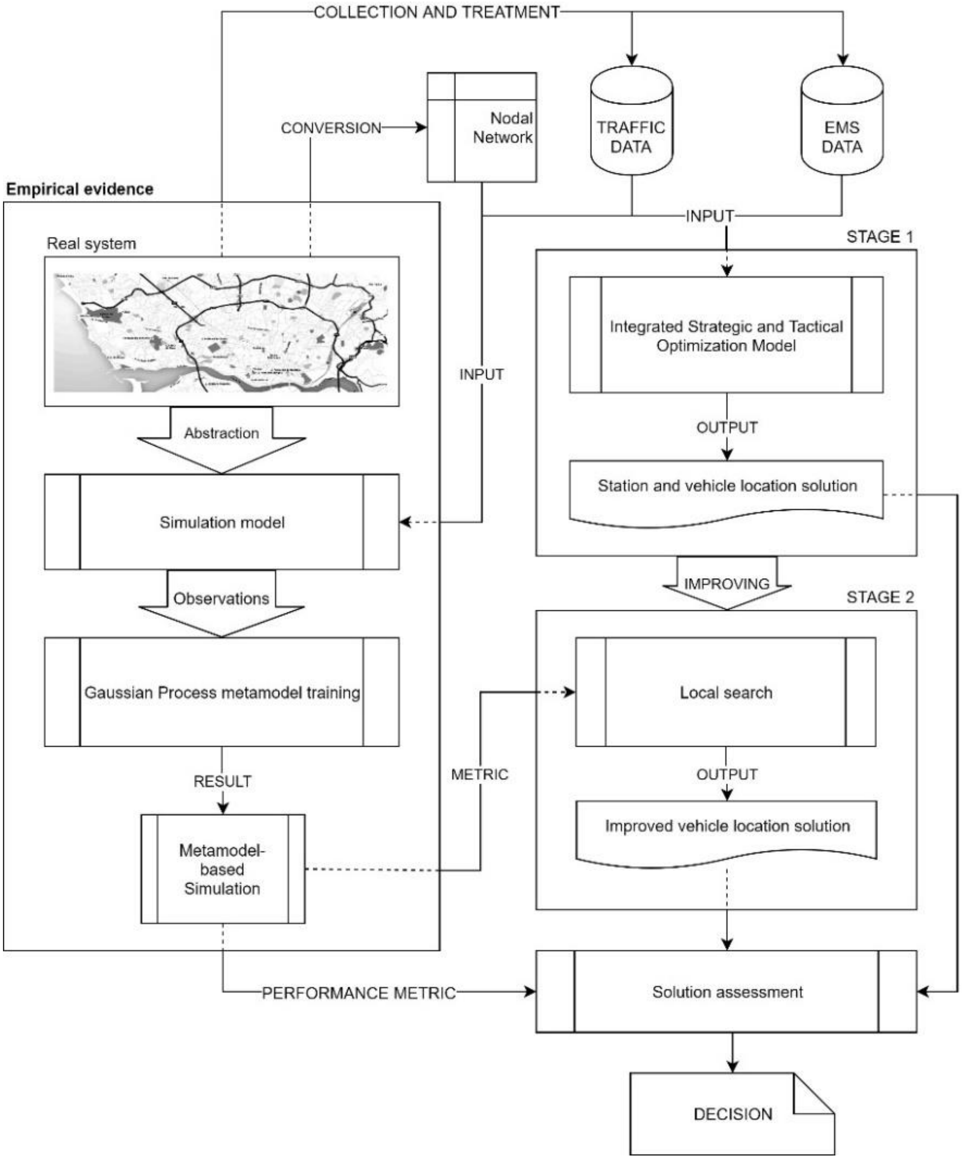


Figure 29. Diagram of the framework.

6.2.2. STRATEGIC AND TACTICAL DECISIONS – INTEGRATED OPTIMIZATION MODEL

The objective of the Strategic and Tactical Integrated Model (STIM) is to provide a tool to combine the strategic and tactical planning of emergency vehicle and station locations and offer the first analytical solution.

The first important step in the solution search is to define the performance metric to be optimized. As was previously discussed, there is a wide range of performance metrics that can be integrated into the objective function. These can focus on the operational performance or on the victims' outcomes. This work follows a survival approach, firstly introduced by Erkut et al. (2008), and afterwards adopted by many other researchers (McCormack and Coates, 2015, Bandara et al., 2014, Mayorga et al., 2013, Knight et al., 2012, Amorim et al., 2018). This approach uses a survival function that, when combined with victims' heterogeneity, takes the form of equation (6.1):

$$S_e = [1 + \exp(C^k + m^k \times t_e)]^{-1} \tag{6.1}$$

Moreover, the system performance, P^s, is calculated through equation (6.2):

$$P^s = \sum_e S_e \tag{6.2}$$

Where,

e is an index of a set of events E

k is the index of a type of event from the set of events type K, e.g. cardiac arrest, road injury.

t is the response time to event e, and

C^k and m^k are the survival function parameters for medical emergency type k.

To capture traffic and demand spatial and temporal heterogeneity, a multi-period scenario-based optimization is preferred, as seen in Dibene et al. (2017) and also explored by Amorim et al. (2016). The proposed formulation derives directly from the strategic planning model for emergency vehicle station locations formulated by Amorim et al. (2016) and is adapted to vehicle locations using the notion of station busy fraction; firstly introduced by Daskin (1983) and adapted in the works of Snyder and Daskin (2005) and Berg et al. (2016).

The formulation of the model uses $s \in S = [1, 2, \ldots, s]$ scenarios that discretize the traffic and demand continuous changes. Adopting performance metric P^s as the objective to maximize and B as the maximum busy fraction of the first order of a station, the new scenario-based strategic and tactical decisions integrated optimization model (STIM) is formulated as follows:

$$\text{maximize} \sum_s \sum_e \sum_l \sum_p y_{s,l,p} \times d_{s,p} \times [1 + \exp(C^k + m^k \times r_{s,l,p})]^{-1} \tag{6.3}$$

Subjected to:

$$\sum_l y_{s,l,p} = 1, \quad \forall\ [s,*,p] \in A \tag{6.4}$$

$$y_{s,l,p} \le x_l, \quad \forall\ [s,l,p] \in A \tag{6.5}$$

$$\sum_l x_l \le M_l, \quad l \in L \tag{6.6}$$

$$\sum_p y_{s,l,p} \times d_{s,p} \times N_s \le B \times \frac{z_l}{b}, \quad \forall\ [s,l,*] \in A \tag{6.7}$$

$$\sum_l z_l \le M_v, \quad \forall\ [*,l,*] \in A \tag{6.8}$$

$$z_l \le M_m, \quad \forall\ [*,l,*] \in A \tag{6.9}$$

$$x_l \in \{0,1\}, \quad \forall\ [*,l,*] \in A \tag{6.10}$$

$$y_{s,l,p} \in \{0,1\}, \quad \forall\ [\mathrm{s},l,p] \in A \tag{6.11}$$

$$z_l \in \{0,1,2,...,M_m\}, \quad \forall\ [*,l,*] \in A \tag{6.12}$$

$$a \in A = \left\{[s,l,p] \mid r_{s,l,p} \le M_r\right\}, \quad \forall\, s \in S,\ l \in L,\ p \in P \tag{6.13}$$

Where the sets are:

$s \in S$ is the set of scenarios for the multi-period approach

$l \in L$ is the set of possible station locations

$p \in P$ is the set of demand points

$a \in A$ is the availability set which consists of a set of tuples [s,l,p] that satisfy maximum response time constraint, e.g. $r_{s,l,p} \le M_r$. Set A can reduce the number of decision variables and constraints, thus reducing the problem size.

The decision variables are:

$y_{s,l,p} = 1$ if a vehicle from station l serves node p during scenario period s, 0 otherwise,

$x_l = 1$ if a station is located at possible station location l, 0 otherwise,

z_l is the number of vehicles deployed at station l.

Furthermore, the parameters are:

$d_{s,p}$ is the estimated demand during period s at demand node p,

$r_{s,l,p}$ is the travel time from station location l to demand node p during scenario period s,

M_l is the maximum number of open stations,

N_s is the demand peak factor during period s,

B is a parameter that translates station busy fraction,

b is the average time a vehicle is busy when responding to an emergency (a sum of the response time, assistance time, travel to the hospital, drop off the victim at the emergency service, return to its station and be ready for the next call),

M_v is the maximum number of vehicles per station,

M_m is the maximum number of emergency vehicles, and

M_r is the maximum response time allowed.

Equation (6.3) maximizes the overall survival considering the demand at each node and at each period. Equations (6.10), (6.11) and (6.12) represent the decision variables and their domain, while equation (6.13) represents the availability set, whose use drastically reduces the size of the problem by considering only the pairs [period scenario, station location, demand node] that respect the constraint $r_{s,l,p} \leq M_r$.

For logistic purposes, constraints represented by equations (6.6), (6.8) and (6.9) limit the total number of open stations, the total number of deployed vehicles and the maximum number of vehicles at one station respectively. Equation (6.4) ensures that for each scenario period only one station is allocated to each demand point and equation (6.5) ensures that there is a vehicle deployed at that station.

Finally, equation (6.7) is the contribution to the station location formulation proposed by Amorim et al. (2016) and it adopts the concept of busy fraction to control the number of vehicles deployed at each station to limit the busy fraction B of the service stations during a scenario period demand peak characterized by N_s. It allows the user to control the station or vehicle busy fraction and to have solutions tailored for specific failure rates, i.e. a solution that will suffice for a certain fraction of the analysed period.

6.2.3. Simulation model

A simulation model is retrieved from Amorim et al. (2018) to generate data points for training the metamodel that will be used for alternative solution exploration. The adopted simulation model is an agent-based simulation that can be quickly adapted for the objective of this work.

The model translates each EMS stakeholder into independent agents that interact within a stochastic environment. The main agent representing the city acts as the EMS provider by answering emergency calls and assigning a response vehicle to them. Each medical emergency is formulated as an agent with certain characteristics that define their type, location, and timestamp. Each vehicle is also an agent that answers to the city agent and has autonomy to travel to the city, assist and pick up patients and transport them to the desired hospital.

The environment abstracts the complexity of the road network by considering a nodal network. Nodes are represented by a coordinate system and each pair of nodes has a directional multi-period travel time associated. This means that travelling from A to B is not the same as travelling from B to A, and travelling from A to B during period s is not the same as travelling from A to B during period s'.

The advantage of the simulation is to have a stochastic representation that resembles the real system it tries to mimic. To meet the objective of this study, two uncertainties are studied; traffic behaviour and how it affects travel times, and the EMS demand and how it fluctuates during the day. Two parameters are implemented to control these two behaviours. The first parameter ε_T captures the traffic stochasticity by introducing an "*error term*", when dispatching a vehicle, in the observed travel time $r_{s,l,p}$, (6.14):

$$r'_{s,l,p} = r_{s,l,p} + r_{s,l,p} \times f\left[N\left(\mu, \frac{\varepsilon^2}{4}\right)\right] \tag{6.14}$$

Where $f(N(\mu,\sigma^2))$ is a function that returns a random value from a normal distribution with mean μ and standard deviation σ. The mean is set to zero and the standard deviation is set to $\varepsilon/2$. This means that there is a confidence of 95% that the estimated travel time has at most an error of $\varepsilon \in [0, 1]$.

The second parameter, w_1, captures the possible randomness in the demand, by reallocating the medical emergency event locations, e_p, through the network with a probability of w_1, equation (6.15):

$$e'_p = p\left(g\left(E_s\right), e_p, w_1, w_2\right) \tag{6.15}$$

Where $g(E_s)$ is a function that randomly picks a location from a bag of possible locations weighted according to period s, and $p(e_1, e_2, w_1, w_2)$ is a function that picks either e_1 or e_2 with a respective probability of w_1 and w_2 where $w_1 + w_2 = 1$.

These two parameters, ε_T and w_1, allow a controlled integration of the randomness that is observed in a real environment by manipulating the analytical data fed to the simulation model. To control the simulation output, two hyperparameters are also implemented. The number of runs, n^r, which controls the number of times the same input and parameters are evaluated by the simulation model; and the number of simulated days, n^d, which controls the number of days the model will simulate to produce each output – solution performance. The simulation output uses equation 2 to calculate the average system performance during all runs n^r, and also returns the average response time for all the calls served.

6.2.4. METAMODEL-BASED SIMULATION AND LOCAL SEARCH

The time-consuming problem of a simulation model to produce relevant outputs invalidates the possibility to test big batches of inputs and different combinations of parameters. A metamodel can overcome this disadvantage and can be used to quickly explore alternative solutions during strategic and tactical EMS decision making.

A metamodel-based simulation using Gaussian Processes is proposed. By definition, and according to Rasmussen (2004), a Gaussian Process (GP) is a stochastic process in which any finite number of random variables forms a multivariate Gaussian distribution. Each GP is solely defined by a mean and a covariance (or kernel) function, respectively, $m_f(x)$ and $k_f(x,x')$, where x and x' are two different input observations. The GP framework is a well-known and widely applied machine learning tool in a variety of research areas.

In terms of notation, a GP is commonly denoted by $GP(m_f(x), k_f(x,x'))$. Within a standard regression problem $y = f(x)$, where e $\varepsilon \sim N(0, \sigma^2)$, the GP places a prior over $f(x)$, equation (6.16):

$$f\left(x\right) \sim GP(m_f\left(x\right),\ k_f(x,x')). \tag{6.16}$$

It is usual for the mean and covariance functions to have a certain number of free parameters. These parameters, also called hyperparameter of the GP, are estimated in order to maximize the marginal likelihood. After this fitting procedure, the conditional distribution for an unobserved data point x_* is given by equation (6.17):

$$f_* \mid X, y, x_* \sim N\left(k_f\left(X, x_*\right)^T \left[k_y\right]^{-1} y, k_f\left(x_*, x_*\right) - k_f\left(X, x_*\right)^T \left[k_y\right]^{-1} k_f\left(X, x_*\right)\right) \tag{6.17}$$

where K_y is the covariance matrix and X is the design matrix. Therefore, the GP makes a prediction in the form of Gaussian distributions.

The local search is then accomplished through an algorithm that generates all possible combinations C_k^n where n is the possible vehicle locations (the maximum number of vehicles per each available station) and k is the total number of vehicles. The exploration space is generated using *Python itertools library* and tested using the aforementioned metamodel-based simulation and the solution with the maximum survival output is chosen.

6.3. CALCULATIONS

6.3.1. CASE STUDY APPLICATION

The proposed methodology and claims of this work are validated using real data from the city of Porto, Figure 30.

The data was collected from the INEM (National Medical Emergency Institute) database and consists of 35,000 emergency medical calls, originated in Porto, between 2012 and 2013, Figure 29.

The data was processed using Google maps geocoding API to translate addresses into coordinates and stored in a SQL database. The national census database sub-section division was used to build a 90-node network. Using a clustering algorithm with a Euclidian distance metric, the medical emergency calls were allocated to the network nodes. From the 90 network nodes, 15 were chosen as possible station locations.

Five scenarios are set corresponding to different periods of the day and days of the week. The working days were separated from weekend days. For the working days, the day was divided into three periods: from 6:00 am to 2:00 pm; from 2:00pm to 10:00pm; and from 10:00pm to 6:00am. The weekend days were divided into two periods: from 6:00am to 10:00pm; and from 10:00pm to 6:00am. These periods were defined after carefully studying the traffic and demand temporal and spatial fluctuations.

Part of the optimization model parameters were calculated according to the resulting database, which are the demand parameters $d_{s,p}$, the travel time matrix with elements $r_{s,l,p}$, and the demand peak factors N_s.

The maximum number of stations, maximum number of emergency vehicles and maximum number of vehicles per station values were set to $M_l = 7$, $M_m = 14$, $N_s = 3$, respectively. From the authors experience, these are values that allow for a service with no significant delays, i.e. there is always an available vehicle to respond to a call. To reduce the problem size, the maximum response time was set to $M_r = 12$min.

Finally, the average time a vehicle is busy was estimated as $b = 120$min and the busy fraction parameter was set to $B = 1$. This is the limit of feasibility since it tolerates a 100% rate of vehicle usage during the peak period characterized by N_s.

With the optimization model solution, the local search algorithm is applied to find an empirically better solution.

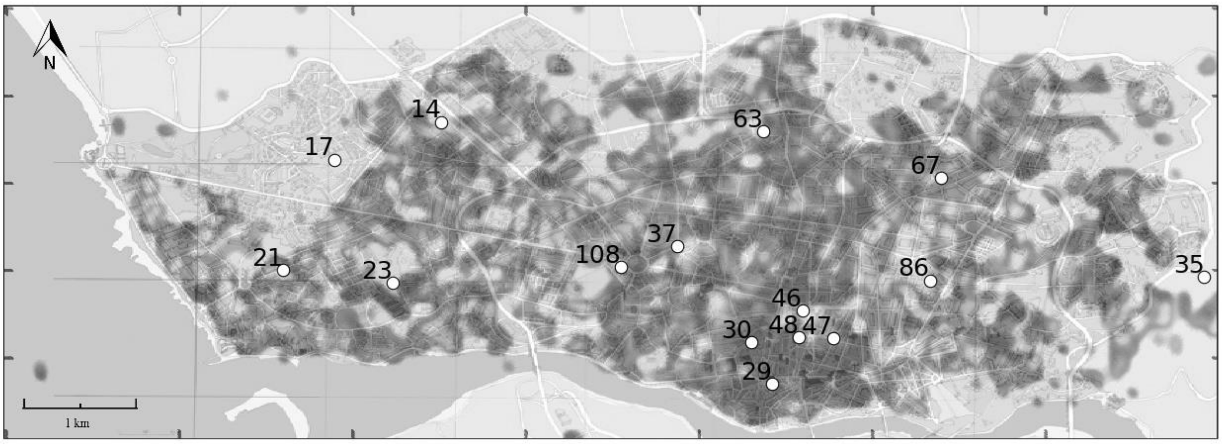


Figure 30. City of Porto and possible station locations with corresponding label. Darker areas represent a higher number of emergency calls.

6.3.2. METAMODEL-BASED SIMULATION TRAINING

To train the metamodel, a batch of 800 simulation runs is used as observations. These observations consist of an input vector with the position of emergency medical vehicles and the stochastic parameters ε and w_1, and the respective average survival output.

The set of input vectors was generated by constraining the number of stations between 7 and 10, the total number of vehicles between 11 and 15, and the maximum number of vehicles per station to 3. From the authors' experience, after several optimization models tested in previous studies, 15 possible station locations were considered to generate the simulation input.

An algorithm randomizes a feasible solution according to the above restrictions, then randomly picks a value for the parameters ε and w_1, and inputs it in the simulation model with hyperparameters $n^r = 20$ and $n^d = 15$. The input vector and the outputs are stored in a database to be used to train the Gaussian Process metamodel parameters.

The GP was trained, for this study, with the mean function set to be a constant function dependent on the target variable average of the training data set. On the other hand, the well-known Squared Exponential, defined by equation (6.18) was chosen to be the covariance function of the GP:

$$k_f\left(x, x'\right) = \sigma^2 \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right) \tag{6.18}$$

Where, σ^2 and l^2 are the signal variance and the length-scale, respectively.

The GP was trained using a data set comprised of 1,098 random simulation results. Each observation has 17 dimensions in total, namely, the location change probability w_1, traffic error ε_T, and 15 locations. The training process was applied for both simulation outputs: average response and average survival rate. The hyperparameters were obtained using marginal likelihood optimization over this training data set, and the predictive power of the metamodel is shown in Figure 31. Validation of the GP metamodel generalization capacity was conducted using a 10-fold cross-validation scheme.

The metamodel achieved a linear correlation of 0.912 and 0.898 for the average response and average survival outputs, respectively. In terms of the root-mean-square error (RMSE), the training process achieved values of 16.05 for the average response time and 0.018 for the average survival. These results show a high predictive and generalization performance of the trained model.

