Optimizing the Network of the Emergency Medical Service in Hanoi

Rui Guilherme Machado Soares

Master's Dissertation

Supervisor: Prof. Carlos Bragança

U. PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Mecânica

2016-07-04

Otimizar a rede do Serviço de Emergência Médica em Hanói

Resumo

Nesta dissertação pretendeu-se analisar e melhorar a rede do Serviço de Emergência Médica em Hanói. Isto implica a alocação de ambulâncias a estações base, a fim de melhorar o nível de serviço. Este é um problema com que a área da Investigação Operacional lida.

A primeira tarefa foi compreender as operações típicas que o SEM envolve e como este está organizado em Hanói. Constatou-se que dispõe de 23 veículos e são operadas 5 estações próprias.

Uma estratégia envolvendo duas técnicas de IO, modelação matemática e simulação computacional, foi utilizada na procura pela melhor configuração para servir a população. No entanto, a prioridade constava em definir os pontos da rede que seriam utilizados como dados de entrada dos modelos, estes representam os postos de socorro e a distribuição da procura.

A fim de aumentar a cobertura, um plano de cooperação foi explorado, e 34 instalações médicas na província foram selecionadas para serem consideradas como possíveis locais base de ambulâncias.

Devido à falta de dados relativos ao histórico de chamadas de emergência, a procura foi modelada com base nos dados demográficos da região. Isto introduziu fontes de erro, o que, aliado à aleatoriedade intrínseca do SEM, levou à decisão de optar pela formulação de vários modelos matemáticos simples que serviram como pontos de partida para as várias iterações simuladas.

O modelo de simulação foi construído no software *AnyLogic* e foi implementada uma interface de GIS para visualizar as operações e fazer os veículos seguir rotas precisas. A distribuição geográfica e taxa de chamadas de emergência foi baseada em mapas de população obtidos na plataforma *ArcGIS* e informações fornecidas pelo SEM de Hanói.

O principal parâmetro de desempenho registado para cada iteração foi o Tempo de Resposta – período desde o pedido de emergência até que se alcance a vítima – uma vez que é o mais fortemente correlacionado com a sobrevivência da mesma. Além disso, as taxas de utilização das estações foram analisadas de forma a orientar as realocações de veículos para as novas iterações.

A melhor configuração obtida dispondo apenas dos recursos originais mostrou uma redução em metade do tempo médio de espera da vítima – de 20 para 10 minutos.

Várias iterações foram realizadas com a adição de ambulâncias em zonas rurais, até um máximo de 5, a fim de reduzir os longos TRs experienciados. Com 28 veículos, 90% da população foi servida em menos de 16 minutos em casos de emergência, 20% menos tempo do que era possível com os recursos originais, para a mesma porção.

Ambas as configurações se mostraram robustas com o aumento da procura visto os seus desempenhos não caírem mais que 30% em qualquer indicador avaliado, mesmo tendo o número de chamadas duplicado.

Abstract

This dissertation meant to analyse and improve the network of the Emergency Medical Service in Hanoi. This entails allocating ambulances to the possible deployment stations in order to better serve the demand, often designated Facility Location Problem.

The first task was to understand the typical operations that EMS involve and how it is organized in Hanoi. It was gathered that they have 23 vehicles and operating in 5 stations.

A two-folded strategy implementing mathematical modelling and computer simulation was used in the pursuance of a better supply layout. However, setting the network nodes to serve as input of the models was necessary first, these represented the ambulance stations and demand distribution.

In order to increase the coverage, a cooperation plan was explored, and 34 medical facilities in the province were selected to be added as possible deployment sites.

As there was lack of data concerning the historic of calls, the demand was modelled based on the demographics of the region. This introduced sources of error, which, allied with the EMS intrinsic stochasticity, led to the decision of opting for the formulation of several simple mathematical models that served as starting points for the iterations with the computer simulation tool.

The simulation model was built in software AnyLogic implementing GIS tools to visualize the operations and follow accurate travel paths. The distribution and rate of emergency calls was based on population maps from the ArcGIS and information provided by the Hanoi EMS.

The main performance metric stored for each iteration was the Response Time, duration reaching the person in need, since it is the one most correlated with survival. Moreover, the utilization rates of the stations were analysed as a mean to guide relocations of vehicles for new iterations.

The best performer design maintaining the original resources proved to reduce the average victim's waiting time in half – from 20 to 10 minutes.

Experiments were conducted with the addition of up to 5 ambulances in rural areas in order to reduce the long RTs experienced there. With 28 vehicles, 90% of the population could be served in under 16 minutes in case of emergency, 20% less than it was possible with the original resources.

Both designs proved robust with increasing demand by underperforming by less than 30% in all metrics, when the rate of calls doubled.

Acknowledgements

I am very grateful to the persons in Hanoi that allowed for this opportunity and helped me along the way.

I would like to express my gratitude to my family and girlfriend, Catarina, for all the support given through these past months.

Finally, I thank to all my friends that shared some hard work moments with me.

Contents

Acronyms

ALS	Advanced Life Support
BLS	Basic Life Support
EMS	Emergency Medical Services
FLP	Facility Location Problem
OR	Operations Research
RT	Response Time
RTT	Round-trip Time

List of figures

List of tables

1 Introduction

1.1 Emergency medical services

Emergency Medical Services are the system that responds to unpredictable health needs of the population. It is clear, then, their importance: they're the first contact a victim has and are responsible for stabilizing its condition and carrying it to a medical facility, if necessary. They save lives.

The service is performed by organizations, usually public, that own a fleet of vehicles, which are dispatched to the incident scene if a help request is received.