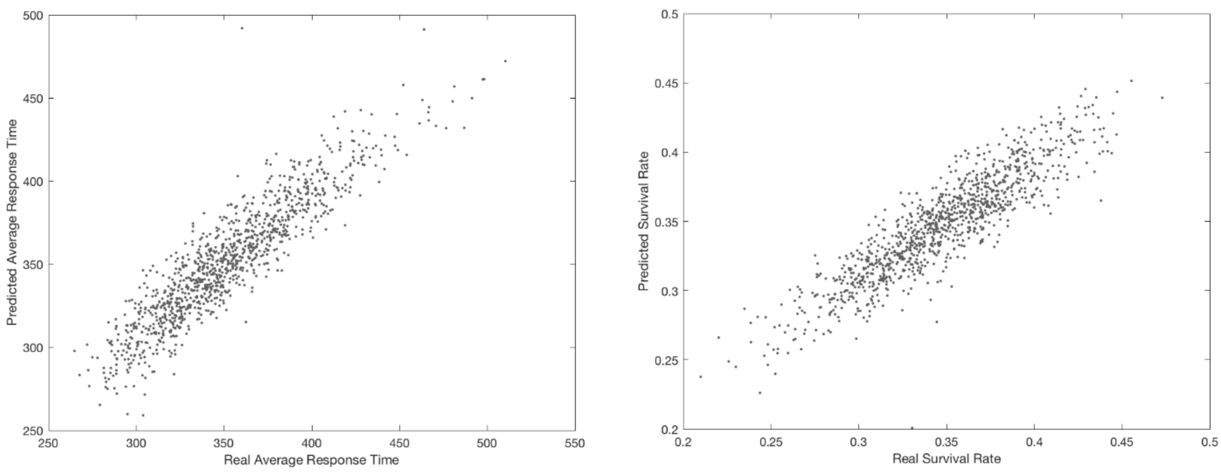


Figure 31. Predicted values using the metamodel versus real values for the survival and average response time outputs.

6.4. RESULTS AND DISCUSSION

6.4.1. STRATEGIC AND TACTICAL PLANNING: INTEGRATED VS NON-INTEGRATED

A discussion on planning strategic and tactical decisions as a combined problem is led in van Essen et al. (2013). To contribute to this question, the methodology of this study is applied to define an integrated solution and afterwards to compare it with a non-integrated solution.

First, the solution achieved solely through analytical optimization is analysed. Using the model proposed by Amorim et al. (2016), a station location solution was obtained. Afterwards this solution was used to restrict the station location in this paper's optimization model and a solution for the vehicle location is calculated. This two-phased optimization process results in the non-integrated solution. For the integrated solution, the STIM optimization model was applied allowing for a solution that defines station and vehicle locations simultaneously.

Afterwards, the local search algorithm was applied to each of the initial solutions resulting in two new solutions (with local search). The metamodel was used for assessing the various solution performance in terms of survival. For reference, and as a collateral consequence of station and vehicle locations, the average response time of each solution is calculated.

The solutions are represented in Figure 32 and the performance result for different values of the stochastic coefficient (with $\varepsilon_T = w_1$) are plotted in Figure 33.

It is obvious that the integrated solution outperforms the non-integrated solution. Nevertheless, higher survival comes at the cost of higher average response times. It is expected that in order to better respond to life-threatening events, emergency vehicles must be allocated closer to the nodes where these types of events dominate. This shows that demand is not homogeneous and that the use of survival functions as a performance metric allows for a system that better serves the victims' outcomes, i.e. better average response times not always lead to higher survival rates.

The local search allows for solutions with better empirical performance as expected. When the stochastic coefficients increase, an obvious decrease in performance is visible although this is not always true when it comes to the response time. It is worth noting that the integrated solution has a lower performance variation when the stochastic coefficients increase. This might indicate that this solution is more robust, thus adapting better in uncertain environments.

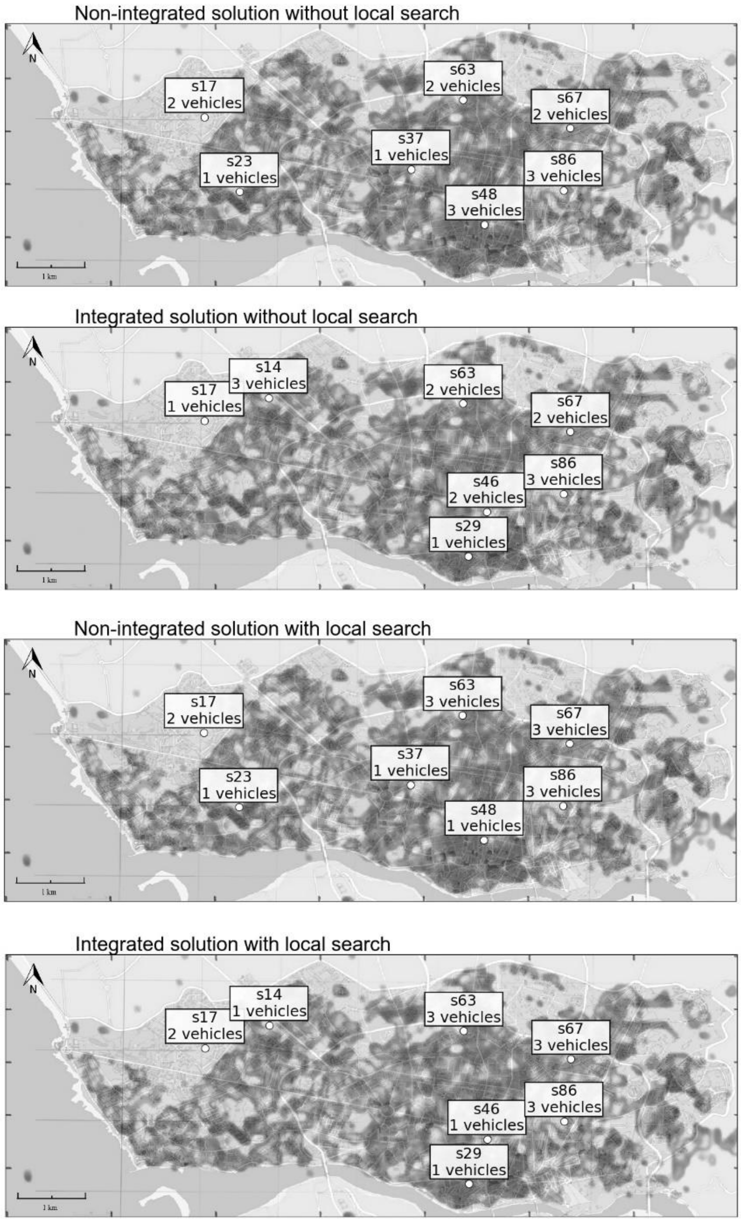


Figure 32. Visualization of the non-integrated and integrated solutions with and without local search.

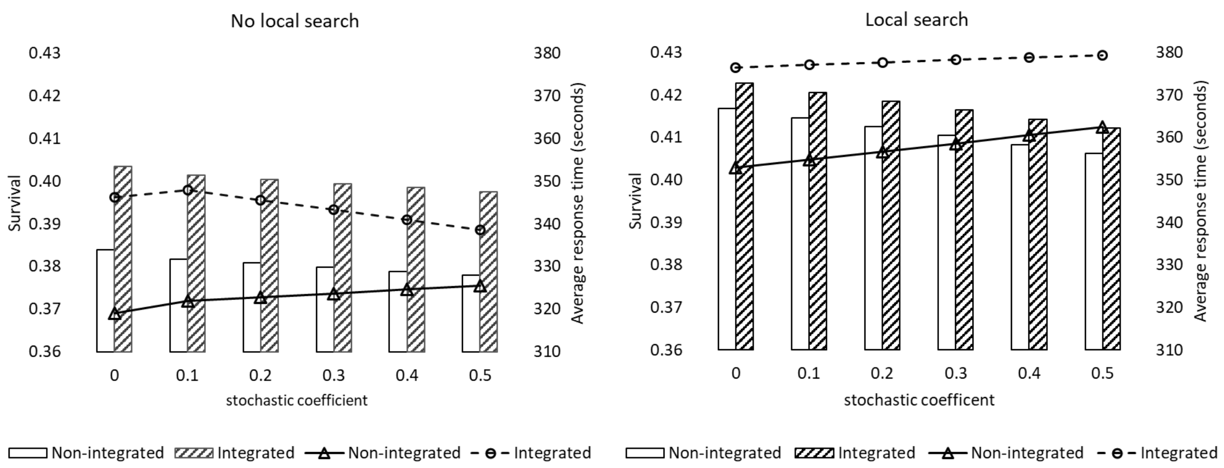


Figure 33. Performance of the integrated and non-integrated solutions with and without local search.

6.4.2. IMPROVING SOLUTION AND SCENARIO ANALYSIS

This section briefly explores the capabilities of the methodology. The dilemma of adding more vehicles at the cost of reducing the number of stations, and the inverse, is analysed.

A set of five solutions are considered:

- IS-7s-14v - Integrated solution with 7 stations and 14 vehicles
- IS-6s-14v - Integrated solution with 6 stations and 14 vehicles
- IS-6s-15v - Integrated solution with 6 stations and 15 vehicles
- IS-8s-14v - Integrated solution with 8 stations and 14 vehicles
- IS-8s-13v - Integrated solution with 8 stations and 13 vehicles

The same previous parameters are used, and the full methodology is applied, i.e. optimization solution followed by a local search followed by a performance assessment. However, for the solution IS-8s-13v the busy fraction parameter *B* had to be raised to 1.1 so that the optimization model could find a feasible solution. The results are compiled in Figure 34.

From the tested solutions, it can be concluded that having seven stations is more advantageous than one less or one more station. When there is an additional vehicle available in the solution with six stations, the solution with seven stations is preferable. Curiously, adding a new station does not lead to higher performances.

In terms of the average response time performance, most of the solutions seem to have smaller variances as the stochastic coefficients increase. Nevertheless, it is interesting to note that the solutions with eight stations have less stability. Having more stations, it is most likely that when dispatching a vehicle, an error in travel time measures leads to the selection of a non-optimal vehicle. The dispersion of stations makes the solution more susceptive to the traffic error.

If *vehicle cost < station cost*, then in terms of economical preference *IS-6s-14v > IS-6s-15v > IS-7s-14v > IS-8s-13v > IS-8s-14v*, it can be concluded that the solutions with six and seven stations are preferable.

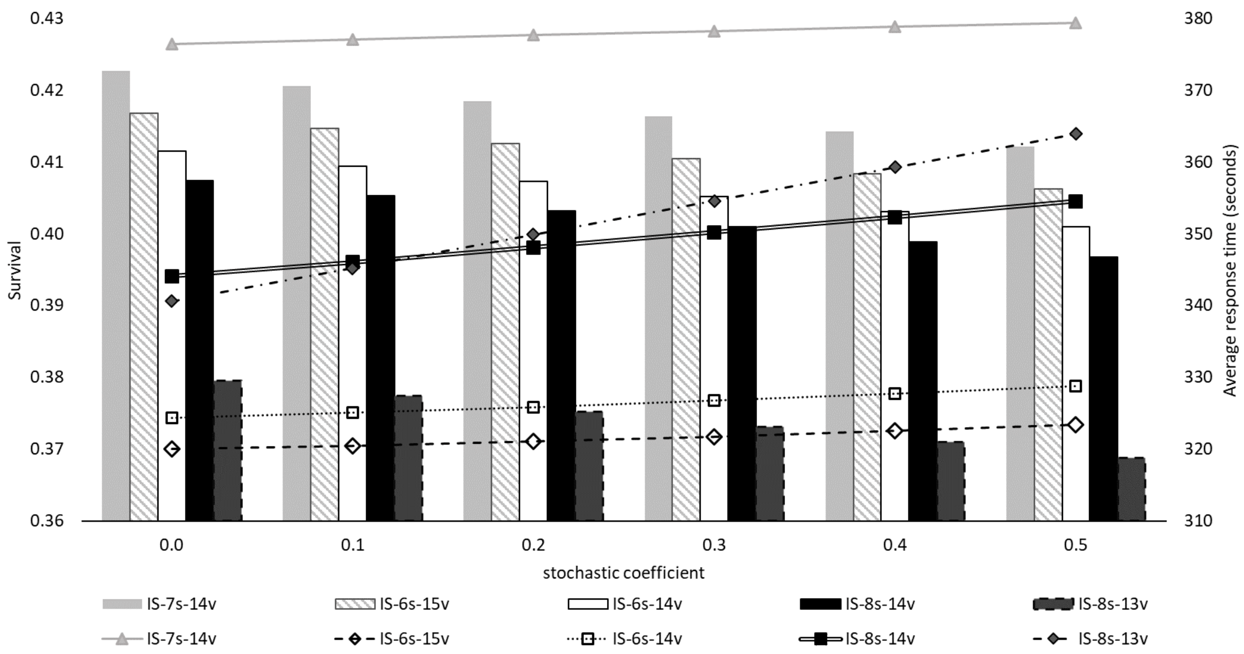


Figure 34. Performance of the different tested solutions. The darker bars represent the most expensive solutions.

6.5. CONCLUSIONS

The study of strategic and tactical planning for station and vehicle locations in EMS has attracted much attention in recent years. With the availability of big data sets and high computer resources, the classical approach for EMS planning becomes obsolete as analytical formulations fail to capture important components of such systems, e.g. traffic changes, demand variability.

This study proposed a framework to tackle strategic and tactical decisions in a way that allows the use of stochastic variables and incorporates the complexity of an EMS system and the urban area it serves.

An optimization model was proposed that can be solved within a reasonable time, and produces solutions adapted to the continuous demand and traffic changes during the day. To complement the optimization model, a metamodel-based simulation was also proposed. It allows the evaluation of solutions based on a simulation model that mimics a real system and the environment it is implemented in. These two tools allow for a chained methodology to plan EMS decisions with robustness and flexibility, ensuring decision makers of the expected performance.

The methodology applied in the city of Porto contributed to the discussion between integrated and non-integrated strategic and tactical decision planning. It was shown, through the metamodel, that the solution performance of a non-integrated approach leads to lower system performance when compared to an integrated solution produced by the proposed methodology.

Furthermore, the methodology was also applied to a station versus vehicle dilemma and was shown to be a practical tool to support decisions or to explore alternatives in uncertain scenarios. The analytical formulation of the metamodel presents a simple yet robust performance function that can be easily used by decision makers or by designers to quickly prototype system improvements. The results show that, in practice, an increase in vehicles is more preferable than an increase in stations.

As both studies illustrate, better average response times not always lead to a higher survival. This is an important finding supporting survival metrics over classical operational ones.

For future research, the study of metaheuristics or mathematical formulation is suggested to allow a direct use of the metamodel function as the objective function of an optimization model for EMS. Implementing robustness measures, such as the variance obtained during the simulation runs, is also worth investigating as this will provide more robust metrics when exploring or evaluating solutions.

6.6. ACKNOWLEDGEMENTS

The authors acknowledge the support of the FCT (Portuguese National Funding Agency for Science, Research and Technology) under grants PD/BD/52355/2013 and PD/BD/128047/2016 during the development of this work.

6.7. REFERENCES

Aboueljinane, L., Sahin, E. & Jemai, Z. 2013. A review on simulation models applied to emergency medical service operations. Computers & Industrial Engineering, 66, 734-750.

Amorim, M., Ferreira, S. & Couto, A. 2016. Urban emergency medical service: Dynamic model for dynamic cities. University of Porto: University of Porto - Faculty of Engineering.

Amorim, M., Ferreira, S. & Couto, A. 2017. Road safety and the urban emergency medical service (uems): Strategy station location. Journal of Transport & Health, 6, 60-72.

Amorim, M., Ferreira, S. & Couto, A. Emergency medical service response: Analyzing vehicle dispatching rules. Transportation Research Board 97th Annual Meeting, 2018 Washington DC, United States.

Antunes, F., Ribeiro, B., Pereira, F. C. & Gomes, R. 2018. Efficient transport simulation with restricted batch-mode active learning. IEEE Transactions on Intelligent Transportation Systems, 1-10.

Bandara, D., Mayorga, M. E. & Mclay, L. A. 2014. Priority dispatching strategies for ems systems. Journal of the Operational Research Society, 65, 572-587.

Barton, R. R. & Meckesheimer, M. 2006. Chapter 18: Metamodel-based simulation optimization. In: Henderson, S. G. & Nelson, B. L. (eds.) Handbooks in Operations Research and Management Science. Elsevier.

Bélanger, V., Kergosien, Y., Ruiz, A. & Soriano, P. 2016. An empirical comparison of relocation strategies in real-time ambulance fleet management. Computers & Industrial Engineering, 94, 216-229.

Berg, P. L. V. D., Essen, J. T. V. & Harderwijk, E. J. Comparison of static ambulance location models. 2016 3rd International Conference on Logistics Operations Management (GOL), 23-25 May 2016 2016. 1-10.

Blanning, R. W. 1974. The sources and uses of sensitivity information. Interfaces, 4, 32-38.

Church, R. & Velle, C. R. 1974. The maximal covering location problem. Papers in Regional Science, 32, 101-118.

Daskin, M. S. 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. Transportation Science, 17, 48-70.

Daskin, M. S. & Stern, E. H. 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. Transportation Science, 15, 137-152.

Dibene, J. C., Maldonado, Y., Vera, C., De Oliveira, M., Trujillo, L. & Schütze, O. 2017. Optimizing the location of ambulances in tijuana, mexico. Computers in Biology and Medicine, 80, 107-115.

Erkut, E., Ingolfsson, A. & Erdogan, G. 2008. Ambulance location for maximum survival. Naval Research Logistics, 55, 42-58.

Fujiwara, M., Kachenchai, K., Makjamroen, T. & Gupta, K. 1988. An efficient scheme for deployment of ambulances in metropolitan bangkok. Operational research, 87, 730-741.

Fujiwara, O., Makjamroen, T. & Gupta, K. K. 1987. Ambulance deployment analysis: A case study of bangkok. European Journal of Operational Research, 31, 9-18.

Gendreau, M., Laporte, G. & Semet, F. 1997. Solving an ambulance location model by tabu search. Location Science, 5, 75-88.

He, Z., Qin, X., Xie, Y. & Guo, J. 2018. A service location optimization model for improving rural emergency medical services.

Iannoni, A. P., Morabito, R. & Saydam, C. 2009. An optimization approach for ambulance location and the districting of the response segments on highways. European Journal of Operational Research, 195, 528-542.

Kleijnen, J. P. C. 1975. A comment on blanning's “metamodel for sensitivity analysis: The regression metamodel in simulation”. Interfaces, 5, 21-23.

Knight, V. A., Harper, P. R. & Smith, L. 2012. Ambulance allocation for maximal survival with heterogeneous outcome measures. Omega, 40, 918-926.

Li, X., Zhao, Z., Zhu, X. & Wyatt, T. 2011. Covering models and optimization techniques for emergency response facility location and planning: A review. Mathematical Methods of Operations Research, 74, 281-310.

Maxwell, M. S., Restrepo, M., Henderson, S. G. & Topaloglu, H. 2010. Approximate dynamic programming for ambulance redeployment. INFORMS Journal on Computing, 22, 266-281.

Mayorga, M. E., Bandara, D. & Mclay, L. A. 2013. Districting and dispatching policies for emergency medical service systems to improve patient survival. IIE Transactions on Healthcare Systems Engineering, 3, 39-56.

McCormack, R. & Coates, G. 2015. 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, 294-309.

Rasmussen, C. E. 2004. Gaussian processes in machine learning. In: Bousquet, O., Von Luxburg, U. & Rätsch, G. (eds.) Advanced lectures on machine learning: Ml summer schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, revised lectures. Berlin, Heidelberg: Springer Berlin Heidelberg.

Repede, J. F. & Bernardo, J. J. 1994. Developing and validating a decision support system for locating emergency medical vehicles in louisville, kentucky. European Journal of Operational Research, 75, 567-581.

Restrepo, M., Henderson, S. G. & Topaloglu, H. 2008. Erlang loss models for the static deployment of ambulances. Health Care Management Science, 12, 67-79.

Revelle, C. & Hogan, K. 1989. The maximum availability location problem. Transportation Science, 23, 192-200.

Schmid, V. & Doerner, K. F. 2010. Ambulance location and relocation problems with time-dependent travel times. European Journal of Operational Research, 207, 1293-1303.

Snyder, L. V. & Daskin, M. S. 2005. Reliability models for facility location: The expected failure cost case. Transportation Science, 39, 400-416.

Song, W., Han, K., Wang, Y., Friesz, T. & Del Castillo, E. 2017. Statistical metamodeling of dynamic network loading. Transportation Research Procedia, 23, 263-282.

Su, S. & Shih, C.-L. 2003. Modeling an emergency medical services system using computer simulation. International Journal of Medical Informatics, 72, 57-72.

Toregas, C., Swain, R., Revelle, C. & Bergman, L. 1971. The location of emergency service facilities. Operations Research, 19, 1363-1373.

Tufuor, E. O., Nam, Y. & Rilett, L. R. 2018. Land suitability analysis for ems posts along state highways-a case study of California.

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

Van Essen, J. T., Hurink, J. L., Nickel, S. & Reuter, M. 2013. Models for ambulance planning on the strategic and the tactical level, Netherlands, TU Eindhoven, Research School for Operations Management and Logistics (BETA).

Yue, Y., Marla, L. & Krishnan, R. 2012. An efficient simulation-based approach to ambulance fleet allocation and dynamic redeployment. In Twenty-Sixth AAAI Conference on Artificial Intelligence.

7. EMERGENCY MEDICAL SERVICE RESPONSE: ANALYZING VEHICLE DISPATCHING RULES[19]

Marco Amorim[20], Sara Ferreira[20], Antonio Couto[20]

Abstract

In an era of information and advanced computing power, emergency medical services (EMS) still rely on rudimentary vehicle dispatching and reallocation rules. In many countries, road conditions such as traffic or road blocks, exact vehicle positions, and demand prediction are valuable information that is not considered when locating and dispatching emergency vehicles.

Within this context, this paper presents an investigation of different EMS vehicle dispatching rules by comparing them using various metrics and frameworks. An intelligent dispatching algorithm is proposed, and survival metrics are introduced to compare the new concepts with the classical ones.

This work shows that the closest idle vehicle rule (classical dispatching rule) is far from optimal and even a random dispatching of vehicles can outperform it. The proposed intelligent algorithm has the best performance in all the tested situations where resources are adequate. If resources are scarce, especially during peaks in demand, dispatching delays will occur, degrading the system's performance. In this case, no conclusion could be drawn as to which rule might be the best option. Nevertheless, it draws attention to the need for research focused on managing dispatch delays by prioritizing the waiting calls that inflict the higher penalty to the system performance.

Finally, the authors conclude that the use of real traffic information introduces a considerable gain to the EMS response performance.

Keywords: Vehicle dispatching, Emergency medical service, Intelligent algorithm, Victims survival, Simulation

---

[19] Amorim, M., Ferreira, S., & Couto, A. (2018). Emergency Medical Service Response: Analyzing Vehicle Dispatching Rules. Transportation Research Record.

Final version at https://doi.org/10.1177/0361198118781645

[20] CITTA, University of Porto – Faculty of Engineering, Porto, Portugal

7.1. INTRODUCTION

7.1.1. MOTIVATION AND CONTRIBUTION

Emergency medical services (EMS) are a vital part of today's modern cities. These are services that must respond promptly to medical emergencies, which may occur at any time and any place.

Depending on the country, the EMS "modus operandi" may vary, but in general, it relies on an emergency phone line (e.g., 911 in the USA, 112 in the EU) that directs all the medical emergencies to an EMS call center. The operator acquires the relevant information from the caller and makes a pre-assessment of the medical emergency needs. Usually, the EMS call center has an implemented algorithm that manually or automatically defines the priority of the emergency and triggers a request for a vehicle dispatching. The most frequently used EMS dispatching rule is to send the closest idle vehicle.

Nevertheless, it is important to investigate why and if the closest vehicle is really the best practice. One might conclude that, individually, the faster the victim is rescued, the better. However, sending that vehicle might debilitate the system's response for the next emergencies compared to sending a different one. This is particularly true if we take into account the type and severity of the case, where the response time may impact the victims' outcome.

Moreover, the urban environment where the system is implemented also influences the quality of the response it can provide. In fact, time-dependent information within this urban environment plays an important role in the system's response, as shown by Schmid (2012). With recent technological improvements, practitioners and researchers are in a position to collect and integrate, in real time, large amounts of diversified data, such as traffic congestion, and real-time emergency calls (Ferrucci and Bock, 2014).

One could claim that dispatching an emergency vehicle is a complex decision that goes beyond the request for the closest unit available. It requires the management of a vital network of vehicles, the assessment of the victims' severity and heterogeneity, and an understanding of the urban environment where the system is integrated.

The objective of this study is to investigate the previous claim by assessing several dispatching rules in different temporal contexts. This is achieved by using various performance metrics to evaluate the versatility of the closest vehicle dispatching rule and its possible substitutes. A simulation model and a dispatching rule are proposed by the authors; different performance metrics are applied and vehicles dispatching rules performance are assessed in different contexts with the objective of providing practitioners with a proof of concept, applicability, and empirical results in different emergency contexts.

7.1.2. LITERATURE REVIEW

The classical emergency service problem can be traced back to 1955, with the fire station location planning study by Valinsky (1955). Additionally, Hogg (1968) along with Savas (1969) completed the base archetypes for this field of investigation, where the latter gives special focus to the EMS. EMS station and vehicle locations, allocation and dispatching have since been widely studied; however, there has always been a gap between theory and practice, owing to the lack of possibilities to test these models in a real environment.

With computers becoming ever more powerful and accessible to everyone, simulation is also becoming an interesting tool researchers to formulate more realistic and complex models, either to assess solutions or to support optimization models (Restrepo et al., 2008, Maxwell et al., 2010, Yue et al., 2012, McCormack and Coates, 2015, Iannoni et al., 2009, Su and Shih, 2003). Haghani and Yang (2007) propose a simulation model to assess EMS performance. To abstract from the complexity of the road system, they adopt a nodal network and assume that real-time traffic information is known. McCormack and Coates (2015) also use simulation to assess vehicle allocation performance with a focus on increasing victims' survival. The use of simulation has proven to be valuable owing to the ability of directly using real EMS call data, compared with other methods where demand must be modeled, and thus simplified.

When assessing EMS performance, researchers fall back on metrics from the two most relevant works on EMS - those of Toregas et al. (1971) and Church and Velle (1974). The former presents a solution to the location set covering problem (LSCP), guaranteeing that all demand is covered within a maximum time or distance radius. Church and Velle (1974) approach the problem with a solution for the maximal coverage location problem (MCLP), which intends to overcome the resource limitations neglected by Toregas et al. (1971). Nevertheless, once a facility is called to service, its allocated demand points are no longer covered. Daskin and Stern (1981), (1983) and Hogan and ReVelle (1986), (1989) tackle this problem by adding facility busy probability and reliability.

In these models, the metric used to evaluate the system performance and/or optimize it is the response time, which is usually simplified by a maximum response time threshold, e.g., percentage of population covered within a response radius of 8 min (Amorim et al., 2017). Although it is obvious that a quicker medical response will always result in improved medical assistance (Blackwell and Kaufman, 2002, Pons et al., 2005), the response time affects different types of medical emergencies in different ways. For instance, Sánchez-Mangas et al. (2010) indicated that a response time reduction of 10 min could result in a 30% reduction of fatalities in traffic crashes, whereas Valenzuela et al. (2000) showed that cardiac arrest fatalities could be reduced by 50% if the victims were assisted no later than three minutes after a collapse. Thus, by relying on homogenous performance metrics that are time- or distance-based, no consideration to the victims' survival and heterogeneity is made. Erkut et al. (2008) point out that the trend in EMS response research is to substitute time- and distance-covering concepts with concepts that account for survival probabilities. This type of metric is used in recent works (Knight et al., 2012, McCormack and Coates, 2015, Amorim et al., 2017) and shown to be more suitable when assessing EMS from the user perspective; additionally, it allows for benchmarking comparisons (Knight et al., 2012, Erkut et al., 2008).

Yet, as of today, vehicle dispatching rules for medical emergency requests follow distance- or time-based metrics such as the classical closest idle vehicle dispatching rule (Haghani and Yang, 2007, Jagtenberg et al., 2017, Yang et al., 2005), which consist of allocating the vehicle that is closest to the emergency occurrence site. This rule is tied to the previously mentioned classical performance metrics that focus on the overall system response time. Thus, there is no consideration for the victims' heterogeneity or survivability. Several studies have started to investigate the fact that the closest idle vehicle is not always the optimal solution. The subsequent closest vehicles could still provide acceptable service while the closest vehicle provides better coverage of the network if it is available in the following hours (Jagtenberg et al., 2016, Carter et al., 1972).

In many countries, the emergency response service does not even consider real-time information when assessing the closest idle vehicle. Instead, they mostly rely on spatial distances or on the operators' skills and experience. Moreover, time-dependent information has been shown to be one of the key factors for better vehicle dispatching rules (Schmid, 2012).

Clearly, real-time information and different vehicle dispatching rules are the most recent topics in EMS optimization. Travel times and changes with respect to EMS call volumes are used by Schmid (2012) to achieve a decrease in the average response time of emergency services. Haghani et al. (2003) propose an optimization model for real-time emergency vehicle dispatching and routing, using real-time traffic information to better support dispatching decisions. Nevertheless, the proposed model presents major issues when applied in real-time situations, owing to its computation time burden. Thus, Haghani and Yang (2007) upgrade this model by drastically reducing the computational burden using the rolling-horizon approach, and adding coverage concerns for future demand. Undoubtedly, demand prediction presents an interesting challenge for researchers (Amorim et al., 2017), as does the use of real-time information to better model dispatching rules. Li et al. (2017) uses the uncertainty theory to deal with uncertain factors, such as demand, when dispatching medical emergency resources. Still, they do not consider environmental randomness, which can be captured using simulation. On the contrary, Knopps and Lundgren (2016) compile the most common dispatching rules the classical closest rule, and the preference rule, which tries to minimize the response time for high-priority calls (Bandara et al., 2014) - and propose new ones. These are the modified preference rule, which adds to the original a maximum response time threshold for lower-priority calls, and the preparedness rule, which employs a function that measures the preparedness of the system for future calls and tries to maximize it when dispatching vehicles to lower-priority emergencies. In the work of Jagtenberg et al. (2017) a benchmark model is proposed and when applied in a case study, the authors were able to show that the closest idle vehicle is a factor of 2.7 away from the optimum.

These state-of-the-art research topics (real-time information, EMS heterogeneity, survival indicators) are scattered and analyzed individually, and thus no relevant research combines them with emergency vehicle dispatching rules performance and assesses their consequences.

7.2. METHODOLOGICAL APPROACH

The literature review clarified that simulation is a superior tool to assess EMS dispatching rules performance with realism and accuracy. Furthermore, it was pointed that although the preferred EMS performance metric is the system response time, survival functions are more suitable for measuring the victims' outcome.

Consequently, to investigate this work claim, a simulation is proposed using an agent-based model. The authors suggest performance metrics related to response time and survival, and dispatching rules based on the classical closest idle vehicle rule and survival functions.

The method consists of building a simulation of a real city with real emergency call data to test different scenarios, in which the different dispatching rules are applied in different conditions, such as the existence or not of real-time information and various during periods of the year. The simulation results are calculated based on the chosen performance metrics and a comparison is made.

7.2.1. SIMULATION MODEL

An agent-based model is used to test the EMS vehicle dispatching rules and simulate urban environment conditions, in which a city agent controls a group of lower-level agents: events, road network, vehicles, and nodes, as shown in Algorithm 1.

The city agent is the main model agent and is responsible for storing and controlling all other agents by giving them update requests. The EMS calls are simulated by an event agent, which is responsible to activate events and keep track of their status. When an event is activated, this agent sends an assistance request to the city agent. When assisted by a vehicle agent, the event agent is also responsible for requesting directions to the nearest hospital from the city agent.

ALGORITHM 1 Simulation algorithm

Definitions:

N = set of nodes n

n = node, where s = node of type station and h = node of type hospital

V = set of vehicles v_s

v_s = vehicle in station s

S = set of stations s

H = set of hospitals h

E = set of events e^t_n

e^t_n = emergency event occurring at node n during t

M = set of matrices M^t

M^p = matrix of real travel times for period p

T = total simulation time

t = time

$step$ = temporal resolution

$f(\,)$ = programming function

While $t < T$:

1. **Update city( )** "set t and activate e^t_n."
2. **Update network( )** "interact through every v_s to travel one $step$ and transfer it to destination nodes"
3. **Update events( )** "activate e^t_n and the vehicle dispatching algorithm"
   - -Network calculates time travel from all stations
   - -Network returns the shortest time travel
   - -Vehicle dispatching algorithm runs
4. **Update vehicles job( )** "updates v_s status"
   - -If v_s arrived to e^t_n, activate assisting timer
   - -If assisting timer ends, request network to be processed to h
   - -If v_s arrived to h, transfers v_s to s.
5. **Update results( )** "calculates the EMS performance at the current $step$"
6. $t = t + step$

To simulate the transport network, a network agent is created with the responsibility of routing all vehicles and assisting in vehicle dispatching and transport. It is also responsible for calculating the fastest origin–destination routes for different temporal scenarios or network status.

A vehicle agent is added to simulate each individual EMS transport/assistance unit. Each vehicle agent keeps track of its position and informs the network agent when it arrives at any destination. This is simulated on a microscale by tracking the distance when traveling between nodes; the delays and assisting times that are randomly generated by the event agent. The vehicle agent transports victims (event agents) from the occurrence node to the hospital node. It is also responsible to keep track of the station where it must return after a job is completed.

The node agent can be of three types: network node, hospital node, or station node. This agent assists the network and city agent by storing vehicles and events, and constitutes the origins and destinations of the network agent. The routing is abstracted by a previous calculation of travel times between origins and destinations for different conditions, which are then stored in the network agent. This simplification does not detrimentally affect the methodology, because this research does not focus on vehicle rerouting. Moreover, by simplifying, the simulation model becomes faster, thereby allowing for more complex dispatching algorithms and longer analyzed periods.

7.2.2. PERFORMANCE METRICS

The performance P_i of an EMS response to an event ***i*** of type ***t*** can be defined by a function that may depend on the time between the event start and the arrival of assistance vehicle, $\boldsymbol{r_i}$, given by equation (7.1).

$$P_i = f^t\left(r_i\right) \tag{7.1}$$

The classical response time metric, $\boldsymbol{P^c_i}$ can then be described as equation (7.2):

$$P_i^c = r_i \tag{7.2}$$

which is then generalized for the overall service either by its summed parts, equation (7.3) or its average, equation (7.4):

$$P^c = \sum_{i=1}^{n} r_i \tag{7.3}$$

$$\overline{P^c} = \frac{\sum_{i=1}^{n} r_i}{n} \tag{7.4}$$

This metric is sometimes simplified into a binary evaluation, $\boldsymbol{P^T_i}$, equation (7.5), where ***T*** is a time threshold:

$$\begin{cases} P_i^T = 0 \text{ if } r_i > T \\ P_i^T = 1 \text{ if } r_i \leq T \end{cases} \tag{7.5}$$

Recently, as pointed out in the literature review, these solely time-based metrics (or distance-based metrics if $\boldsymbol{r_i}$ is substituted by the distance between ***i*** and the station that responds to it) can be replaced by a survival function that measures the likelihood of a victim to survive a medical emergency of type ***e*** when assisted within $\boldsymbol{r_i}$, in line with the work of Erkut et al. (2008), equation (7.6):

$$P_i^s = \left(\mathrm{e}^{c^e \times r_i}\right)^{-1} \tag{7.6}$$

where $\boldsymbol{c^e}$ is a survival coefficient for an event of type $\boldsymbol{e}$. This metric can also be generalized to the system by taking the form of a sum or an average, equation (7.7):

$$P^s = \sum_{i=1}^{n} \left(\mathrm{e}^{c^e \times r_i} \right)^{-1} \text{ or } \overline{P^s} = \frac{\sum_{i=1}^{n} \left(\mathrm{e}^{c^e \times r_i} \right)^{-1}}{n} \tag{7.7}$$

7.2.3. DISPATCHING RULES

Three dispatching rules are proposed for this investigation, as shown in Algorithm 2. The classical rule in which the closest vehicle is dispatched, is named ClosestDR. When this rule uses real-time information of traffic it becomes RT-ClosestDR. A random vehicle dispatching rule, RT-RandomDR is considered to determine how choosing the closest vehicle might compare with choosing any other vehicle, proving the inadequacy of the classical rule.

Finally, an intelligent survival dispatching rule, RT-InteligentSurvivalDR, is proposed by the authors. The survival performance metric is used to calculate the system status after dispatching a certain vehicle, and the intelligent algorithm maximizes the system status, $\boldsymbol{S^p}$, at period $\boldsymbol{s}$ by dispatching to non-life-threatening emergencies the vehicle that penalizes $\boldsymbol{S^p}$ the least. This is calculated using the busy fraction $\boldsymbol{q_s}$ (Daskin, 1983) and the expected response time by Berg et al. (2016), according to Snyder and Daskin (2005):

$$S^p = \sum_i \sum_s \sum_k \sum_e d_i^{es} \left(\mathrm{e}^{c^e \times r_{is}^p} \right)^{-1} \times (1 - q_s) \times q^{k-1} \times z_{kis} \tag{7.8}$$

where:

$$q = \frac{\sum_e d_i^{ep}}{a_s} \tag{7.9}$$

$$q^{k-1} = \prod_{j=1}^{k-1} q_j \tag{7.10}$$

$\boldsymbol{d^{es}_i}$ is the predicted number of events of type $\boldsymbol{e}$ during period $\boldsymbol{s}$ at demand point $\boldsymbol{i}$, $\boldsymbol{a_j}$ is the number of idle vehicles allocated at station $\boldsymbol{s}$, and $\boldsymbol{z_{kis}}$ is 1 if $\boldsymbol{s}$ is the $\boldsymbol{k^{th}}$ nearest station to $\boldsymbol{i}$, 0 otherwise.

Thus, the decision of dispatching a vehicle from a station $\boldsymbol{s}$ to a certain event can be scored as $\boldsymbol{Q_s}$:

$$Q_s = \left(e^{c^e \times r_{is}^p} \right)^{-1} + S^{p'} \tag{7.11}$$

where $\boldsymbol{S^{p'}}$ is the system survival status when an ambulance from $\boldsymbol{s}$ was dispatched to $\boldsymbol{i}$, updating $\boldsymbol{a_s}$ to $\boldsymbol{a_s - 1}$.

ALGORITHM 2 Dispatching rules

While no v_s is dispatched:

If ClosestDR or RT-ClosestDR is active:

1. Sort S by ascending r_i

2. From index 0 to size of S, dispatch v_s if idle

If RT-RandomDR is active:

3. Select all s within r_i < maximum response limit which have at least one idle v_s. If none, go step 5.

4. Dispatch a random v_s from step 3.

5. Go step 1

If RT-InteligentSurvivalDR:

6. Select all s within r_i < maximum response limit which have at least one idle v_s. If none, go step 10.

7. For every s in step 6, calculate system status using equation (7.11)

8. Sort solutions from step 7 by descending Q_s

9. From index 0 to size of solution from step 7, dispatch v_s if idle

10. Go step 1

Both the RT-RandomDR and RT-InteligentSurvivalDR algorithms can be controlled by a threshold time to ensure that any emergency is assisted within a maximum response time limit. If no vehicle within this maximum response time limit is available, then the closest vehicle is dispatched. Additionally, they are only activated for non-life-threatening emergencies.

7.3. APPLICATION OF THE MODEL

To assess the validity of this work's arguments and the performance of the different dispatching rules, the methodology is applied to a simulation created with real data from San Francisco, Figure 35.

The data was collected from the U.S. Government's open data strategic American resources and consists of a collection of Calls-For-Service database, which includes all the fire units' responses to calls in a total of 4.4 million vehicles dispatched between 2000 and 2017. The fire department is responsible for managing the EMS calls and responses, requesting a private unit when required. Thus, the dispatch of private units (to 911 calls) is also recorded in the database.