As emergency incidents may arise from anywhere and at any moment, EMS operations are difficult to plan due to their stochasticity. Ever growing complex approaches and methods borrowed from engineering and applied mathematics fields are often employed, then, to match the challenge of providing the best possible service level.

1.2 Purpose

Hanoi, the capital of Vietnam, has an estimated population of 7.5 million persons and the emergency response system isn't able to fulfil such demand (General Statistics Office of Vietnam, 2014). In an interview to Thanh Thien News, in 2015, the deputy director of the public EMS admitted it takes over one hour to reach patients sometimes and that it lacks efficiency overall. Furthermore, the country has one of the world's largest rate of road fatalities, which may be a consequence of the poor system and urges for actions being taken in the field (WHO, 2015).

1.3 Objectives

Facing the information of section 1.2, the EMS of Hanoi demonstrated interest in supporting a study on the Logistics of their organization.

The goal of this research was to investigate and improve the network design of the EMS in the province of Hanoi. This entails choosing the facilities in which ambulances may be deployed from and how to allocate them, this challenge is usually called Facility Location Problem in the literature. More often than not, this problem also deals with the demand allocation to each facility though in this case, emergency operations, victims should be served by the fastest way, which depends on the resources availability and can't be predicted.

FLP are considered strategic level planning in most applications, especially if the construction of facilities is intended – it involves large costs and long time-horizons. However, the Hanoi EMS requires a solution in a shorter period and less costly which was the task of this study to achieve.

Overall the main focus was to reach a layout for the network that increased the service level, which meant saving more lives, and was roughly translated into shorter response times. Moreover, it should be of non-expensive implementation and the value of new vehicles added to the fleet was important to be tested.

Lastly, a tool was to be developed and handed to the organization that could help them make more insightful plans and that support future tactical decisions.

1.4 Research strategy

The field of engineering that concerns with problems such as the one described is Operations Research, or, more specifically, Logistics which borrows OR techniques for practical applications. This research made use of such methods to tackle the problem in hands.

Mathematical modelling was first utilized in order to find the optimal distribution of vehicles over a set of possible deployment stations to better serve a group of demand nodes. However, the models created did not translate reality perfectly, which is very difficult given the randomness of the system in hands and the lack of precise starting data, which means the solution might not hold the best in practice as well.

For that very reason, simulation was brought up as a tool that is able to mimic operations of the EMS more accurately and ultimately point out the fittest design by testing several iterations. It was also used to further study changes in the fleet and in demand.

Beside these main areas of study, other work that was necessary involved the extensive analysis of the organization and the area, by researching online information and conducting interviews.

1.5 Thesis overview

The next chapter sums most of the literature written on the EMS field tackled with OR techniques, both mathematical modelling and simulation tools (section 2).

Afterwards, the text investigates deeper into the EMS operations in general, following a view on the Hanoi's organization in particular. This raises the problems that it started with and sets improvements to chase (section 3).

The 4th chapter describes the first part of the mentioned two-folded research method: the mathematical modelling. It includes the gathering of data to serve as input and the actual solving.

The second part of the solution follows, the simulation model created is described in section 5 and different scenarios are tested in that environment. After being analysed and compared, further vehicles relocations and additions are attempted, based on the collected data, in order to reach better designs.

The thesis wraps up with some remarks on the results obtained and recommendations to the EMS of Hanoi.

2 Literature review

In this section, two key areas for this research are reviewed: the mathematical modelling for optimization and computer simulation. Although the applications of these tools extend across many fields, mostly EMS cases are focused in this text for their similarities with the present study.

2.1 Mathematical models

In the last 30 years, the literature in FLP applied to EMS has been keeping up with advancements in computer technology and algorithms. This is observed has models evolved from static and deterministic to dynamic and stochastic, which means they have into account time and probabilities. As these later ones are more complex, only the appearance of digital tools allowed reaching solutions in practical duration (Brotcorne *et al*, 2003).

The models presented here are all discrete, this kind is the most suited to the problem in question since the facilities are restricted to a set and demand appears in nodes of a network. Other types of location models include analytic, continuous and network models, these differ from the first type in terms of demand distribution or facilities' location options and are not mentioned further (Daskin, 2008).

2.1.1 P-centre models

In short, P-centre models aim at minimizing the maximum distance between any demand node and its closest supply point, this explains why another term for this kind of model is "minmax". In practical terms, it improves the worst link in a network. The number of facilities here is predetermined, as a constraint.

The centre problem dates back to Sylvester (1857), called the smallest circle problem then, when it was first proposed in a continuous space. It searched for the location of the centre of a circle, one facility, which had the shortest possible radius and involved a set of dots, the demand points. In more recent times, the model lent itself well to the planning of networks in EMS and other industries.

Garfinkel *et al* (1977) solved the problem successfully, modelled in integer programming, and further explored its properties. Later, ReVelle and Hogan (1989b) applied it to EMS, in a model that had in consideration the probability of ambulances being busy. The model minimized the maximum distance for a certain level of reliability of service, this depended on how busy they were.

A case study with emergency helicopters in the Alps was introduced by Talmar (2002), it reached a solution making use of heuristics and minimized the slowest time to respond to victims.

More approaches using this model were developed besides these and some examples are Brandeau *et al* (1995), Daskin (2000) and Current *et al* (2004).

2.1.2 P-median models

Unlike the previous kind of model, the P-median tackles the overall performance of a network, rather than its weakest link. It minimizes the average distance between demand nodes and supply points, with a constraint on the number of facilities. This was introduced by Hakimi (1964) and translated into linear programming by ReVelle and Swain (1970), who successfully found a solution method for it.

The median problem had a more extensive use than the centre one since, in practical cases, it is usually more important to improve the whole system than to soften the worst performance.