The data were processed and filtered into a SQL database for easy access and data manipulation. From the same open data source, fire station locations were acquired and added to the SQL database.

Figure 35. San Francisco experiment area with the available stations on the left and the total number of calls (2000–2017) on the right. Each circle in the left map represents a station, and each square on the right map represents one grid cell.

The city was divided in a grid of 500 m × 500 m cells (right map in Figure 35) and each unit is represented by a node corresponding to its center, with a total of 518 nodes. The fire stations, a total of 45, were then assigned to the closest node as well as three fictional hospitals representing the major San Francisco hospitals. The lack of information on the destination hospital for each call necessitated a random allocation of calls to hospitals based on their proximity. It would be possible to identify the most likely hospital to receive the victims using the San Francisco EMS assignment algorithm. However, the required information for this allocation is also not present in the database.

To implement the system performance metrics and the simulation model, travel time matrices were built using real travel times collected from Google through its Directions API. This specific API allows the calculation of travel times for different days and hours. Because of the burden of this process and the API limitations, a resolution of week vs weekend with 3 h intervals was used when building the OD matrices.

The intelligent vehicle dispatching algorithm coefficients were trained with calls from 2010 to 2015 and the priority calls were defined as those tagged as “Potentially Life-Threatening” in the original data.

Finally, several runs of the simulation model were computed for different months of the year of 2016, different numbers of vehicles, and the different dispatching rules.

All the scripts for collecting the data and for the simulation model were programmed in Python 2.7 and were run on a machine with an Intel core i7 quad core processor at 1.73 GHz and 8 GB of RAM running a Windows 10 64-bit operating system. The running time for each experiment depends on the dispatching algorithm used and the number of simulated days, but each experiment took between 20 s and 4 h.

The relevant results were processed and compiled in comparative graphs, which are analyzed in the next chapter.

7.4. RESULTS

To analyze the performance of each dispatching rule for “Potentially Life-Threatening” emergency events, the ClosestDR algorithm is set as the base rule for comparison. Each rule is then tested for different scenarios, and different metrics are calculated: the survival rate, average response time, and number of assisted victims within 8 min. The use of the 8 min threshold is considered because this is a common threshold used in several works (Amorim et al., 2017, McCormack and Coates, 2015). For a better understanding of the results, several graphs were compiled, in which the accumulative performance of each dispatching rule is calculated as a gain over the base rule.

The methodology was first applied to the month of February and tested for each dispatching rule in configurations of one vehicle per station and two vehicles per station. This allows for an understanding of how the number of available vehicles might influence the EMS response and how each dispatching rule performs in each condition.

Table 12 presents a summary of the different performance metrics for the different configurations and dispatching rules, and Figure 36 shows the gain of each dispatching rule when compared with the base rule. There is an obvious increase in performance when the number of vehicles doubles. One of the reasons for this, is the fact that with the existence of two vehicles per station, there are no delays in responding to emergency calls. This means that at least one vehicle is always idle and available to be dispatched.

Interestingly, the random dispatching rule, RT-RandomDR, outperforms all other rules when resources are scarce (one vehicle per station). Nevertheless, when measuring the performance by the number of assisted victims within 8 min, the RT-InteligentSurvivalDR is shown to be optimal, and as expected, the RT-ClosestDR algorithm provides the overall lower average response time when all calls are considered.

From Figure 36, a break is visible in the tendency of each rule’s performance around the 4000th event. Further research revealed that this disruption corresponds to the period when the “Superbowl 50” event was held near the San Francisco area, which leads to the occurrence of several smaller events across the city, and probably an increase in the number of tourists/visitors. This also shows that in such an unpredictable situation, the ClosestDR algorithm performs better than the nonrandom rules, as it minimizes the delays felt during demand peaks. This can be ascribed to the fact that this algorithm acts as a restricted random dispatching rule, because it does not use real-time information.

To further investigate on the problem of sporadic events that may overload the EMS response and produce system delays, a simulation was placed for the period around New Year's Eve (from 30 December, 2016 to 3 January, 2017), which is believed to be an event likely to generate a high demand peak and may be relatable to the reader.

The results, shown in Figure 37, indicate that when the EMS resources (number of vehicles) are low, the system performance degrades independently of the rule applied, which is explained by the existence of delays. In a situation where resources are not scarce, each dispatching rule seems to perform similarly, recovering its usual performance when the demand peak vanishes. Moreover, it is important to note that the time response metric is the most sensitive one in encumbered periods and under low resources; because it is the only metric that captures delays.

Finally, the seasonal adaptability of each dispatching rule performance is analyzed, as shown in Figure 38, by running simulations of different months. For the simulations where the number of vehicles is scarce, the intelligent survival algorithm underperforms in nearly all of the analyzed months, owing to delay propagation. The exception is February, in which the closest dispatching rule is the weaker rule. In the other months, the best performance is achieved either by the closest rule or the random rule. When the number of vehicles per station is doubled, the intelligent survival dispatching rule again becomes the best practice, although the random dispatching rule performs in a similar fashion.

Apparently, the lack of sufficient resources to respond in a timely manner to every life-threatening emergency significantly affects the system's response by creating delays. This becomes more obvious for the RT-InteligentSurvivalDR algorithm and when the performance is measured by the accumulated response time. This gap is less noticeable in the other performance metrics.

Table 12. Summary of the results for the month of February with two different configurations: one vehicle per station and two vehicles per station. Different performance indicators are presented.

Dispatching rule	All calls				Possible life-threatening					
	average response (s)		average delay (s)		average response (s)		total events assisted in less than 8 min		average survival rate per event	
Nº vehicles/station	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
ClosestDR	585	337	24	0	634	348	5116	9463	42%	59%
RT-ClosestDR	520	287	64	0	601	295	6086	10397	45%	64%
RT-RandomDR	563	396	26	0	536	290	6617	10526	48%	64%
RT-InteligentSurvivalDR	609	329	36	0	558	288	6447	10571	47%	64%

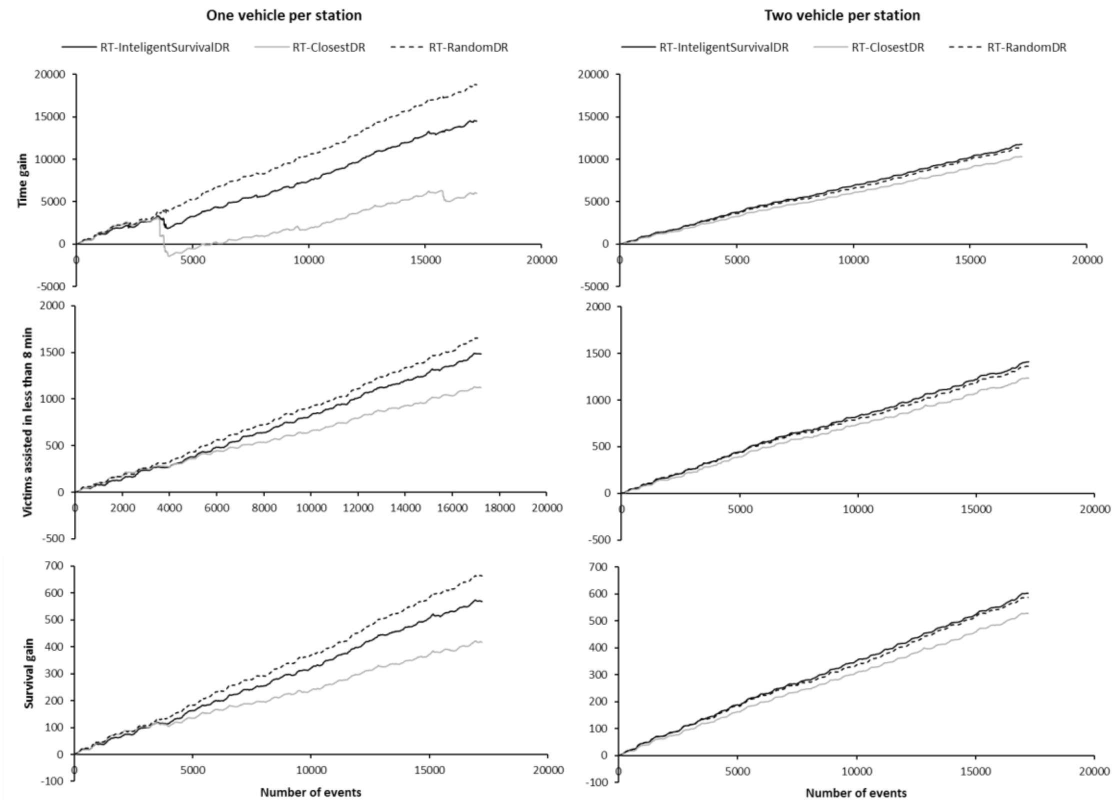


Figure 36. Gain of each dispatching rule when compared to the base rule: ClosestDR. The left side presents the dispatching rules' performance for a configuration of one vehicle per station, whereas the right side shows that for a configuration of two vehicles per station.

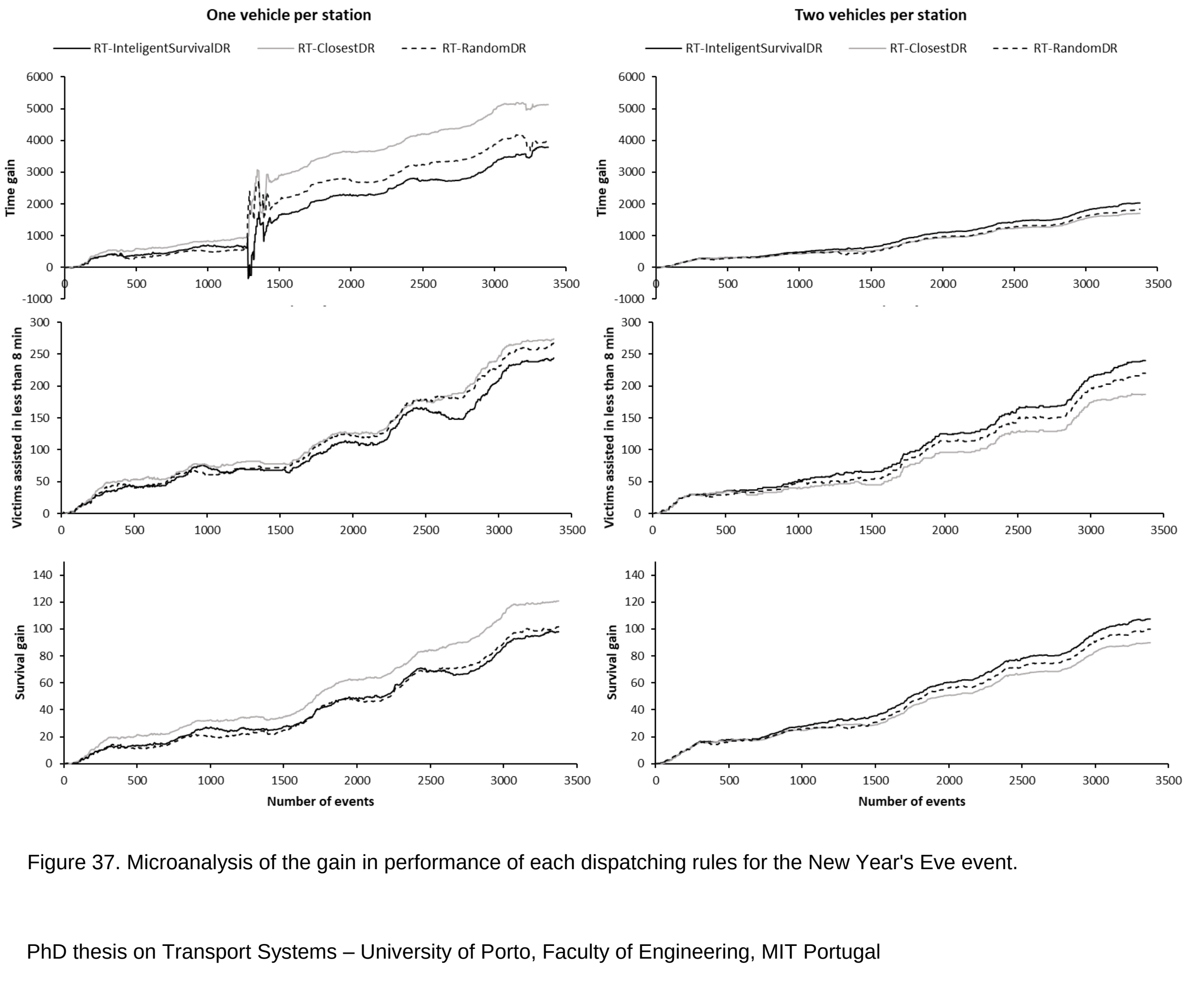


Figure 37. Microanalysis of the gain in performance of each dispatching rules for the New Year's Eve event.

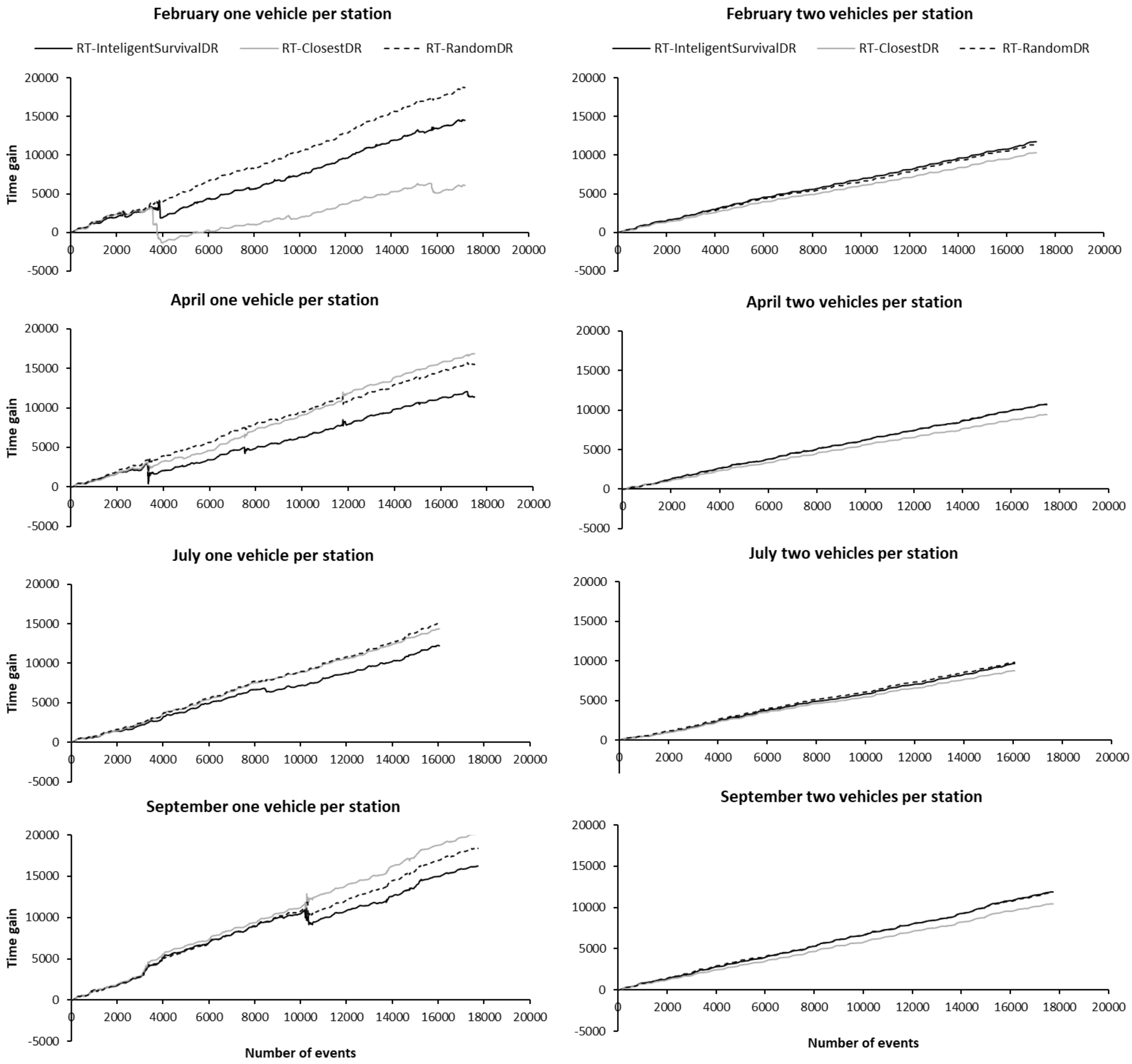


Figure 38. Comparison of the response time performance gain when compared with the base rule for different months of the year and different vehicle configurations.

7.5. CONCLUSIONS

This paper presents a methodology to assess EMS dispatching rules performance, providing several performance metrics and dispatching rules, including a dispatching rule proposed by the authors that aims to increase victims' survival, without excessively penalizing non-life-threatening emergencies. Furthermore, the method is simulated with real data from the city of San Francisco.

The presented results open new pathways to the study of EMS vehicle dispatching rules, their consequences, and the characteristics that influence their performance. This work makes an important contribution to the state of the art by providing a proof of concept and empirical results, which illustrate that the optimal dispatching rule is not the obvious one — sending the closest vehicle. They also show that having a single rule is never the optimal in the long term.

In an EMS system in which resources are not scarce, the proposed intelligent survival dispatching rule is the obvious choice in every scenario. Moreover, a random dispatch of vehicles (without penalizing non-life-threatening events by more than 15 min) can outperform the classical rule in almost every tested case.

This work proves that the use of real-time traffic information is essential in every situation, and significantly improves the system's performance in terms of response time, survival rate, and the total number of emergencies assisted within 8 min.

It was found that demand peaks, owing to special occasions on which the number of people in the same area increases or when people engage in more risky behaviors, present an interesting challenge, and the optimal solution such cases is difficult to determine with the presented dispatching rules. It came to the authors' attention that the lower performance of the dispatching rules in this situation is due to the poor handling of delays. It is also possible that different dispatching rules may fit better to different configurations of demand: local demand peaks, or dispersed demand peaks. This is a topic worth investigating.

It is important to note that the victims in a waiting list for a vehicle were processed using a greedy algorithm that tries to minimize the total number of delays by avoiding delay propagation. This is clearly not the optimal approach, and thus the authors propose the creation of delay-handling algorithms that can consider the same EMS performance principles that were discussed through this paper.

The simulation model accounts for "in situ" treatment time. Nevertheless, the survival function studies found on the state of the art do not account for hospital care and how pre-hospital time (which includes transport to a hospital) might further affect victims' outcomes (Kereiakes et al., 1990, Curtis et al., 2006, Gaspoz et al., 1996). Rapid diagnoses and hospital interventions are key factors affecting victims' survival. This is an interesting topic of investigation and could allow EMS planners to better manage resources, i.e., the type and quantity of vehicles to dispatch to each event, allowing a phased dispatch of a treatment and transport units, and a reallocation of the treatment unit to another event after "in situ" assistance.

The presented topic could be further advanced with the use of additional details in the simulation model and response policies, such as the use of different vehicle types, vehicle rerouting, and a wider discretization of victims' heterogeneity. Moreover, with the application of these concepts to other cities to further support this study's conclusions, practitioners will have more tools to implement better dispatching policies and/or identify critical problems in their response systems and make better tactical decisions.

This work highlights the fact that sending the closest vehicle (for non-life-threatening emergencies) to a certain demand point tends to debilitate that same demand point and its surroundings. This is because, statistically speaking, the same demand point proves to have a higher chance of having more emergencies, and thus the obvious dispatching rule is not always the optimal one.

7.6. Acknowledgments

We acknowledge the support of FCT (Portuguese national funding agency for science, research and technology) under the grant PD/BD/52355/2013 during the development of this work.

7.7. Author Contribution Statement

The authors confirm contribution to the paper as follows: study conception and design: Marco Amorim and António Couto; data collection: Marco Amorim and Sara Ferreira; analysis and interpretation of results: Marco Amorim, António Couto and Sara Ferreira; draft manuscript preparation: Marco Amorim. All authors reviewed the results and approved the final version of the manuscript.

7.8. References

Amorim, M., Ferreira, S. & Couto, A. 2017. Road safety and the urban emergency medical service (uEMS): Strategy station location. Journal of Transport & Health.

Bandara, D., Mayorga, M. E. & Mclay, L. A. 2014. Priority dispatching strategies for EMS systems. Journal of the Operational Research Society, 65, 572-587.

Berg, P. L. V. D., Essen, J. T. V. & Harderwijk, E. J. Comparison of static ambulance location models. 2016 3rd International Conference on Logistics Operations Management (GOL), 23-25 May 2016 2016. 1-10.

Blackwell, T. H. & Kaufman, J. S. 2002. Response time effectiveness:Comparison of response time and survival in an urban emergency medical services system. Academic Emergency Medicine, 9, 288-295.

Carter, G. M., Chaiken, J. M. & Ignall, E. 1972. Response areas for two emergency units. Operations Research, 20, 571-594.

Church, R. & Velle, C. R. 1974. The maximal covering location problem. Papers in Regional Science, 32, 101-118.

Curtis, J. P., Portnay, E. L., Wang, Y., Mcnamara, R. L., Herrin, J., Bradley, E. H., Magid, D. J., Blaney, M. E., Canto, J. G. & Krumholz, H. M. 2006. The pre-hospital electrocardiogram and time to reperfusion in patients with acute myocardial infarction, 2000–2002: Findings from the national registry of myocardial infarction-4. Journal of the American College of Cardiology, 47, 1544-1552.

Daskin, M. S. 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. Transportation Science, 17, 48-70.

Daskin, M. S. & Stern, E. H. 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. Transportation Science, 15, 137-152.

Erkut, E., Ingolfsson, A. & Erdogan, G. 2008. Ambulance location for maximum survival. Naval Research Logistics, 55, 42-58.

Ferrucci, F. & Bock, S. 2014. Real-time control of express pickup and delivery processes in a dynamic environment. Transportation Research Part B: Methodological, 63, 1-14.

Gaspoz, J. M., Unger, P. F., Urban, P., Chevrolet, J. C., Rutishauser, W., Lovis, C., Goldman, L., Héliot, C., Séchaud, L., Mischler, S. & Waldvogel, F. A. 1996. Impact of a public campaign on pre-hospital delay in patients reporting chest pain. Heart, 76, 150-155.

Haghani, A., Hu, H. & Tian, Q. An optimization model for real-time emergency vehicle dispatching and routing. 2003. Citeseer.

Haghani, A. & Yang, S. 2007. Real-time emergency response fleet deployment: Concepts, systems, simulation & case studies. In: Zeimpekis, V., Tarantilis, C. D., Giaglis, G. M. & Minis, I. (eds.) Dynamic fleet management: Concepts, systems, algorithms & case studies. Boston, MA: Springer US.

Hogan, K. & Revelle, C. 1986. Concepts and applications of backup coverage. Management Science, 32, 1434-1444.

Hogg, J. M. 1968. The siting of fire stations. J Oper Res Soc, 19, 275-287.

Iannoni, A. P., Morabito, R. & Saydam, C. 2009. An optimization approach for ambulance location and the districting of the response segments on highways. European Journal of Operational Research, 195, 528-542.

Jagtenberg, C. J., Bhulai, S. & Van Der Mei, R. D. 2016. Dynamic ambulance dispatching: Is the closest-idle policy always optimal? Health Care Management Science, 1-15.

Jagtenberg, C. J., Van Den Berg, P. L. & Van Der Mei, R. D. 2017. Benchmarking online dispatch algorithms for emergency medical services. European Journal of Operational Research, 258, 715-725.

Kereiakes, D. J., Weaver, W. D., Anderson, J. L., Feldman, T., Gibler, B., Aufderheide, T., Williams, D. O., Martin, L. H., Anderson, L. C., Martin, J. S., Mckendall, G., Sherrid, M., Greenberg, H. & Teichman, S. L. 1990. Time delays in the diagnosis and treatment of acute myocardial infarction: A tale of eight cities report from the pre-hospital study group and the cincinnati heart project. American Heart Journal, 120, 773-780.

Knight, V. A., Harper, P. R. & Smith, L. 2012. Ambulance allocation for maximal survival with heterogeneous outcome measures. Omega, 40, 918-926.

Knopps, L. & Lundgren, T. 2016. Modeling ambulance dispatching rules for EMS-systems.

Li, H., Peng, J., Li, S. & Su, C. 2017. Dispatching medical supplies in emergency events via uncertain programming. Journal of Intelligent Manufacturing, 28, 549-558.

Maxwell, M. S., Restrepo, M., Henderson, S. G. & Topaloglu, H. 2010. Approximate dynamic programming for ambulance redeployment. INFORMS Journal on Computing, 22, 266-281.

Mccormack, R. & Coates, G. 2015. 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, 294-309.

Pons, P. T., Haukoos, J. S., Bludworth, W., Cribley, T., Pons, K. A. & Markovchick, V. J. 2005. Paramedic response time: Does it affect patient survival? Academic Emergency Medicine, 12, 594-600.

Restrepo, M., Henderson, S. G. & Topaloglu, H. 2008. Erlang loss models for the static deployment of ambulances. Health Care Management Science, 12, 67-79.

Revelle, C. & Hogan, K. 1989. The maximum availability location problem. Transportation Science, 23, 192-200.

Sánchez-Mangas, R., García-Ferrrer, A., De Juan, A. & Arroyo, A. M. 2010. The probability of death in road traffic accidents. How important is a quick medical response? Accident Analysis & Prevention, 42, 1048-1056.

Savas, E. S. 1969. Simulation and cost-effectiveness analysis of new york's emergency ambulance service. Management Science, 15, B-608-B-627.

Schmid, V. 2012. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. European Journal of Operational Research, 219, 611-621.

Snyder, L. V. & Daskin, M. S. 2005. Reliability models for facility location: The expected failure cost case. Transportation Science, 39, 400-416.

Su, S. & Shih, C.-L. 2003. Modeling an emergency medical services system using computer simulation. International Journal of Medical Informatics, 72, 57-72.

Toregas, C., Swain, R., Revelle, C. & Bergman, L. 1971. The location of emergency service facilities. Operations Research, 19, 1363-1373.

Valenzuela, T. D., Roe, D. J., Nichol, G., Clark, L. L., Spaite, D. W. & Hardman, R. G. 2000. Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos. New England Journal of Medicine, 343, 1206-1209.

Valinsky, D. 1955. Symposium on applications of operations research to urban services—a determination of the optimum location of fire-fighting units in new york city. Journal of the Operations Research Society of America, 3, 494-512.

Yang, S., Hamedi, M. & Haghani, A. 2005. Online dispatching and routing model for emergency vehicles with area coverage constraints. Transportation Research Record: Journal of the Transportation Research Board, 1923, 1-8.

Yue, Y., Marla, L. & Krishnan, R. 2012. An efficient simulation-based approach to ambulance fleet allocation and dynamic redeployment. In Twenty-Sixth AAAI Conference on Artificial Intelligence.

8. Emergency Vehicles Dispatching Technological Advantages: Implementing Survival and Real-Time Information[21]

Marco Amorim[22], Sara Ferreira[22], Antonio Couto[22]

Abstract

In an era of fast technology improvements and easy access to real-time data from various sources, intelligent transport systems, ITS, emerge. Emergency medical services (EMS) response is one of the transport systems that can take advantage of technological advances. Moreover, cities present themselves as dynamic environments where traffic flows change during the day as well as people's location. Therefore, a static EMS response is inappropriate and unable to give a proper response at every period of a day with a reasonable ambulance fleet size.

This paper studies technological improvements to be applied at dispatching time and how these improvements will translate in the service performance. We propose a methodology that assesses theoretical concepts through a simulation model that intends to provide empirical evidence using operational and victims focused metrics. We propose a dispatching algorithm that combines drivability and demand predictions to ensure that the service area is always covered with the highest survival status.

We apply our methodology to a case study, Porto city, to validate it and assess the impact of ITS in dynamic environments.

Keywords: emergency medical service response; vehicle dispatching; intelligent EMS; real-time data, ITS.

[21] Emergency vehicles dispatching technological advantages: implementing survival and real-time information. Submitted to Transportation Research Part B: Methodological.

[22] CITTA, University of Porto – Faculty of Engineering, Porto, Portugal

8.1. INTRODUCTION

8.1.1. MOTIVATION

Emergency medical services (EMSs) fulfill a vital service in modern societies; they respond to medical aid calls, protect, and ensure public health and safety. The decision-making process in the EMS response management is a challenge with growing interest in the last decades. This process involves balances at all levels: between efficiency and efficacy, quality and costs, and health policy and resource allocation. EMS planning involves decisions from both strategic and tactical viewpoints. Strategic decisions include the number and location of response teams to attain overall system goals, while tactical decisions focus in responding to situations that arise given a fixed number of resources. Within the latter, the main decision is to select a response vehicle to dispatch to an emergency and possible reallocation of the remain fleet to better prepare for the next service hours.

When deciding on which vehicle to dispatch, both literature and EMS operators opt for the closest available vehicle policy (Aringhieri et al., 2017). This policy always decide for the closest available vehicle despite the system features, the environment where it operates, and the EMS call characteristics. Moreover, many EMS operators, and even researchers calculate closeness using distance metrics, many times approximated, and thus disregarding real-time drivability conditions. Two questions arise when this solution is at use: Can emerging technologies and new tools improve service performance at dispatching time by offering more accurate travel time predictions? Moreover, how much can be improved if vehicle dispatching considers the emergency, environment and system characteristics?

This work addresses these two questions by analyzing the use of real-time drivability conditions and study a dispatching rule that focuses on a survival performance metric and system status. The proposed dispatching rule tries to balance life-threatening and non-life-threatening emergencies improving response time to the former by allowing higher response times to the latter. Nevertheless, it is important to note that vehicle reallocation is out of the scope of this study because, first we want to solely assess improvements that do not require a change in resources management nor that require extra resources, and second, in many EMS systems, vehicles are tied to a specific location or belong to a specific institution. Therefore, they must always return or stay at the original location (McLay and Mayorga, 2013).

8.1.2. BACKGROUND

The most common problem studied in EMS response focus on the strategic decisions. Several models have been thoroughly studied over the last decades to allocate emergency vehicles to stations (Erkut et al., 2008, Gendreau et al., 2005, Marianov and Serra, 1998, Gendreau et al., 1997, ReVelle and Hogan, 1989, Hogan and ReVelle, 1986, Daskin, 1983, Church and Velle, 1974, Toregas et al., 1971, Araz et al., 2007, Mitsakis et al., 2014, Geroliminis et al., 2011). When it comes to tactical decisions (i.e., vehicles dispatching) the conventional rule is to dispatch the closest idle vehicle (Haghani and Yang, 2007, Jagtenberg et al., 2017, Yang et al., 2005). However, the closest idle vehicle is not always the optimal solution. Nearby vehicles could still provide an acceptable response while the closest vehicle provides better coverage of the network if it is available for the next hours (Jagtenberg et al., 2016, Carter et al., 1972). Nevertheless, when assessing the closest vehicle, EMS decision makers usually rely on average travel times which weakly characterize the drivability at the response instant (Ingolfsson et al., 2008).

The recent technological improvements have enabled practitioners and researchers to collect, in real-time, more substantial amounts of diversified data, which fueled a new interest to the classic dispatching and vehicle control problem in operational research and transportation. Sáez et al. (2008) formulated a pick-up and delivery problem considering future demand, and prediction of travel times. They measured the formulation benefits using a system cost function. More recently, Ferrucci and Bock (2014) used real-time control in the pick-up and delivery problem, integrating real-world aspects such as real-time requests, traffic congestion, and vehicle disturbances.

This improvement can be applied to the EMS dispatching problem, with the difference that in the EMS case we pick-up medical emergency victims, consequently the victims' survivability gains priority over operation cost, making waiting time for life-threatening emergencies a crucial factor in the solution. Thus, better integration of real-world aspects and accounting for the victims' characteristics in the system performance will lead to a better service. However, little work is available regarding real-time information and survival metrics for the EMS vehicles dispatching problem.

Haghani et al. (2003) study real-time traffic information to better support dispatching decisions by proposing an optimization model for real-time emergency vehicle dispatching and routing. Nevertheless, the proposed model computation time did not allow its application in real-time. Haghani and Yang (2007) tackled this problem by reducing the computational burden and further improve its scope by adding coverage concerns for future demand. Demand prediction presents an interesting challenge for researchers, as is the use of real-time information. Further arguments to support the importance of using time-dependent information for better vehicle dispatching rules have been shown by Schmid (2012). The author uses travel times and assumes demand changes to achieve a decrease in the average response time of emergency services. Noticeably, average response time or response time thresholds are still the favorite metrics of researchers (Schmid, 2012, Jagtenberg et al., 2017, Lam et al., 2015).

However, by relying on performance metrics that are based on time or distance, most of these models have little to no regards to the victims' survival and heterogeneity. In the last decade researchers started to focus on the emerging of new performance metrics, motivated by victims' outcomes, which use survival functions to assess EMS response, and showing how this type of performance metrics are more suitable when assessing emergency medical services (Knight et al., 2012, Erkut et al., 2008).

McLay and Mayorga (2013) study a model for optimally dispatching EMS vehicles by prioritizing emergency calls with a higher risk of being life-threatening situations. With another view, Knopps and Lundgren (2016) present several dispatching rules with the goal of having the system better prepared to respond to life-threatening or higher priority medical emergency calls by readjusting the way non-life-threatening or low priority calls are responded. Bandara et al. (2012), Bandara et al. (2014), and Mayorga et al. (2013) study different policies for dispatching EMS vehicles and use survival probabilities as a metric to better capture victims' outcomes. These works show that the closest vehicle policy is not always optimal and that a dispatching rule should consider victims' heterogeneity, particularly by prioritizing life-threatening situations. Similar conclusions with regards to the closest idle vehicle policy are made by Jagtenberg et al. (2016), Jagtenberg et al. (2017).

The research questions previously presented remain unanswered. The literature does not offer active dispatching rules that benefit from the new concepts of survival function and heterogeneity, nor directly assess the benefit of introducing real-time information or estimations (Vlahogianni et al., 2014) in EMS dispatching policies.

8.1.3. Objectives

In short, there is a huge gap between practice and theory when looking at dispatching policies and the implementation of the available state of the art technologies. With the new computer age and the rise of smart cities, tremendous amount of real-time data is easily available which can be used to make the jump from theory to practice. In this context, we suggest a follow up on dispatching policies by assessing the use of real-time information and proposing a vehicle-dispatching rule that focus on the victims' survival, IRTADA (Intelligent Real-Time Ambulance Dispatching Algorithm), and contribute to state of the art by:

- Implementing survival functions in the dispatching rule;
- Acknowledging daily urban changes such as changes to traffic and to population location in dispatching policies, and
- Assessing the contribution of real-time information on drivability and medical emergency events at decision time.

With this work, we claim that the classic dispatching policy, still overly used in EMS practices (Sung and Lee, 2016), and dispatching practices that neglect real-time information are myopic when survival is at stake; particularly when applied to dynamic urban areas. This is supported by testing the use of real-time drivability information in a dispatching policy, and by comparing a real-time survival dispatching algorithm with the closest vehicle policy through simulation.

8.2. FRAMEWORK

We explore the technological advances at the EMS tactical level in two parts, as shown in Figure 39. In a first part- methods, we propose the tools to simulate and test dispatching policies in accordance with the technological advances. In a second part - calculations, we measure the advantage of technological advances such as real-time drivability information and survival when implemented in the dispatching policies.

The methods propose a platform to test how the presented concepts perform, a metric that can capture the emerging concepts of survival, and a dispatching policy that takes patients heterogeneity and survival as the goal to optimize. Because it is infeasible to test these concepts in a real environment, a simulation model is proposed. With this simulation model, technological advances can be implemented, and tests can be made to understand the influence of real-time drivability information in the response time and how accurate this information must be. Afterward, based on a performance metric that focuses on survivability, a dispatching rule is proposed and tested in the simulation model against the closest vehicle rule.

The mentioned tests are made during the calculations' step through a study case and sensitivity analysis of key parameters as described in Figure 40.

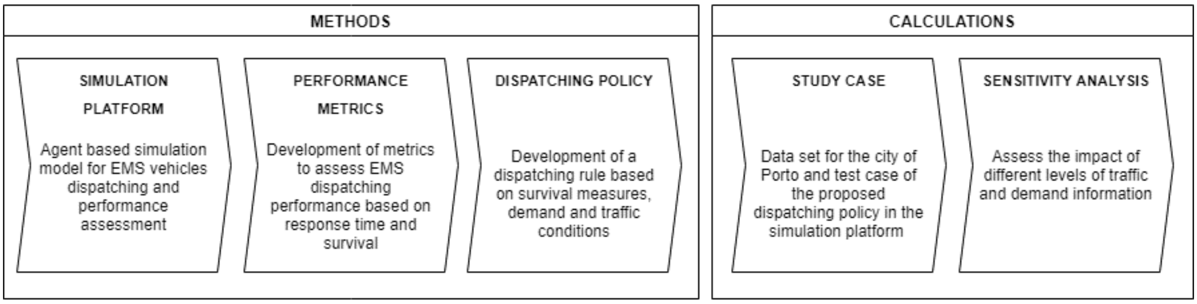


Figure 39. Framework flowchart

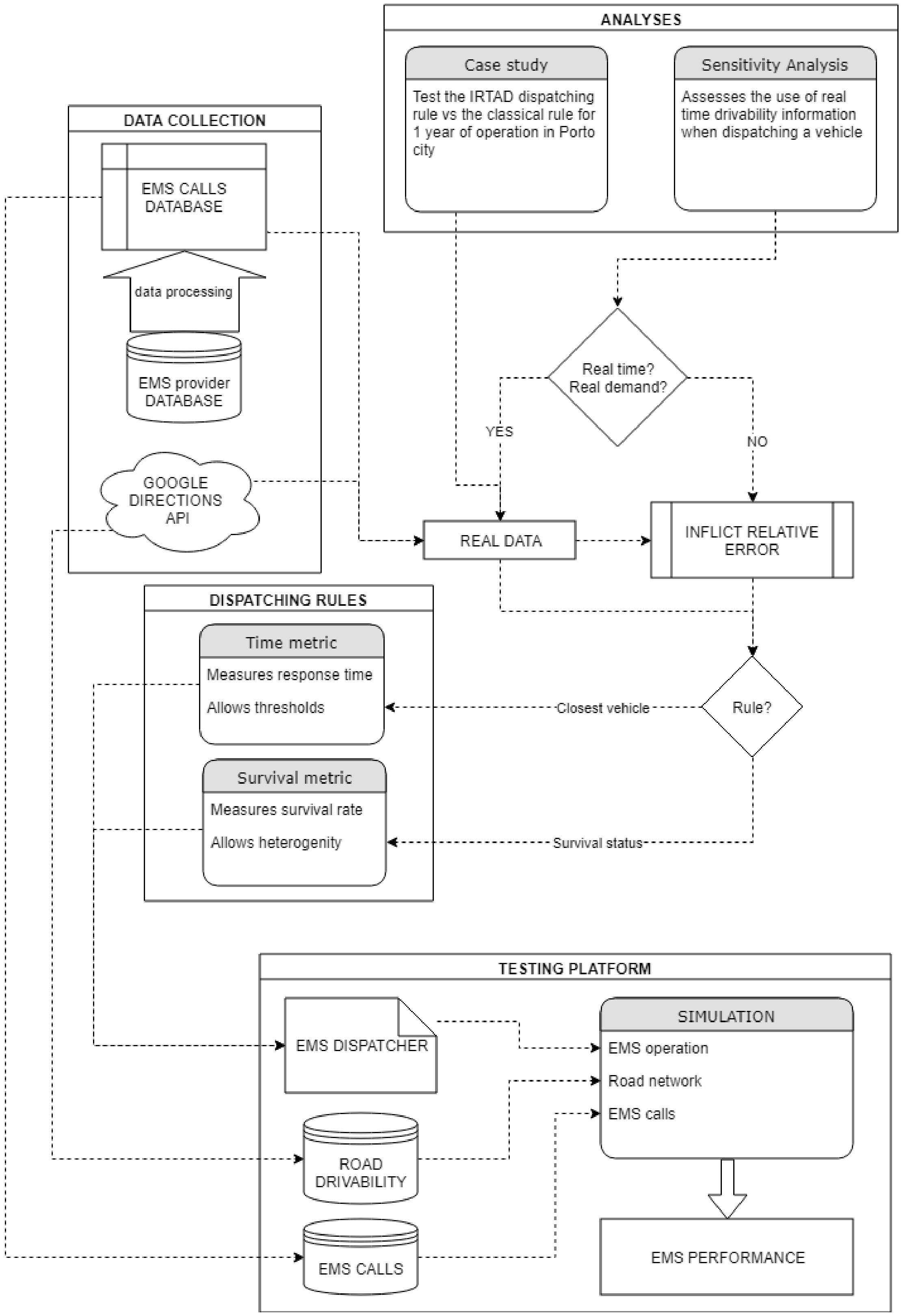


Figure 40. Framework pipelines.