Carbone (1974) applied the model to day-care centres, the goal was to minimize the distance covered by the users to reach them. It was further improved from a deterministic problem to a probabilistic one in order to contemplate the randomness of demand.

A case study in Carbondale, IL, in the USA, was researched by Paluzzi (2004) and dealt with the fire department. It attempted to locate a station so that the average distance to the different demand points was minimal. This study proved the model usefulness since the solution matched others from different methods.

Focusing on EMS applications, which is an area that suits well this model, Berlin *et al* (1976) first used it with three objectives: minimize the average distances between hospitals and demand, from ambulance stations to demand and from stations to hospitals. Mandell (1998) later studied the problem with different vehicle types, which implied different priorities in calls. A more advanced research, considering dynamic allocation of ambulances, was proposed by Carson and Batta (1990). The model tested different scenarios in which the vehicles were positioned to minimize average response time.

2.1.3 Covering models

The most widely utilized models in EMS literature are the ones that address coverage. This concept means that a demand node has an emergency station close enough such that a rescue vehicle would take less time to reach it than a predefined Time Standard or Coverage Standard (it is considered covered if so).

Toregas *et al* (1971) introduced the Location Set Covering Model (LSCM) which ensures coverage for every demand node and minimizes the number of facilities required to accomplish that. Another early model was developed by Church and ReVelle (1974), called Maximal Covering Location Problem (MCLP), which aims at maximizing the amount of demand covered (weighted points) with a limited pre-set number of facilities.

These two formulations represent a major division within the covering kind of models. While the first type serves better as a long-term planning tool to figure out an upper bound of facilities and their allocation, the later regards a very crucial aspect of real organizations – limited resources – and seeks to make the best use out of them. Despite being both very simple, they were the foundations to evolve from and many inputs were added by researchers since, the first advancements were still in the deterministic static domain and were followed by probabilistic or dynamic models.

A weakness of the MCLP tackled early on was the lack of thought given to the availability of vehicles. The model considers a node covered if at least one facility is supplying it, therefore it will never place two ambulances at the same station. The problem is that, once a vehicle is deployed, the whole population relying exclusively on it is left unserved. In order to solve this problem, Schilling *et al* (1979) proposed the Tandem Equipment Allocation Model (TEAM). It considered a demand point covered only if two types of emergency vehicles were supplying it. Hogan and ReVelle (1986) presented a different approach and added an objective to the MCLP – to maximize the demand covered twice. Later, Gendreau *et al* (1997) developed a model using two different time standards, the Double Standard Model (DSM). It stated that all demand must be covered by the longest one and a pre-set percentage of the population should lie within the shortest. The objective function aimed at maximizing the demand covered twice within the shortest coverage standard.

As far as probabilistic models, Daskin (1983) was a pioneer with the Maximum Expected Covering Location Problem (MEXCLP), by considering the probability q of an ambulance not being available. It was designated the busy fraction and was calculated dividing the duration of every demand request by the number of vehicles. These were considered independent from each other. The goal was to maximize the expected covered demand. The success of this model was proved by Fujiwara *et al* (1987) by applying it to Bangkok. A further improvement was made by Repede and Bernardo (1994), who included variations in the travel speeds of vehicles according to different periods of the day, and later by Goldberg *et al* (1990b), who added stochasticity to the travel time.

Still in probabilistic models, ReVelle and Hogan (1989) had a different approach by seeking to maximize the demand covered within a probability set in advance. They formulated two versions of the Maximum Availability Location Problem (MALP I and MALP II). The second is more sophisticated since it accepts a distinct busy fraction for each station while the first assumes the same for all. It is difficult to predict these probabilities, mainly since they depend on the model results, but they may be estimated using several techniques (such as queueing theory), this was the focus of many articles in the literature and required a fairly precise demand forecast.

Another aspect that improves the EMS performance is having changes over time into account. This meant ambulances could relocate to where they were needed the most. On the other hand, a model would have to be solved very frequently to mind changes during operation time. This was made possible by the increase in computational capacity witnessed in recent years. Gendreau *et al* (2001) developed a dynamic model based the previously proposed DSM and adding constraints from dynamic nature such as avoiding: successive redeployments for the same vehicle and long relocation travels. This model was meant to be solved after each emergency call, in order to prepare the network design accordingly. A tabu search heuristic was put forward by the same authors to find solutions quickly.

2.2 Computer simulation

The advantage of simulation tools against other methods such as mathematical modelling or queueing theory is allowing the description of the system in more detail and handling sources of randomness with no need for the simplifying assumptions that the others require (Aboueljinane *et al*, 2013).

A pioneer in this field was Savas (1969), who researched the New York emergency service in order to improve it, since then many innovations were proposed.

Regarding demand data, a key characteristic to model is the arrival rate of calls in time, this is approximated by a Poisson distribution in most papers, which means the inter-arrival time is exponentially distributed. It was suggested by Goldberg et al (1990) since each individual in a bigger population has a small chance of needing the service. Some researchers vary this rate through time or space to mimic demand patterns better.

A second characteristic of demand is its distribution in the map, here there’s a trade-off between accuracy and complexity, depending on the amount of node aggregation. Often in the literature, calls from historic data are merged based on their proximity, in areas and, in the centre of each, a demand point is placed. Although the great majority of case studies did have demand data available, some studies relied on demographic information, by considering population as the estimated demand (Aboueljinane *et al*, 2013).

One last aspect of demand is the priority of calls, due to different urgency levels, and how it affects time delays during operations. In simulation models from the literature, Lubicz and Mielczarek (1987) considered different durations of assistance on-spot and whether there was the need for hospital transportation depending on the severity of each case. For the same reasons, Repede and Bernardo (1994) varied the dispatching time in their model according to the situation.