8.3. METHODS

8.3.1. SIMULATION MODEL

A simulation algorithm to empirically compute solutions performance is constructed similarly to several models found on the literature (Haghani and Yang, 2007, McCormack and Coates, 2015, Su and Shih, 2003) but adapted to focus solely on the influence of the dispatching rule. This means that the downstream behavior of the simulation is simplified, thus not accounting for decisions at the event site or the hospital. Furthermore, the necessary processes that precede the decision of which vehicle to dispatch are omitted as they are irrelevant for the dispatching policy performance. A brief resume of the simulation is presented in Algorithm 1. The simulation uses an agent-based approach to simulate the actions of the autonomous agents, i.e., vehicles, EMS dispatcher/city, events. This provides a simulation platform that allows for complex dispatching rules and prioritization of active emergency calls.

Algorithm 1 Simulation algorithm

Definitions:

$\boldsymbol{N}$ = *set of nodes* $\boldsymbol{n}$

$\boldsymbol{n}$ = *node*, where $\boldsymbol{s}$ = *node of type station* and $\boldsymbol{h}$ = *node of type hospital*

$\boldsymbol{V}$ = *set of vehicles* $\boldsymbol{v_s}$

$\boldsymbol{v_s}$ = *vehicle allocated to station* $\boldsymbol{s}$

$\boldsymbol{S}$ = *set of stations* $\boldsymbol{s}$

$\boldsymbol{H}$ = *set of hospitals* $\boldsymbol{h}$

$\boldsymbol{E}$ = *set of events* $\boldsymbol{e^t_n}$

$\boldsymbol{e^t_n}$ = *emergency event occurring at node* $\boldsymbol{n}$ *during* $\boldsymbol{t}$

$\boldsymbol{M}$ = *set of matrices* $\boldsymbol{M^t}$

$\boldsymbol{M^p}$ = *matrix of real travel times for period* $\boldsymbol{p}$

$\boldsymbol{T}$ = *total simulation time*

$\boldsymbol{t}$ = *time*

step = *time resolution*

f( ) = *programming function*

While $t < T$:

7. *Update city( )* "set $\boldsymbol{t}$ and activate $\boldsymbol{e^t_n}$."
8. *Update network( )* "interact through every $\boldsymbol{v_s}$ to travel one ***step*** and transfers it to destination nodes"
9. *Update events( )* "activates e^t_n and the vehicle dispatching algorithm"

    -Network calculates time travel from all stations

    -Network returns calculations to the EMS dispatcher

    -Vehicle dispatching algorithm runs

10. *Update vehicles job( )* "updates $\boldsymbol{v_s}$ status"

    If $\boldsymbol{v_s}$ arrived at $\boldsymbol{e^t_n}$, activate assisting timer

    If assisting timer ends, request network to be processed to $\boldsymbol{h}$

If $\boldsymbol{v_s}$ arrived at $\boldsymbol{h}$, transfers $\boldsymbol{v_s}$ to $\boldsymbol{s}$.

11. *Update results( )* “calculates the EMS performance, ***P***, at the current *step*”
12. $\boldsymbol{t} = \boldsymbol{t} +$ ***step***

This model is controlled by a city *agent* that takes the role of the emergency medical service entity and allocates and dispatches vehicles using the closest dispatching rule (Haghani and Yang, 2007, Jagtenberg et al., 2017, Yang et al., 2005) or any proposed rule. A network agent is used to simulate road drivability conditions and the EMS vehicle movements by using a nodal network as an abstraction of the road network. This agent uses pre-computed travel times for different daily conditions obtained using Google’s Directions API. An Event agent simulates the population request medical emergency assistance to the EMS entity. The network tracks the vehicle agent’s movement from its origin to its destination and the vehicle agent assists and transports the occurrence from its location to a hospital. The required assistance time is a property of the event which is generated by the event agent. A data agent is implemented with access to every agent, being able to request any information and using this information to calculate the service performance. This agent can provide the necessary information to the city agent if it is necessary to the dispatching rule.

8.3.2. Performance metrics

In the EMS strategical plan, decisions are made towards the maximization of a certain goal. When dealing with vehicle dispatching decisions, this goal usually focuses on providing the quickest response to an active medical emergency, ***m***. Therefore, one can assume that the vehicle response is the metric to optimize, be it measured as driving time or driving distance. This is represented by response performance metric $\boldsymbol{P^r_m}$ defined as the symetric of a driving measurement (time or distance) that defines the physical or temporal separation between the location of the dispatched vehicle (***j***) and the location of the event (***i***), $\boldsymbol{r_{ij}}$:

$$P^r_m = -\sum_j r_{ij} \times v_j \tag{12}$$

Where $\boldsymbol{v_j}$ is 1 if vehicle at location ***j*** is dispatched and $\boldsymbol{\Sigma v_j = 1}$.

Therefore, the response performance of the system, $\boldsymbol{P^r}$, can be assumed as the sum of all responses or their average, where ***M*** is the set of all events ***m***:

$$P^r = \sum_m P^r_m \text{ or } P^r = \frac{\sum_m P^r_m}{M} \tag{13}$$

The latter performance, $\boldsymbol{P^r_m}$, can also be translated into a binary metric which will measure if a certain event is responded within a time threshold $\boldsymbol{T^D}$:

$$P_m^{T^D} = \begin{cases} 1 \text{ if } P_m^r \leq T^D \\ 0 \text{ if } P_m^r > T^D \end{cases} \tag{14}$$

Moreover, the system performance is then either a sum or the average of all individual performances:

$$P^{T^D} = \sum_m P_m^{T^D} \text{ or } P^{T^D} = \frac{\sum_m P_m^{T^D}}{M} \tag{15}$$

Then, we assume that events are heterogeneous and survival is the goal. Therefore, the survival performance $\boldsymbol{P^s_m}$ of the EMS response to an event $\boldsymbol{m}$ of type $\boldsymbol{e}$ is rather defined by a survival function that depends on the time between the event starts and the arrival of the assistance team, $\boldsymbol{r_m}$. The state of the art survival approach in EMS points exponential functions as seen in Erkut et al. (2008). Let $\boldsymbol{c^e}$ be a survival coefficient and $\boldsymbol{k^e}$ a survival constant for the event of type $\boldsymbol{e}$. Thus, $\boldsymbol{P^s_m}$ comes as:

$$P_m^s = \left(1 + e^{k^e + c^e \times r_m}\right)^{-1} \tag{16}$$

As for the response performance, the survival performance of the system, $\boldsymbol{P^s}$, can be assumed as the sum of all survival performances or their average:

$$P^s = \sum_m P_m^s \text{ or } P^s = \frac{\sum_m P_m^s}{M} \tag{17}$$

It is important to understand that these performance metrics can be used to assess solutions during dispatch time, but also to evaluate the dispatching decision at the end of the event or events. Clearly, depending on the available information at decision time, both evaluations can lead to different results i.e. at dispatching time only distance-based measurement exist, while at event closure the real travel time (or response time) is available.

Moreover, these differences can mean that the selected solution at dispatching time might not be the optimal one even at the single event view. A mismatch between the estimated time or distance and the time or distance the vehicle actually traveled can be enough that a vehicle from a different station would be preferable.

8.3.3. DISPATCHING POLICIES – CLASSIC VERSUS IRTADA

The dispatching of medical emergency vehicles is part of a complex management process that exists within EMS. Taking the Portuguese example when an emergency call arrives at the call center the first filter is applied to reroute medical emergency calls to the INEM (Portuguese National Institute of Medical Emergencies) service. There, an operator responds to the call and, using a proprietary algorithm, goes through a sequence of multiple-choice questions that automatically triggers a request for the appropriate response. This response arrives at a dispatching operator that enquires the dispatching system for vehicles location and availability. Then, the system responds with an ordered list of the closest idle vehicles to the event location – closest idle policy. This is the *modus operandi* of most of the European and American systems, and although the pipeline for a call to arrive at the dispatching decision may vary, usually the closest idle policy is at practice (Aringhieri et al., 2017). Nevertheless, it might be the case where vehicles are strictly allocated to certain areas thus responding solely within their boundaries, but this is out of the scope of this paper, and we assume all vehicles are available to respond to any location (within a defined urban area) if idle.

The closest idle vehicle policy decides which vehicle to dispatch to an event, ***m***, based on the maximization of a response time metric at the individual level, i.e., per case basis:

$$\text{maximize } P_m^r = -\sum_j r_{ij} \times v_j \qquad (18)$$

However, the closest idle vehicle might not be the best solution when medical emergency heterogeneity and survival are at stake, nor if the measurements used at dispatching time differ from real-time conditions. This allows for two improvements: (1) use of real-time drivability data; (2) account for the victims' heterogeneity and survival under dynamic environments, which implies that drivability conditions, the victims' location, and typology differs in both time and spatial dimensions.

Let us assume that a technological improvement allows the system to access and collect traffic (or any drivability metrics) and demand (EMS calls or any demand metric) information from the urban area where it is implemented. One can define a drivability matrix $\boldsymbol{R}$, where each item $\boldsymbol{r^s_{ij}}$ represents the drivability metric (e.g., travel time) between demand point $\boldsymbol{i}$ and vehicle station $\boldsymbol{j}$ for traffic conditions at period $\boldsymbol{s}$, and a matrix of demand metric $\boldsymbol{D}$, where each item $\boldsymbol{d^{es}_i}$ represents the demand metric for emergency of type $\boldsymbol{e}$ for period $\boldsymbol{s}$ at demand point $\boldsymbol{i}$.

We define the system Survival Status $\boldsymbol{S^s}$ using the survival metric, equation (16), the concept of station busy fraction $\boldsymbol{q_j}$ (Daskin, 1983) and the expected response time as seen in Berg et al. (2016) based on Snyder and Daskin (2005), as:

$$S^s = \sum_i \sum_j \sum_k \sum_e d_i^{es} \left(e^{c^e \times r_{ij}^s} \right)^{-1} \times \left(1 - q_j^s\right) \times q_{k-1}^s \times z_{kij} \tag{19}$$

Where:

$$q_j^s = \frac{\sum_{i \in T_j^s} \sum_e d_i^{es} \times w}{a_j^s} \text{ and } q_{k-1}^s = \prod_{j=1}^{k-1} q_j \text{ ,} \tag{20}$$

d_i^{es} is the demand metric for events of type ***e*** during the period ***s*** at demand point ***i***,

w is a constant that converts ***d_i^{es}*** to the number of vehicles requests within the next period,

a_j is the number of idle vehicles located at station ***j***,

z_{kij} is 1 if ***j*** is the ***k^{th}*** nearest station of ***i***, and zero otherwise, and

T_j^s is a subset of demand points ***$i \in I$*** indexed to station ***j*** for which ***z_{1ij}*** = 1 during ***s*** and is defined as the subset of demand points ***i*** which can be reached by any vehicle at station ***h*** quickest than any other vehicle located at a station ***$i \in I$*** during period ***s***, i.e. ***$i \in T_j^s \ni r^s_{ih} \leq r^s_{ij} \ \forall j \in J \setminus \{k\}$***. Therefore, ***$q_j^s$*** is the busy fraction of station ***j*** at period ***s*** consequence of the demand for emergency vehicles within its coverage area, ***T_j^s***.

Contrary to the station location problem where ***z_{kij}*** is a decision variable, in the dispatching problem the station locations are already defined, thus ***z_{kij}*** is known. Consequently, ***S^s*** can be simplified as:

$$S^s = \sum_i \sum_j \sum_e d_i^{es} \left(e^{c^e \times r_{ij}^s} \right)^{-1} \times P\left(j \rightarrow i\right)^s \tag{21}$$

Here, ***$P(j \rightarrow i)^s$*** is the probability of demand point ***i*** be served by the ***k***-nearest station ***j*** at period ***s*** and depends on the busy fraction, ***p***, of each ***g***-nearest station, with ***g*** < ***k***. This probability can be defined as:

$$P(j \rightarrow i)^s = \prod_{g \epsilon G}^{k-1} q_g^s \times (1 - q_k^s) = q_{1^{st}}^s \times q_{2^{nd}}^s \times \ldots \times (1 - q_{k^{th}}^s) \tag{22}$$

Where:

q_g^s is the busy faction of the ***g***-nearest station of ***i*** at period ***s***,

g is a position of a station in the set ***G^i***,

G^i is an ordered g-nearest station to demand point i, and

k^{th} is the k position of station j in set G^i.

From the previous performance function which measures the Survival Status of the EMS dispatching network and with the use of the survival function to measure the performance of a singular event, the decision of dispatching a vehicle from a station j to a certain event can be scored as Q_j:

$$Q_j = s^e + S^{s-j} \tag{23}$$

$$s^e = \left(e^{c^e \times r_{ij}^s}\right)^{-1} \tag{24}$$

Where $S^{s\text{-}j}$ is the Survival Status of the system after a vehicle from j is dispatched to i resulting in a predicted survival S^e for the responded event e, and making $a_j = a_j - 1$. The dispatching rule for this approach can be simplified as the decision that maximizes the remainder Survival Status assuming that an emergency medical event is always covered by a vehicle, i.e., the solution that maximizes Q_j is also the same that maximizes $S^{s\text{-}j}$ because:

- We assume that for any life-threatening event the closest vehicle is dispatched;
- The dispatching rule that maximizes Survival Status only activates for non-life-threatening events;
- A non-life-threatening event never puts survival at risk.

Therefore, $Q_j = 1 + S^{s\text{-}j}$, for any r^s_{ij}, hence argmax(Q_j) = argmax($S^{s\text{-}j}$).

Accordingly, we propose an intelligent real-time dispatching rule which can benefit from real-time drivability information, EMS demand predictors and the proposed Survival Status metric, and relies on the following optimization problem, equations (25) (26) (27) (28), where decision variable v_j identifies the station that will dispatch a vehicle:

$$\text{maximize } S^{s-j} = \sum_i \sum_j \sum_e d_i^{es} \left(e^{c^e \times r_{ij}^s}\right)^{-1} \times P\left(j \to i\right)^s \tag{25}$$

$$\sum_j v_j = 1 \tag{26}$$

$$a_j = a'_j - v_j \; \forall \, j \in J \tag{27}$$

$$v_j \in \{0, 1\} \tag{28}$$

$$a_j \geq v_j \; \forall \, j \in J \tag{29}$$

$$P\left(j \rightarrow i\right)^s = \prod_{g \in G}^{k-1} q_g^s \times (1 - q_k^s) \quad \text{and} \quad q_j^s = \frac{\sum_{i \in T_j^s} \sum_e d_i^{es} \times w}{a_j^s} \tag{30}$$

Where:

a_j is the number of vehicles at station ***j*** after the dispatching decision,

a'_j is the number of vehicles at station ***j*** before the dispatching decision,

v_j is de decision variable that takes 1 if a vehicle is dispatched from station ***j***.

The intelligent real-time vehicle dispatching algorithm (IRTADA) makes three assumptions:

- (1) the system performance can be measured using survival functions as a performance metric;
- (2) a real-time drivability metric is available and/or there is a history of travel times for different day periods, ***r_{ij}^s***;
- (3) The algorithm receives the EMS calls in real-time and there is an EMS demand indicator, ***d^{es}_i***, that specifies how the demand will behave in the next hours.

To solve IRTADA we propose a heuristic that searches the full space of feasible solutions, Algorithm 2. In step 1 the algorithm interacts through the list of possible stations and sorts them by closeness to the event site. In this process, it uses a function *travelTime(o → d)* that calculates the expected travel time between two points. In practice, this reflects a technological improvement that can be: (1) a historical database with drivability metrics for different periods of the day; (2) online platforms such as *google maps* or an *in-house* system that gives correct travel times for the real traffic conditions. Step 2 compiles solutions (from which station to dispatch a vehicle) and updates the number of idle vehicles if the current solution is used. Finally, step 3 iterates through the solutions proposed by step 2 and computes the corresponding Survival Status remainder. Step 4 chooses a vehicle from the station that provides the highest Survival Status remainder.

Algorithm 2 Real-time intelligent dispatching greedy algorithm

Definitions:

i ϵ ***I*** *"vector containing all demand points* ***i****"*

j = [id, time, #vehicles] ϵ ***J*** *"object representing vehicle stations with labels id, time and number of idle vehicles"*

J' = Ø *"auxiliary vector of stations* ***j*** *when vehicles from station* ***j'*** *is dispatched"*

g = [***J'***, ***j'***] *"vector holding* ***J'*** *and the corresponding possible dispatching solution"*

G = Ø *"vector of possible solutions: a solution is a vector containing stations **J'** with label updated if a vehicle from station **j'** will be dispatched"*
travelTime(o → d) "is a function that calculates real-time travel time between o and d"
append(object) "*is a function that adds an argument to a vector"*

1. *For* ***j*** in ***J***:
   r_{ij} = *travelTime(**i** → **j**)*
   a_j = *count(* idle vehicles allocated to ***j)***
   append(***j***[id, time = r_{ij}, #vehicles = a_j] → ***J'***)
   *sort **J'** by time in ascending order*
2. *for* ***j*** in ***J***':
   if $a_j \neq 0$:
   G' = Ø
   for ***j'*** in ***J***:
   if ***j'*** = ***j***: *append(* ***j'***[id, time, a_j – 1] → ***G')***
   else: *append(* ***j'***[id, time, a_j] → ***G')***
   append(***G'*** → ***G)***
3. *for each* ***g*** in ***G*** *calculate*:
   Q_j
4. chose vehicle from station ***j*** corresponding to solution ***g*** with the higher Q_g

8.4. CALCULATIONS

8.4.1. CASE STUDY

To further validate our hypothesis and support our claim, we apply the proposed framework to a real-life case that comprises of one year of EMS calls dataset from Porto city. This dataset contains all the medical emergencies that took place between the 10th of May of 2012 and the 10th of May of 2013 with information on the event occurrence time, the type of medical emergency, the priority assigned by the INEM proprietary algorithm and the address of the occurrence (which was geocoded afterward). For the drivability metric r_{ij}^s, we programmed an algorithm to collect, from *google maps* API, real travel times for morning, afternoon, night, workdays and weekend traffic peak.

Survival functions have been well defined for cardiac arrests, however little is known about survival decays for a wide range of medical emergency typologies. Therefore, we approach the problem by dividing emergencies into two types, the ones where the victims' survival is at stake (Type 1), and the ones where it is not (Type 2).

Several methods can be used to define the demand metric d^{es}_i. As a first approach to this subject, a statistical approach is favored because it simplifies the understanding of the presented concept and dispatching rule, e.g., estimating the number of events for a certain period $\boldsymbol{s}$ and calculating the probabilities of an event taking place at demand point $\boldsymbol{i}$. Thus $\boldsymbol{d^{es}_i}$ can be decomposed as $\boldsymbol{d^{es} * d^{s}_i}$, where $\boldsymbol{d^{es}}$ is a metric that reflects the fraction of events of type $\boldsymbol{e}$ during period $\boldsymbol{s}$, and $\boldsymbol{d^{s}_i}$ is a metric that reflects the contribution of location $\boldsymbol{i}$ during period $\boldsymbol{s}$ for $\boldsymbol{d^{es}_i}$.

For type 1 we assume that our algorithm always sends the closest vehicle. For type 2 emergencies we assume that survival is not at stake, thus a farther vehicle can be dispatched without interfering with the victims' survival, as explained in the dispatching policies chapter. This means that equation (23) is simplified to $\boldsymbol{Q_g} = \boldsymbol{S^{s'}}$ and $\boldsymbol{c^e}$ simplified to $\boldsymbol{c}$, which assumes that all emergencies of type 1 follow a similar survival decay parameter.

Using *Python* and the collected data, a simulation was programmed, and our algorithm implemented. This simulation model runs through a timeline and activates events as they happen. When an event is activated a protocol is triggered and a dispatching rule implemented. The model uses traffic conditions collected from *google maps*, tracks the location of vehicles and their status (either operating mode or idle mode) and for each event captures the response delays (in case no vehicle is available) and the response time.

Four dispatching strategies are compared for the study case:

- Strategy 1: Classic dispatching policy, CDP - dispatch of the closest vehicle calculated with no drivability metric available.
- Strategy 2: Classic dispatching policy but assuming the existence of real-time drivability information, RL-CDP.
- Strategy 3: Intelligent real-time vehicle dispatching algorithm, IRTADA – assuming drivability and average demand metrics are available in a sufficient resolution.
- Strategy 4: same as IRTADA but with a maximum response time of 15 minutes, IRTADA15 – this constrains ensures that even non-life-threatening emergencies are responded within a reasonable time.

For the vehicle location, we use a 10 station configuration (see Figure 41), previously calculated to force each demand point to be reachable at least by one station within 8 minutes at any day period (Amorim et al., 2016). We assume that there is double the number of vehicles, 20, and they can be allocated in two different configurations: (Configuration 1) vehicles are homogeneously distributed by the available stations; (Configuration 2) vehicles are concentrated in the city center; this is: stations 108, 48 and 86 have three vehicles each; stations 24, 11 and 10 have one vehicle each; and the remaining stations have two vehicles each. Further literature on the problem of emergency vehicle location can be found in van Essen et al. (2013), Liu et al. (2016) and more recently in Akdoğan et al. (2018). The idea underlying different vehicle configurations is to try to understand how an intelligent vehicle allocation could contribute to a better Survival Status of the EMS system.

The simulation model runs for the different strategies and configurations, and several metrics are computed for comparison. A first analysis focuses on the standard EMS response metrics such as the response time, the number of emergencies responded within a certain threshold and the number of delays. A second analysis focuses on the survival approach, and it is calculated in the form of the survival gain over a base solution. The gain measures the accumulated improvement a strategy can reach when compared with strategy 1 and is calculated regarding survival, survived victims and victims lost. The accumulative gain of survived victims is calculated as the number of victims that survived in a certain strategy but did not survive in strategy 1. A similar method is applied for the accumulated gain in victims lost – victims that did not survive in a certain strategy but did survive in strategy 1. This binary option – survived or not survived – is calculated as a response time threshold (or survival threshold) such that if a vehicle arrives at a life-threatening-event within 8 minutes, the victims are considered saved, otherwise it is assumed that the victim is lost.

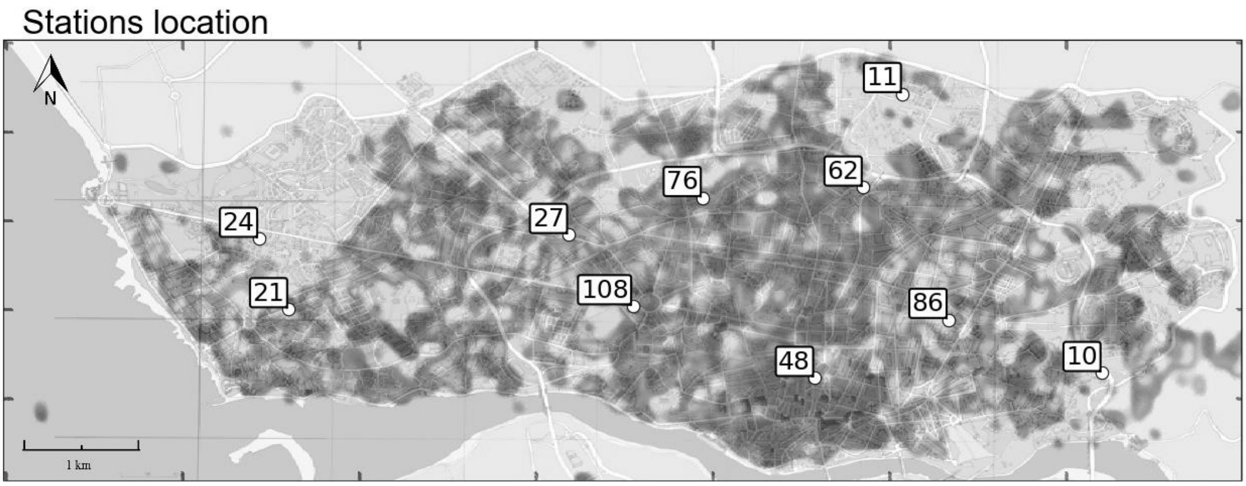


Figure 41. Stations location considered for Porto study case. Darker zones represent areas of higher demand.

8.4.2. Sensitivity analyses

To further conclude on the case study results and better assess the integration of new technologies in the EMS at the tactical level, a sensitivity analysis on the drivability and demand predictors is made. We assume that these technological improvements can provide real-time drivability information and a short-term demand indicator to implement in the proposed dispatching algorithm. Thus, it is important to understand how the accuracy of these two technological improvements will reflect in the EMS dispatching performance.

The accuracy of the drivability predictor is controlled by parameter ε. This parameter introduces an error term in any observed travel time r_{ij}, which measures the time it takes to travel from point i to point j, according to $r_{ij}' = r_{ij} + r_{ij} \times N(\mu, \varepsilon^2/4)$. When the mean is equal to zero and the standard deviation ε/2, a confidence of 95% exists that the travel time between *I* and j has at most an error of $\varepsilon \in [0,1]$.

The accuracy of the demand predictor is measured regarding how well it predicts the location of a certain event. An error *l* translates the probability of the predicted event location differs from the observed location. Accordingly, the location *i* of event *e* is $e_i = P(g(Es), e_i, l, 1 - l)$. Here, *g(Es)* is a function that chooses a location from a bag of possible locations weighted according to the demand observed in the past, and $P(e_1, e_2, w_1, w_2)$ is a function that picks either e_1 or e_2 with probability w_1 and w_2, respectively, where $w_1 + w_2 = 1$ must hold.

Different scenarios are tested using diverse values of ε and *l* to assess the impact of these errors in during the vehicle dispatching operation. Both average time and average survival performance metrics will be used for this assessment.

8.5. RESULTS AND DISCUSSION

8.5.1. APPLICATION TO PORTO CITY

When analyzing EMS response, researcher and practitioners focus on the average response time or on the number of calls that are answered within a chosen threshold. However, as we have been showing, the same response time in different emergencies typologies can have different consequences. Table 13 and Figure 42 shows the classical performance metrics statistics for the different strategies and network configurations.

Smaller average response is achieved when we use the configuration that concentrates vehicles in the denser urban areas (city center). Moreover, the use of real-time drivability information improves the overall response time, as visible when one compares strategy 1 with strategy 2.

Nevertheless, it is evident and expected that the proposed intelligent strategies, IRTADA and IRTADA-15, have higher response times. This is part of their characteristics because they sacrifice the response time to non-life-threatening emergencies to offer a faster response to life-threatening emergencies. If we take a threshold of, e.g. 12 minutes, we see that both strategies 3 and 4 have around 8% and 7% of uncovered emergencies for configuration 1, and 7% and 6% for configuration 2. However, for a threshold of 15 minutes, strategy 4 performs better than all other strategies. The reason is twofold: one, the maximization of the survival status will ensure that vehicles are available close to the more demanding areas; two, the response time constraint will ensure that, if available, a vehicle within a maximum of 15 minutes' driving distance will respond to non-life-threatening events.

These initial results show how the introduction of real-time information benefits the EMS response. However, a change in the dispatching policy towards a survival focus policy debilitates the average response time of the system. Nevertheless, the goal of such policy (i.e., IRTADA) is to augment victims' survival and for that, the analysis needs to focus on survival performance metrics.

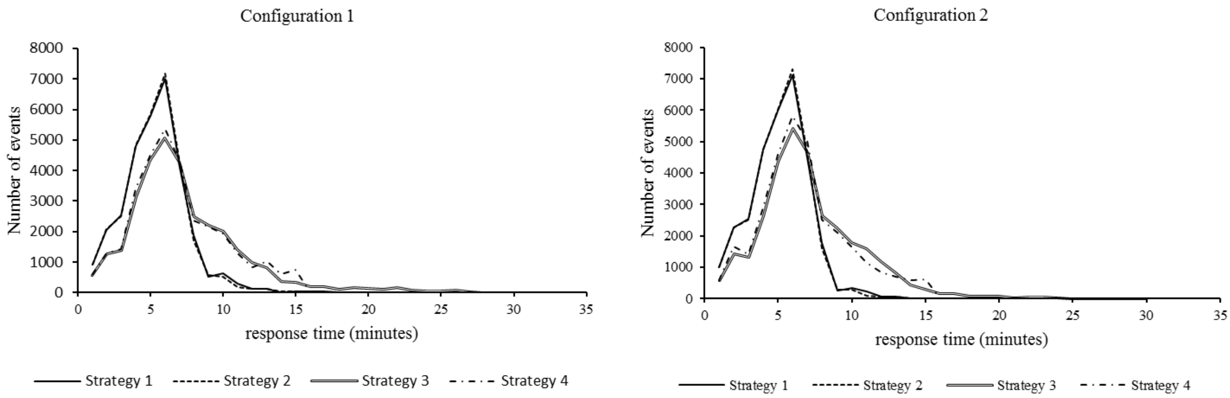


Figure 42. Number of events per response time for the different configurations and strategies.

Table 13. Response time results for the different strategies and configurations.

		time in minutes (Configuration 1)				time in minutes (Configuration 2)			
		Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 1	Strategy 2	Strategy 3	Strategy 4
Response time	Average	4.50	4.45	6.60	6.15	4.36	4.32	6.45	6.02
	Max	58.22	62.03	56.17	51.02	55.60	58.03	49.97	54.17
	Std	2.51	2.43	4.14	3.41	2.33	2.28	3.67	3.24
% over threashold	>8 min	5.45%	4.76%	28.23%	25.66%	3.08%	2.53%	27.35%	22.71%
	>12 min	0.80%	0.62%	8.54%	7.19%	0.41%	0.32%	7.05%	5.75%
	>15 min	0.30%	0.10%	3.99%	0.09%	0.09%	0.04%	2.35%	0.04%
	>20 min	0.03%	0.03%	1.63%	0.04%	0.01%	0.01%	0.53%	0.01%
delays	Total number	10	10	18	45	7	7	12	21
	Average	9.80	9.80	7.56	7.07	9.57	9.57	8.83	7.05
	Max	48.00	48.00	42.00	44.00	38.00	38.00	38.00	41.00

Figure 43 three analyses are presented. A first analysis computes the survival gain of the different strategies when compared with strategy 1. Here, a substantial survival gain is visible when both IRTADA and IRTADA15 (strategies 3 and 4) are in use. For configuration 1, these strategies achieve a survivability rate that can be approximately converted to 60 and 80 saved victims, respectively. For configuration 2 the gain is lower, but still, both strategies achieve around 50 saved victims. Strategy 2, RT-CDP, has a slight improvement of ten saved victims for configuration 1 and five for configuration 2.

A final performance metric uses an 8 minutes' time threshold to count how many victims (life-threatening events) survived in strategies 2, 3 and 4 but did not in strategy 1. The opposite is also analyzed resulting in the number of victims that were lost in strategies 2, 3 and 4 but not in strategy 1.

In configuration 1 the number of survived victims reaches almost 200 in both strategies 3 and 4. When a non-homogenous configuration is used (Configuration 2) the gain is lower and achieves the 100 mark. Even the simple fact that real travel-time is used, strategy 2, guarantees values between of 40 and 50 victims that survived the medical emergency.

In contrast, strategies 2, 3 and 4 lead to a loss of around 10 to 30 victims in configuration 1 and 10 in configuration 2. This loss is due to the use of an average demand metric in the Survival Status function. This metric provides the algorithm with an estimation of what is expected to happen in the next hours. It is however expected that if a more representative metric of the demand fluctuations is used the verified gains tend to increase. Nevertheless, in Amorim et al. (2017) it was shown how certain land use, social and demographic variables correlate with EMS demand. The available data in smart cities can provide a base to define more representative and dynamic demand indicators allowing IRTADA to increase its performance.

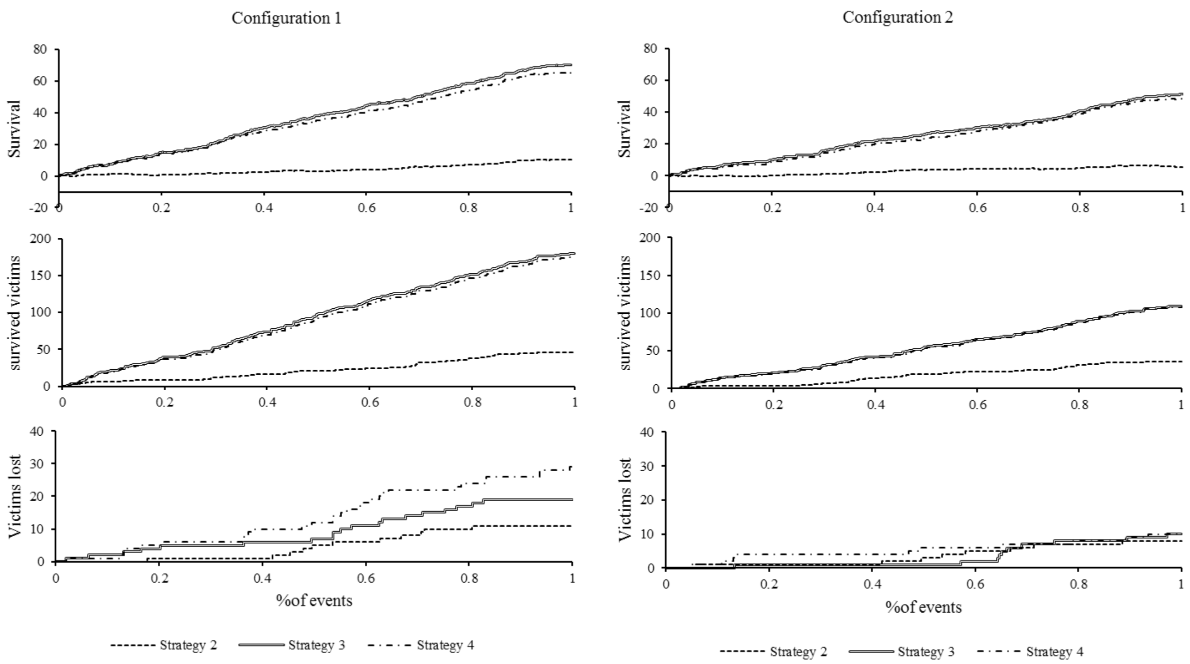


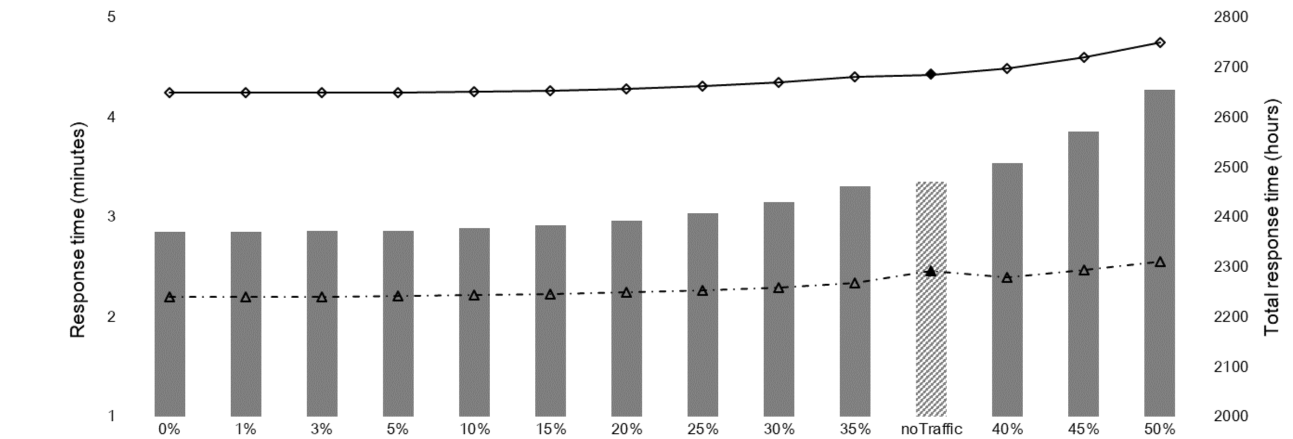
Figure 43. Strategies 2, 3 and 4 gain and lost compared with strategy 1 for the different configurations.

8.5.2. TECHNOLOGICAL IMPROVEMENT IMPACT: SENSITIVITY ANALYSIS

To assess the possible uEMS response improvement after the implementation of a drivability predictor, such as real-time traffic monitoring or short-term prediction models (Zhang, 2014), we assume that the online travel-time, from a station to a uEMS demand point, can be estimated and has at most an error of *e*. We tested several scenarios with error *e* ranging from 0% to 50%. Each of these scenarios represents a uEMS system that dispatches vehicles using the travel-time estimation with the associated error *e*. A scenario “noTraffic” is created using average travel-times to simulate the standard practice, i.e., there is no technological improvement.

Figure 44 shows the average, standard deviation and total response time for different scenarios. It is observed that the average response time almost maintains its value up to an error of 20%, slowly increasing to 0.51 minutes slower for an error of 50%.

The standard practice is equivalent to an error between 35% and 40%. However, the standard deviation shows a higher instability of the standard practice scenario average response compared to the scenarios with similar average response performance.



	0%	1%	3%	5%	10%	15%	20%	25%	30%	35%	noTraffic	40%	45%	50%
Total	2371	2371	2371	2372	2377	2383	2393	2408	2429	2461	2471	2508	2572	2655
Average	4.24	4.24	4.24	4.25	4.25	4.27	4.28	4.31	4.35	4.40	4.42	4.49	4.60	4.75
Std. Dev.	2.20	2.20	2.20	2.20	2.22	2.23	2.24	2.26	2.29	2.34	2.46	2.40	2.47	2.56

Figure 44. Response performance of the uEMS for different confidence scenarios of a possible traffic predictor, and the scenario where there is no traffic information available (noTraffic).

Furthermore, to assess the possible uEMS response improvement upon the implementation of a demand predictor in the Intelligent real-time vehicle dispatching algorithm, IRTAD, we assume that for the next hours the uEMS calls location and volume, $\boldsymbol{d_i^{es}}$, is known with a precision *l* for the location. This precision, *l*, means that there is a chance *l* that the predicted location is not the actual location but still follows the tendency observed at the occurrence period, i.e., follows the annual demand fraction at demand point *i* and period *s*.

We tested six scenarios with *I* ranging from 0% to 50% and measured the average survival, the metric of interest for this algorithm, and the average response time. Each scenario performance was run for 10 simulations of 10 random periods with a length of 7 days (a week) and the results averaged, Figure 45.