The dispatching rules are one of the most important elements to model, these refer to the choice of which vehicle to use for a rescue, minding the location of each team and the victim. Most of the literature utilizes the nearest available ambulance, this is the case of Fitzsimmons (1971), Wears and Winton (1993), Inakawa *et al* (2010) or Van Buuren *et al* (2012), among many others. Other, less common, rule assigns the rescue task to the closest station (with available resources), such is what did Iskander (1989). A more elaborated choice, opted by Silva and Pinto (2010), is the rule of lower response vehicle. It selects the ambulance with the lowest estimated response time, this includes busy ones too since they may be expected to become available sooner and be faster than a more distant, and available, one takes to make the rescue.

More complex algorithms were attempted by Gendreau *et al* (2001) and Andersson and Varband (2007), which account for future required relocations and the preparedness of teams, though the simple rules remain the most widely used and have been proven effective.

The travel times influence greatly the simulation outputs, according to Henderson and Mason (2005), thus it is of utmost importance to model them accurately. Older studies assuming deterministic speeds and using algorithms to compute the distances were accomplished by Fitzsimmons (1971) and Lubicz and Mielczarek (1987). In recent years, however, new technology, such has Geographical Information System, made it easier and more accurate to estimate travel times.

Concerning the destination hospital, the chosen one is usually the nearest, it was the case for Fujiwara *et al* (1987), Su and Shih (2002) and Van Buuren *et al* (2012). However, according to Ingolfsson et al (2003), other factors must weight in, such as the available capacity of the medical facilities, services offered or patients preference. This made Ingolfsson *et al* (2003) and Henderson and Mason (2005) estimate empirical distributions from historical data for hospital destinations for different areas.

The last step in building a simulation model must be to validate it, this means assuring it represents the reality accurately. Fitzsimmons (1971) opted for comparing the model with another pre-established one with known results. A very common alternative, taken by Repede and Bernardo (1994) in Louisville, is to match results with empirical data from the real organization, if extensive records exist.

Graphical interfaces were also utilized by Henderson and Mason (2005) to visualized and confirm the correctness of vehicles behaviour and routes. Ingolfsson *et al* (2003) tracked the movement of ambulances likewise.

Other validation techniques include the examination and approval from experts in the EMS system in study or sensitivity analysis which tests if the model outputs behave as predicted facing changes of the input values, as done by Uyeno and Seeberg (1984).

One last distinction between different kinds of computer simulation is important to make, models falls into one of three categories, or in more than one. These are agent based simulation, discrete events and system dynamics. The first two are the most used in EMS cases since they work from the bottom up, as does the real world operations. These methods lend themselves well to capturing the behaviours of individuals, such as victims or rescue teams, that create the macro patterns in the model.

3 EMS operations overview

This section is intended to explain clearer how the Emergency Medical Services work and how their performance is measured, both in general cases and in the specific situation of Hanoi, and take a critical remark on the current state of the organization in study.

3.1 Typical system

As mentioned previously, EMS have the mission of attending unpredictable medical issues with the goal of reducing mortality in the population, by assisting victims and transporting them, thus increasing the recovering chances. For that, they must have a very coordinated plan in place to act fast and flawlessly at each occurrence, it usually involves two types of operations: central and external. Detailed tasks are presented in Figure 3.1.

The former kind occur in a fixed facility and are responsible for receiving incoming calls, evaluating the severity of the victim's condition and deciding the best resources to dispatch, if necessary.

The external operations begin once a vehicle starts to the patient, its team is entitled to aid the person on spot and assure transportation to a medical facility when needed. The choice of this last destination is usually decided centrally though. A final step required is that the vehicle and team return to a station if there are no more emergencies to take care of.

The time between receiving a call and arriving to the victim's location is referred as Response Time, while the period including this plus reaching the hospital is known as Round-trip time.

Every single situation EMS faces is unique which means that in some cases the sequence of tasks does not include all the ones mentioned, for the patient may be fully recovered on the spot, or an ambulance dispatching may not be necessary.

Many systems have different vehicles to deal with varying degrees of illness. For land vehicles there are commonly three types: simple ambulances that transport patients only (predominant in developing countries), Basic Life Support vehicles which are able to stabilize minor conditions and Advanced Life Support ones, equipped with better technology to attend more severe health needs. Helicopters are sometimes part of the organization's fleet, or at least of a cooperating entity's, and are held for remote rescues or very urgent situations.

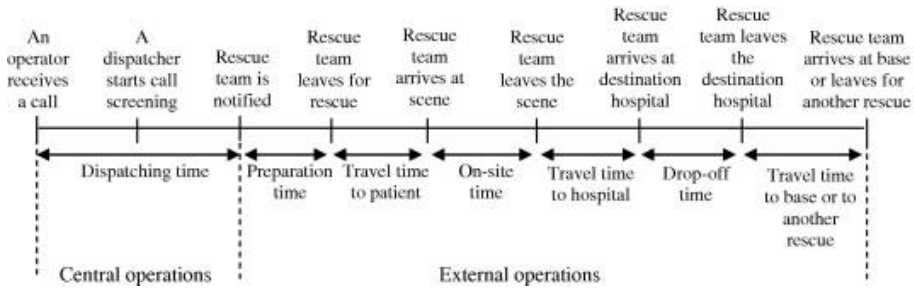

Figure 3.1 - Task sequence of typical EMS in Aboueljinane et al (2013)

Some private hospitals or other entities may operate their own EMS, however, governments maintain a public organization to ensure the availability of this so crucial service. In order to increase efficiency, cooperation between medical bodies and fire departments is common to extend the network and increase coverage and resources. In developed countries, fire stations' vehicles are often the first to arrive.