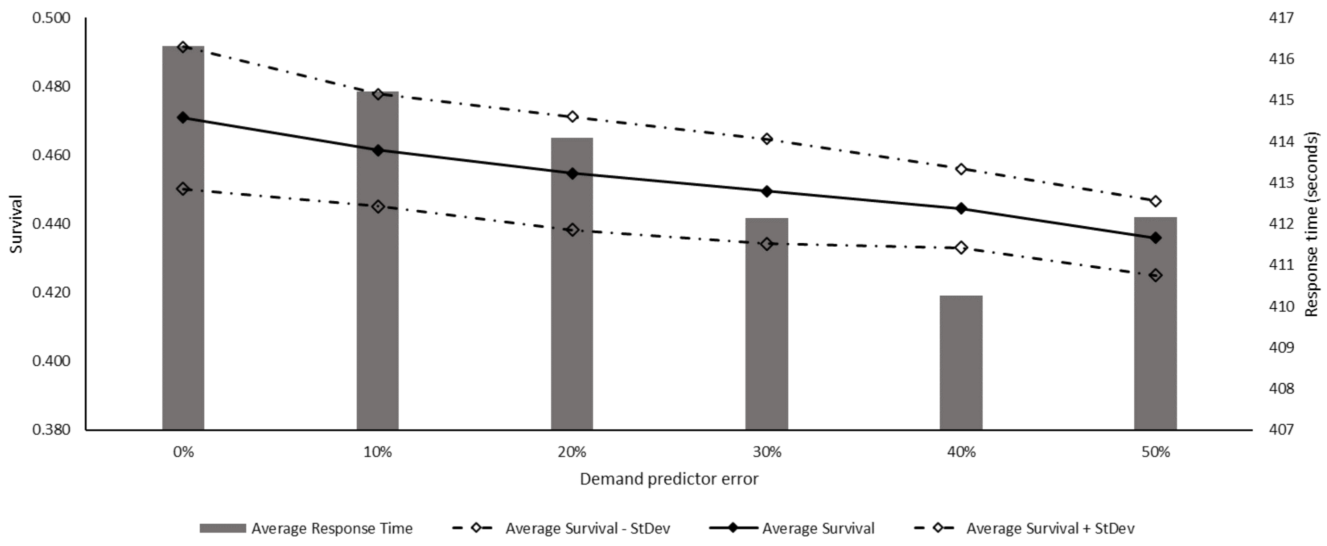


Figure 45. . Response performance of the uEMS for different confidence scenarios of a possible demand predictor.

The results show a clear impact of the predictor accuracy with the survival decreasing 0.03 points when *I* reaches 50%. There is also a decrease in the average response time which shows a performance improvement if this would be the metric of interest. Nevertheless, it is important to remember that the IRTAD aims to reach life-threatening events as fast as possible (i.e. having a vehicle available at the closest station) at the expense of providing a slower response to non-life-threatening. The inaccuracy of the scenario with *I = 40%* shows that, when compared with the scenario *I = 0%*, an average response of +6 seconds can improve the average survival in 0.01 (which can approximately understand as 1 more survival in 100).

8.6. CONCLUSIONS

In this work, we present an intelligent real-time dispatching algorithm with the goal to guarantee the survival of medical emergency victims and the implementation of technological improvements in the uEMS transport system response.

We compare our proposed algorithm with the classic dispatching policy in practice and show how real-time information, survival functions, and intelligent dispatching algorithms can drastically improve victims' survival, leading to a higher number of saved victims. However, the counterpart is an increase in response time for non-life-threatening situations but keeping most of the EMS calls assisted within acceptable thresholds.

Further analysis of the vehicles' location shows that it is advantageous to concentrate the available vehicles on areas with a higher volume of EMS calls. This leads to shorter response times and reduces the survival performance gap between the classic dispatching policy and the intelligent and real-time dispatching policies.

Regarding the technology improvements, the results show that even a system with errors could outperform the current standard practice. The use of traffic predictors to calculate travel-times increases benefit both average response and average survival performance while the implementation of intelligent dispatching algorithms that favor life-threatening events increases victims' survival at a small cost of response time to non-life-threatening events.

Lower response times and the victims' survival are two different approaches in EMS vehicle dispatching. A system with lower response time or lower response threshold does not necessarily lead to a reduction of deaths. It is important to account for victims' heterogeneity and understand that survival is a complex subject. Nevertheless, even with a simplification of these concepts, one can achieve a better service that equilibrates waiting times and saves victims.

For future research, we point out the study of real-time vehicle reallocation and rerouting using survival functions. Furthermore, machine learning can be used to better model short-term demand predictors, a key to the success of intelligent dispatching algorithms, or even model the vehicle dispatching problem directly. We also stress the fact that effort should be placed in understanding the heterogeneity and survival of EMS victims so that a more accurate metric can be used when assessing uEMS response systems.

8.7. ACKNOWLEDGMENTS

We acknowledge the support of FCT (Portuguese national funding agency for science, research, and technology) under the grant PD/BD/52355/2013 during the development of this work. We also express gratitude to INEM (Portuguese Institute of Medical Emergency) for providing us with insightful information and extensive data.

8.8. REFERENCES

Amorim, M., Ferreira, S. & Couto, A. 2017. Road safety and the urban emergency medical service (uems): Strategy station location. Journal of Transport & Health.

Bandara, D., Mayorga, M. E. & Mclay, L. A. 2014. Priority dispatching strategies for ems systems. Journal of the Operational Research Society, 65, 572-587.

Berg, P. L. V. D., Essen, J. T. V. & Harderwijk, E. J. Comparison of static ambulance location models. 2016 3rd International Conference on Logistics Operations Management (GOL), 23-25 May 2016 2016. 1-10.

Blackwell, T. H. & Kaufman, J. S. 2002. Response time effectiveness:Comparison of response time and survival in an urban emergency medical services system. Academic Emergency Medicine, 9, 288-295.

Carter, G. M., Chaiken, J. M. & Ignall, E. 1972. Response areas for two emergency units. Operations Research, 20, 571-594.

Church, R. & Velle, C. R. 1974. The maximal covering location problem. Papers in Regional Science, 32, 101-118.

Curtis, J. P., Portnay, E. L., Wang, Y., Mcnamara, R. L., Herrin, J., Bradley, E. H., Magid, D. J., Blaney, M. E., Canto, J. G. & Krumholz, H. M. 2006. The pre-hospital electrocardiogram and time to reperfusion in patients with acute myocardial infarction, 2000–2002: Findings from the national registry of myocardial infarction-4. Journal of the American College of Cardiology, 47, 1544-1552.

Daskin, M. S. 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. Transportation Science, 17, 48-70.

Daskin, M. S. & Stern, E. H. 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. Transportation Science, 15, 137-152.

Erkut, E., Ingolfsson, A. & Erdogan, G. 2008. Ambulance location for maximum survival. Naval Research Logistics, 55, 42-58.

Ferrucci, F. & Bock, S. 2014. Real-time control of express pickup and delivery processes in a dynamic environment. Transportation Research Part B: Methodological, 63, 1-14.

Gaspoz, J. M., Unger, P. F., Urban, P., Chevrolet, J. C., Rutishauser, W., Lovis, C., Goldman, L., Héliot, C., Séchaud, L., Mischler, S. & Waldvogel, F. A. 1996. Impact of a public campaign on pre-hospital delay in patients reporting chest pain. Heart, 76, 150-155.

Haghani, A., Hu, H. & Tian, Q. An optimization model for real-time emergency vehicle dispatching and routing. 2003. Citeseer.

Haghani, A. & Yang, S. 2007. Real-time emergency response fleet deployment: Concepts, systems, simulation & case studies. In: Zeimpekis, V., Tarantilis, C. D., Giaglis, G. M. & Minis, I. (eds.) Dynamic fleet management: Concepts, systems, algorithms & case studies. Boston, MA: Springer US.

Hogan, K. & Revelle, C. 1986. Concepts and applications of backup coverage. Management Science, 32, 1434-1444.

Hogg, J. M. 1968. The siting of fire stations. Journal of the Operational Research Society, 19, 275-287.

Iannoni, A. P., Morabito, R. & Saydam, C. 2009. An optimization approach for ambulance location and the districting of the response segments on highways. European Journal of Operational Research, 195, 528-542.

Jagtenberg, C. J., Bhulai, S. & Van Der Mei, R. D. 2016. Dynamic ambulance dispatching: Is the closest-idle policy always optimal? Health Care Management Science, 1-15.

Jagtenberg, C. J., Van Den Berg, P. L. & Van Der Mei, R. D. 2017. Benchmarking online dispatch algorithms for emergency medical services. European Journal of Operational Research, 258, 715-725.

Kereiakes, D. J., Weaver, W. D., Anderson, J. L., Feldman, T., Gibler, B., Aufderheide, T., Williams, D. O., Martin, L. H., Anderson, L. C., Martin, J. S., Mckendall, G., Sherrid, M., Greenberg, H. & Teichman, S. L. 1990. Time delays in the diagnosis and treatment of acute myocardial infarction: A tale of eight cities report from the pre-hospital study group and the cincinnati heart project. American Heart Journal, 120, 773-780.

Knight, V. A., Harper, P. R. & Smith, L. 2012. Ambulance allocation for maximal survival with heterogeneous outcome measures. Omega, 40, 918-926.

Knopps, L. & Lundgren, T. 2016. Modeling ambulance dispatching rules for ems-systems.

Li, H., Peng, J., Li, S. & Su, C. 2017. Dispatching medical supplies in emergency events via uncertain programming. Journal of Intelligent Manufacturing, 28, 549-558.

Maxwell, M. S., Restrepo, M., Henderson, S. G. & Topaloglu, H. 2010. Approximate dynamic programming for ambulance redeployment. INFORMS Journal on Computing, 22, 266-281.

Mccormack, R. & Coates, G. 2015. 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, 294-309.

Pons, P. T., Haukoos, J. S., Bludworth, W., Cribley, T., Pons, K. A. & Markovchick, V. J. 2005. Paramedic response time: Does it affect patient survival? Academic Emergency Medicine, 12, 594-600.

Restrepo, M., Henderson, S. G. & Topaloglu, H. 2008. Erlang loss models for the static deployment of ambulances. Health Care Management Science, 12, 67-79.

Revelle, C. & Hogan, K. 1989. The maximum availability location problem. Transportation Science, 23, 192-200.

Sánchez-Mangas, R., García-Ferrrer, A., De Juan, A. & Arroyo, A. M. 2010. The probability of death in road traffic accidents. How important is a quick medical response? Accident Analysis & Prevention, 42, 1048-1056.

Savas, E. S. 1969. Simulation and cost-effectiveness analysis of new york's emergency ambulance service. Management Science, 15, B-608-B-627.

Schmid, V. 2012. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. European Journal of Operational Research, 219, 611-621.

Snyder, L. V. & Daskin, M. S. 2005. Reliability models for facility location: The expected failure cost case. Transportation Science, 39, 400-416.

Su, S. & Shih, C.-L. 2003. Modeling an emergency medical services system using computer simulation. International Journal of Medical Informatics, 72, 57-72.

Toregas, C., Swain, R., Revelle, C. & Bergman, L. 1971. The location of emergency service facilities. Operations Research, 19, 1363-1373.

Valenzuela, T. D., Roe, D. J., Nichol, G., Clark, L. L., Spaite, D. W. & Hardman, R. G. 2000. Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos. New England Journal of Medicine, 343, 1206-1209.

Valinsky, D. 1955. Symposium on applications of operations research to urban services—a determination of the optimum location of fire-fighting units in new york city. Journal of the Operations Research Society of America, 3, 494-512.

Yang, S., Hamedi, M. & Haghani, A. 2005. Online dispatching and routing model for emergency vehicles with area coverage constraints. Transportation Research Record: Journal of the Transportation Research Board, 1923, 1-8.

Yue, Y., Marla, L. & Krishnan, R. An efficient simulation-based approach to ambulance fleet allocation and dynamic redeployment. AAAI, 2012.

9. FINAL REMARKS

9.1. GENERALITIES

Emergency Medical Services (EMS) in the urban context rely on a transport system that must accommodate the environment where it operates. In an urban environment, traffic and demand oscillates both in space and time, thus, a static overview when planning EMS transport system might limit its proper operation. Furthermore, victims of medical distress, who require immediate assistance, rely on a quick response of the EMS but each emergency call for aid has its own characteristics, i.e. naturally, the EMS demand covers all types of distress incidents - logically the response time will reflect differently for different types of medical emergencies.

This thesis addressed the above-mentioned aspects of the EMS transport system to provide a better understanding on how planning the system can be improved in the future. We developed a full methodology that allows the selection of performance metrics by initially studying the impact of response time in different medical typologies. These metrics are then integrated in theoretical strategic and tactical planning models to allow for a more realistic comparison of solutions. The methodology uses a simulation model to serve as an experimental tool to obtain empirical proof.

Our results show that the EMS transport system should not be desegregated or tailored for specific medical typologies. Mostly because such analyzes would require detailed performance metrics for each medical typology and the state-of-the-art review shown that for most of the types of medical emergencies it is not possible to offer quantitative metrics. Furthermore, we analyzed the case of road crashes and our conclusions show no clear evidences of substantial gain when segregating the service, i.e. specifically tailor strategic decisions to respond to injuries in road crashes. The several analysis and comparisons of performance metrics shown that EMS authorities should shift their performance objectives from operational to victims' outcome. Even when talking about strategic and tactical decision, as is the case of station location and vehicle allocation, an integrated model for decision-making is closer to the optimal solution. Separating these decisions leads to sub optima. Finally, through simulation, we were able to show how the classical dispatching rule (the one at practice and assumed by Operational Research (OR) researchers) is far from optimal in terms of survival. Empirical proof shows that not only new dispatching rules can improve EMS response performance, but also the adoption of new technologies.

The key findings of the work here developed together with its contributions provide solid ground for future research as well as valuable insight for practitioners.

9.2. KEY FINDINGS

The key findings of this thesis are resumed in the following points:

- The relation between response time and medical emergency victim's survival is not clear. There is an obvious correlation when talking about cardiac arrest, but the literature shows that for other types of medical emergencies these correlations are weak and hard to define.
- With reference to the segregated demand predictors, we shown that in urban areas with higher percentage of population with no economic activity, i.e. elderly, unemployed and under legal age inhabitants, are more likely to request for medical emergency aid. Whilst motorways and urban areas with high traffic volumes are rather demanding medical assistance for road crash related injuries.
- Predicting EMS demand through statistical models is possible but its accuracy drops when we aim for detailed prediction such as the segregation of the medical typology. This fact complicates the creation of a segregated EMS.
- There is no significant difference when strategically planning with or without the intention to better respond to road crashes. The main finding is obvious for this case: emergency vehicle should be parked close to the main road arterials.
- The use of a scenario-based optimization model for strategic decisions significantly improves the EMS performance both in terms of average response time and survival. The use of average or total values when solving location models are not optimal as proven by empirical analysis through simulation.
- Simulation is an interesting tool for solution analysis when real experiments are prohibitive. However, it is resources demanding, specifically, it requires long running times. Metamodels are able to integrate the benefits of simulation with relatively small loss of predictive power but allow for huge amount of solutions analysis.
- The use of a survival performance metric or an operational performance metric has a huge impact on the obtained station locations solution. With the use of more detailed survival functions more tailored solutions can be achieved.
- A non-integrated strategic and tactical planning leads to lower system performance when compared with an integrated planning. Further, an integrated analysis allows for a better understanding on decisions where strategic and tactical resources might collide, e.g. an integrated analysis will allow decision-makers to assess the hypothesis of building new stations together with the hypothesis of buying more vehicles in a unique framework which could make the relation between investment and performance gains even clearer.

- The *classical dispatching rule* is still far from optimal. A simple random assignment of vehicles to emergency calls can outperform the classic rule in terms of survival metrics. The use of survival functions in the dispatching rule to assess system survival status can become a substitute for the classical rule.
- The assumption of city dynamics is more than plausible, specially during strategic decisions. We found that at certain periods of the day, week, month and even year, the best dispatching rule is not always the same.
- The use of traffic predictors and demand predictors to support dispatching decisions increases performance significantly, both operation and social (through higher survival rates). It was found that these predictors do not need to be highly accurate to outperform the state of the art and state of practice.
- Tactical decisions are highly related with the urban area at stake. What might work well in one city might perform lower in another. How land is used and how the urban area is designed drives the models to be used. This can be seen when dispatching rules performance were tested in two distinct case studies, Porto city and San Francisco city.

9.3. CONTRIBUTIONS

Overall, this thesis contributes to the research line of emergency medical service transport system in the urban context by providing a broad analysis on different planning and decision levels. During the development of our research, methodologies and guidelines were produced and disseminated worldwide.

Two main impacts are highlighted. First, the *“Institute of Medical Emergencies of Portugal”* (INEM), after our feedback and continuous share of results widened their view on the dispatching policies and in the implementation of new technologies. Second, the research developed on dispatching policies analysis, presented in TRB annual meeting of 2018, promoted, together with other researchers’ work, high interest within the *National Association of State EMS Officials* (NASEMSO) and lead to the formation of a subcommittee focusing on medical emergency service response and new technology *“Joint Subcommittee for Emergency Response”*.

The thesis contributions to the *state of art and practitioners* are summarized below:

- A framework to assess road crash impact, or other types of medical emergencies, and how to implement EMS performance metrics. The framework was an initial task of this thesis and included a methodology for road crash data linkage from different sources, analysis on different injury measures and finally an approach on how to relate EMS response time with the injury outcomes. The final findings exposed the difficulty of defining clear relationships between traumatic injuries, survival rate and medical response. For these reasons, as previously explained in Chapter 1, the thesis focus shifted to a broader approach. Nevertheless, it is important that these difficulties and findings be registered, and, although they are not part of the main body of this thesis, they were cited when appropriated.

- An important analysis on medical emergency response segregation. We focused on implementation of road crash focused EMS strategic decisions. The framework utilized can be replicated for different urban areas at different stages. Nevertheless, we show evidences that such segregation of the service might not be operational worthy because of the wide range of medical typologies and the randomness underlying, particularly how hard it is to predict or define demand patterns at such detailed scales.
- A complete methodology that allows for a systematic adoption of theoretical models and comparison using several performance metrics. The main advantage of the methodology is to rely on empirical evidences to ensure that simplifications or assumptions defined during model conception correctly translate into practice.
- The thesis gives support to new EMS research directions and assumptions. Particularly, we added relevant proof to the suboptimality of a two-stage decision process for stations and vehicle location. This supports the line of research that integrated strategic and tactical planning decisions. We provided analysis on performance metrics and how survival functions can replace operational performance but still achieve reasonable response standards. Finally, we shown how the classical dispatching rule is far from optimal and proposed a new dispatching rule based on victims' survival that can be upgraded with new technologies.
- We offered several optimization, simulation and dispatching models that can be quickly implemented in practice or for research and that require low computation resources but still provide empirically proven good solutions. These models allow for mass analyses and assessments because of their computation time and simplicity when it comes to their parameters and data requirements.

9.4. FUTURE RESEARCH

The work developed in this thesis presents a continuation of the classical approach made in past and a turn to new approaches and ways to look at the transportation problem in the Emergency Medical Service in the urban context. The different methods and models that were proposed can be further developed or improved in accordance with the topics below:

- Multidisciplinary research becomes fundamental in the development and investigation of high complex problems and systems. On one hand, the topic that we investigated can be supported by the expertise of the medical and emergency medicine research communities by focusing in the development of more detailed survival functions for the different medical emergency topologies. On the other hand, it becomes more and more important to elaborate how this evidence-based but still rather theoretical research which - by using new data processing approaches - essentially helps to extend the set of available decisions, can be better absorbed practically by policy makers and implementers (i.e., by city hall planning departments, emergency services, hospitals) or as integral part of public policies in general.

- The location models that were proposed can be assessed or improved to allow for reallocation of vehicles during the day. However, this is a very sensitive topic because when we aim to provide useful decision tools that are practice ready, we need to do it accordingly to what is realistic. Dynamic reallocation of EMS resources through the day requires that the vehicle personal be also reallocated which might not be possible. In cooperation with emergency medicine experts, boundaries and guideless should be defined to provide realistic boundaries to the reallocation problem.
- Integrated strategic and tactical models should further be developed according to the proposed methodology. Several simplifications were assumed to allow fast computational runtimes. Nevertheless, there is still a lot of space for improvement, for example by merging classical approaches and work on algorithms that are able to solve, in reasonable time, the extra complexity that the integrated approaches bring.
- A dispatching algorithm was proposed which allows for the implementation of demand predictors. With the development of machine learning techniques, particularly deep neural networks, the proposed algorithm can be improved with better predictive models both for demand and travel times.

These are the main topics where future research can focus and that were shown to have potential through the work we developed.