To ensure the quality of the system, many performance indicators are monitored such as the few presented below:

- Response time (RT): measured from the moment the call was received until the first aid vehicle arrives on scene;
- Round-trip time (RTT): response time plus the period until the victim is delivered at the hospital;
- Dispatching time: moment between receiving call and the team leaving the station;
- Waiting time: duration in queue until a resource is available to answer emergency request;
- Loss ratio: failure to answer demand due to exceeding waiting time limit;
- Survival rate: patients who lived through the reported incident;
- Cost-effectiveness: amount of capital invested for increased performance.

Although the survival rate is the most accurate translation of the EMS goal – save lives – it is difficult to incorporate in theoretical models used for planning. Hence it is often replaced with metrics such as RT and RTT which are very correlated.

To further add complexity to the decision making process in this area, the demand it serves is vast and unpredictable – any individual may need urgent care in any place at any time. Other sources of stochasticity are the operations steps themselves – preparation time, delays in traffic and duration of the medical assistance.

3.2 Hanoi EMS

This research had the contribution of the Emergency Medical Services of Hanoi and the data about their organization and operations was gathered from two interviews with the Vice Director Thanh Khan and from their public website.

In Vietnam, the public EMS is split into provinces, each operating independently from each other. The province of Hanoi is populated by 7,587,000 persons and ranges over an area of 3,329 km^2.

For this demand, they rely on 5 deployment stations (owned and operated by them) and 23 BLS vehicles, distributed as follows:

- Hoan Kiem (HK) – 8;
- Tu Liem (TL) – 5;
- Thanh Tri (TT) – 4;
- Gia Lam (GL) – 3;
- Ha Dong (HD) – 3.

The Hoan Kiem building is the central one and includes the department which receives the calls and makes deployment decisions. Figure 3.2 - Currently operated deployment stations provides a better idea of the area in study and the current distribution of facilities.

These are all the resources available to perform EMS for there’s no cooperation with hospitals or other entities.

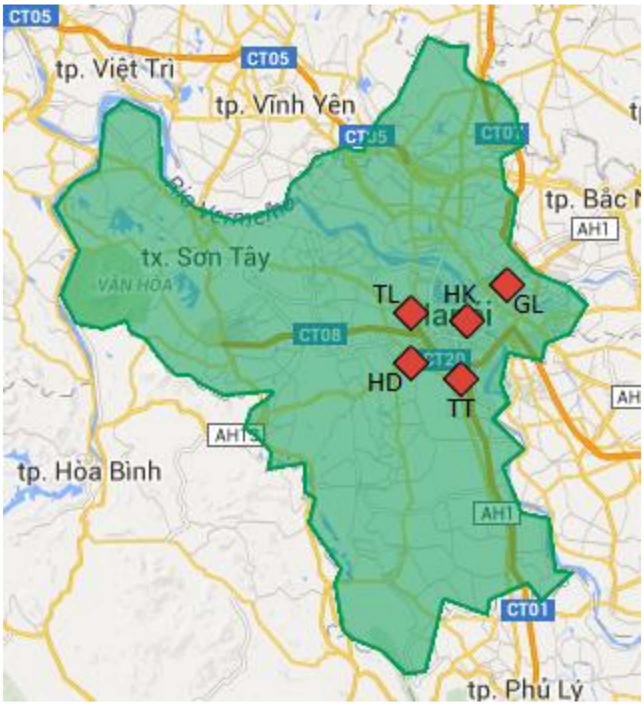

Figure 3.2 - Currently operated deployment stations

The average number of calls each day is 300, from which only one third (100) are real emergencies. Around 90% of these are made from homes in the city and there's not known frequency changes over the different periods of the day. The little demand may be explained by the lack resources and poor efficiency. Moreover, Vietnamese population often turn to self-medication.

Roughly, their average response time is 15 minutes, despite including almost exclusively urban rescues.

Although they do keep data on received calls, these are not stored digitally, which prevented extensive data analysis from demand history.

The ultimate objective of this study was to improve the service level of this organization just described by exploring with the layout of deployment stations. This improvement was decoded in shortening the RT and RTT for the entire province population. On the other hand, there were budget constraints, which were not specified but should stay reasonable. A solution utilizing the available resources was priority and further research on the impact of adding new fleet was a secondary goal. The approach should be robust in terms of demand variation and operations uncertainty.

3.3 Improvement opportunities

An early remark on this information took a special note on the spatial distribution of facilities. These are concentrated towards the province centre leaving suburban and rural areas poorly covered.

Moreover, the resources available seemed too few for the demand, both the number of stations and vehicles. The former could be tackled by cooperating with other entities, such as medical buildings.

4 Mathematical modelling

This chapter is dedicated to explain the first part of the actual solution developed – the mathematical modelling of the problem and its optimization. As mentioned earlier in this document, the operations of EMS have many sources of randomness. Although, models have been developed that tackle this issue by using probabilities, it’s still not the most accurate way to proceed. Furthermore, to estimate them, extensive data on demand and its accurate forecast is necessary, which isn’t available.

Instead, in this research, very simple models were solved in order to get different designs that served as starting points for simulations. This tool is able to mimic the stochastic reality better and was, therefore, used as the tool to ultimately find the best solution for the problem.

Before describing the actual models, the steps to gather the inputs they require are overviewed in the following sub-sections. The actual solving of the problems was accomplished with the tool IBM ILOG Cplex, the full results are presented in Appendix C.

4.1 Network nodes

In order to solve a FLP, a set of supply and demand locations is necessary. In this context, these are the deployment stations and aggregated population nodes and the first step is to study the province of Hanoi to choose them.

4.1.1 Deployment stations

On the supply side of the network there are the locations where ambulances wait idle. They may be any site with room for the vehicles, though the Hanoi EMS advised locations equipped with medical material. This allows ambulances to return to the same spot rather than having to resupply in other station.

In line with the plan of a lower-cost and shorter-term solution, it wasn’t suggested building new facilities. Thus, it left choosing the set of supply nodes from existing ones. A clear possibility was to implement cooperation with other entities and use their network. It was decided, at this stage, to screen the existing medical facilities (hospitals and smaller practices) for candidates.

After some research in Vietnam’s government and universities websites, it was compiled a list with the names of 663 public medical facilities in the province of Hanoi alone. However, many of these are very small practices with no room for vehicles, and some are different specialities in the same building, thus it wouldn’t make sense to consider more than one network node for those.

The task was, therefore, to reduce the list to a set of independent locations which were able to serve as a station for ambulances. In order to accomplish it, the kinds of existing medical facilities had to be understood to create filters to the list, these would have to be in Vietnamese since it was the language the names were in.

- Benh vien (BV) – “hospital”, is a large medical facility, usually specialized in a given field of medicine and containing emergency department
- Benh vien da khoa (BVDK) – “general hospital”, is a large medical facility with multiple services and areas, most have emergency department
- Phong kham (PK) – “clinic”, is a smaller office, often specialized in a field and with no big facilities to keep vehicles
- Phong kham da khoa (PKDK) – “polyclinic”, is a clinic slightly bigger and with services across more areas of medicine, though usually too small to have ambulances stationed in
- Phong kham da khoa khu vuc (PKDKKV) – “regional polyclinic”, is a polyclinic serving a region in rural areas, mostly smaller than the others

- Tam y te (TYT) or Trung tam y te (TTYT) – "health centre" or "medical centre", are the smallest medical practices listed, these are too small to be stations

After overviewing the characteristics of each kind of medical facility, it was decided that only the hospitals (general or not) are suitable candidates for emergency deployment stations.

The selected keywords to find these were "benh vien" and "bv" since they include both desired types of hospital.

From the original 663 names, a list of 27 possible stations emerged. These, however, do not include private and special kinds of hospitals.

The privates were found in a listing from the UK embassy healthcare advice for expats living in Hanoi and with an interview with the Director of the Customer Service Department of the Hanoi French Hospital – Sau Nguyen. Finally, international, state and military hospitals were researched in the website of Vietnam's department of health.

Adding the 5 owned and operated emergency stations, it was left a list with 42 names of medical facilities. The goal was, though, to find their geographical coordinates. In order to do that, a map was created in Google Maps MyMaps inputting an Excel spreadsheet with the list of names as the search field.

At that point, the problem of shared facilities between hospitals was solved by examining the obtained map – the private Hanoi French Hospital shares buildings with the state hospital Bach Mai and two previously existing deployment stations belong to the facilities of general hospitals. These were merged and the final set of possible deployment stations counted with 39 locations, 5 of which were the existing ones and 34 were new cooperation.

The list went through the personnel of the Hanoi EMS to be approved, furthermore, all these entities are included in a cooperation deal that is in place, though not being used.

Having the locations in the map, it was left to extract the geo-coordinates of each. It was done by exporting it to a KML file, which was processed in a spreadsheet to achieve a table with the names and respective coordinates in the format "Latitude, Longitude", as presented in Appendix A. The Figure 4.1 presents the selected possible deployment stations in the map of Hanoi.

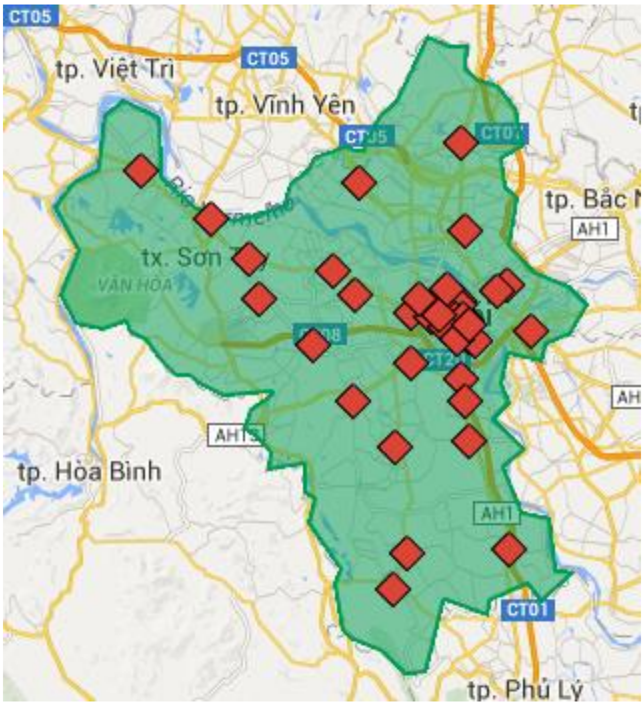

Figure 4.1 - Map of possible deployment stations

It's possible to state how much more coverage this scenario provides compared with the original scenario, especially in the rural areas.

4.1.2 Demand points

While the supply side is more straightforward, the demand adds a lot of complexity to the research due to its vastness and stochasticity. For instance, the need for emergency assistance may arise from every individual, thus the whole population are possible clients. Furthermore, unlike businesses where marketing or other factors influence customers' desires, health incidents are more unpredictable and require quick action, which in turn implies ever ready rescue operations. Despite these difficulties, demand isn't completely random and there are methods to forecast it.

Common practice in the field is to study patterns in the history of stored calls. However, this wasn't an option since, as mentioned, this data was not available digitally and it would be impractical to draw from paper the amount of records necessary to have statistical meaning. Another reason to deviate from this method in the research had to do with this information being biased. As laid out in previous sections, the EMS of Hanoi is inefficient, especially providing for the outer areas. It turns out that demand is, to a certain extent, elastic to supply, which means people underserved by the system turned away from it and most of the population in Hanoi (and Vietnam in general) are used to self-help. Forecasting the people's needs from this faulty data seemed unwise then.

As an alternative, the demand nodes were set based on demographic data. However, this information isn't available, or up to date, at the ward level (smallest political division) and the population of each district was used instead. As these areas are very large and variate (from 5.29 km^2, the smallest, to 428.00 km^2, the largest), they had to be split into several points that represented the locations' inhabitants.

A trade-off necessary at this stage had to do with the amount of demand aggregation. On one hand, the more nodes used, the more accurate the model would be, however, this accuracy grows at the expense of increased complexity which, at some stage, hinders or even prevents finding a solution. At the very limit, a point could be drawn on the map for each individual, this was impractical though. It was decided that one node would represent an area small enough so that the difference of the travel time of an ambulance to the centre of it and to an extreme of it would be unimportant.

It was used a region of 25 km^2 for the purpose. In this case, and considering a circular shape for the area, the radius could be covered in 3 minutes travelling at 50 km per hour, 2.5 at 68 or 2 at 85 (approximating as a straight line). This served as a lower limit to the accuracy of the data since each district's area was divided by 25 and that number, rounded up, was the number of demand nodes representing its population. An approximation was made here by splitting the number of people evenly through the nodes, which, in turn, were distributed through a map created in Google Maps MyMaps. In total, 151 points belong to the demand set, as presented in Appendix A.

The set of coordinates of the nodes was reached in the same way as in the deployment stations' case.

Even though sources of error were introduced in this procedure, it has been proven in past research that demand aggregation inaccuracies are overweighed by the choice of appropriate models (Brotcorne et al, 2003). Moreover, the simulation part of this research was responsible for mirroring the people's needs in a more realistic manner.

The demand points were distributed as displayed in Figure 4.2.

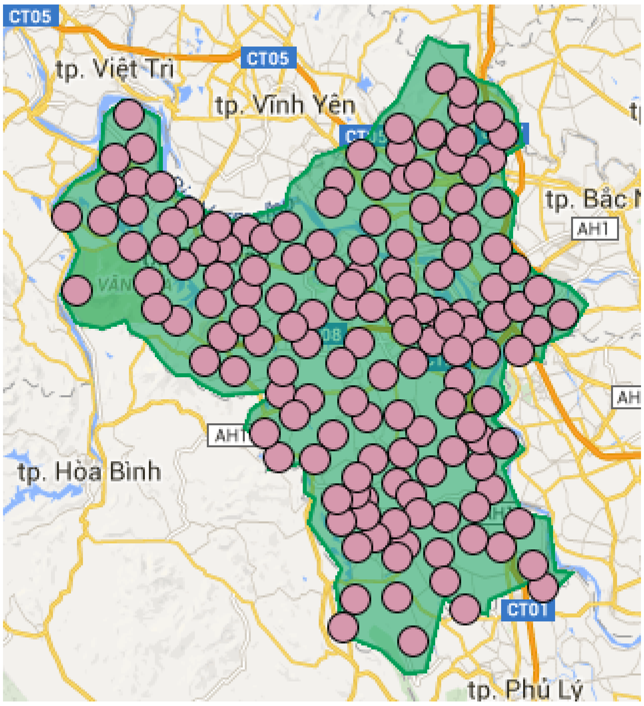

Figure 4.2 - Map of demand nodes

4.1.3 Time matrix

A final piece of the network was left – the time, or distance, between every station and every demand node, this is key input to de model. In order to get that, it was used the Google Maps API Distance Matrix running with the Python IDE Spyder.

This tool takes a set of origins and destinations and outputs two matrixes: one with the travel distance between each two points by the recommended road, and the other containing the time a car would spend on the same path.

The sets of origins and destinations coordinates were translated into text files to serve as inputs for the program.

The time matrix was targeted since it was thought to be more accurate than considering distances. This is especially important in Hanoi due to the poor condition of many roads and to the heavy traffic in some areas, and Google Maps directions times have these into account. Also, the final unit wanted was time, since the main performance metrics of EMS are RT and RTT, thus, if using distances, these would have to be converted into time and, for that, uniform speed would be considered, resulting in less authentic values.

The programming script written to accomplish the task is presented in Appendix B. The resulting matrix is not included in the document since it is too big (39*151).

4.1.4 Network overview

In order to have a better idea of how distant the demand nodes are from stations, the travel time to each demand point from its nearest facility (RT) was gathered from the matrix and a histogram was built with this data. As a comparison, the same method was applied using only the 5 original stations.

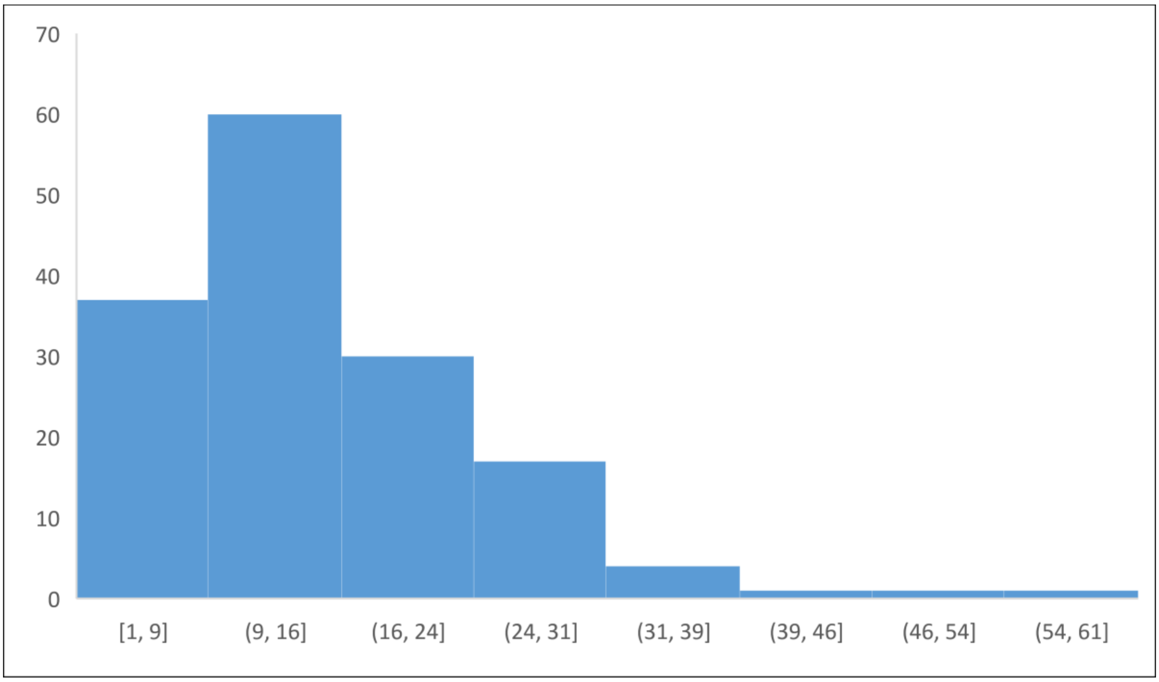

Figure 4.3 - Distribution of shortest RTs with 39 stations (minutes)

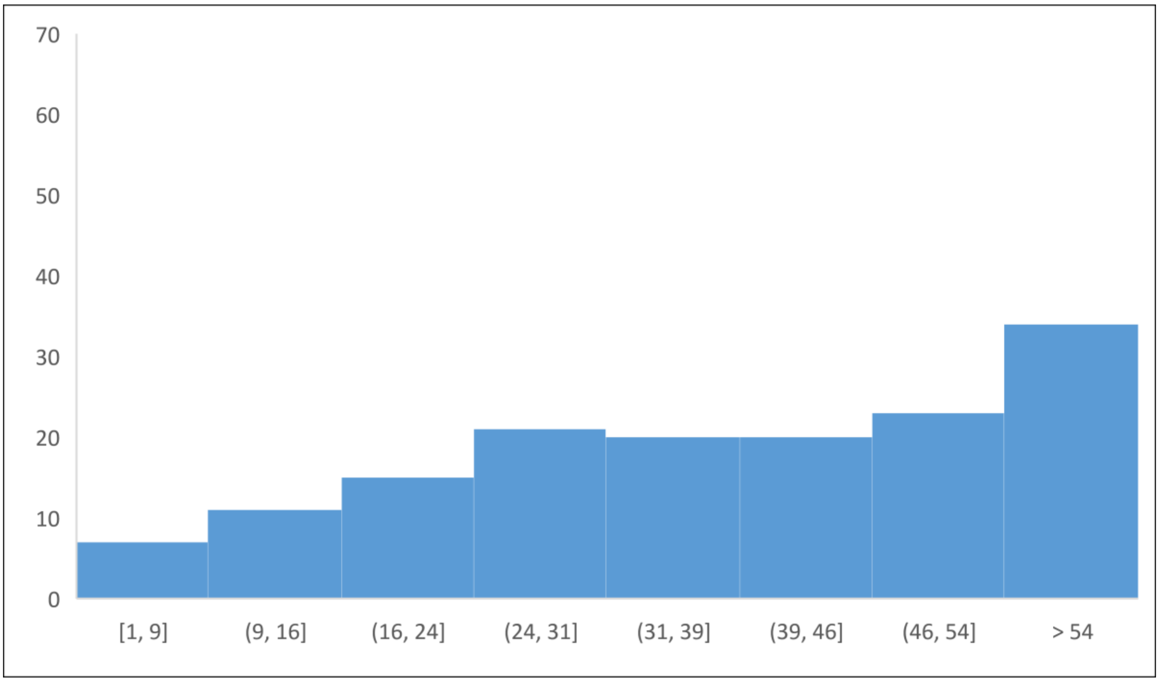

Figure 4.4 - Distribution of shortest RTs with 5 stations (minutes)

At first glance, there was a major shortening in the overall and maximum estimated RTs by adding the new facilities. With 39 stations, the longest one was 61 minutes while previously it was 97, also the average time dropped from 40 to 15 minutes.

4.2 Costs

Information regarding the costs of the EMS of Hanoi were not provided by the organization. However, the approach taken does not incur in high investment, therefore costs weren't as relevant as if the creation of new facilities was intended. These topic is discussed in a more qualitative perspective then, in this section.

In their owned facilities, it was assumed they have a fixed cost (F) and a cost to operate each rescue team (x). F includes the central service of receiving calls and dispatching, and also the bills of owning the 5 stations. The cost x is completely dependent on the number of teams working there, for the 23 vehicles it would be $23x$ then. If one team was dismissed, the variable cost would be reduced to $22x$.

The cost of operating the same team stationed in a hospital, for instance, would be greater than x, it was designated y. Besides the logistical complexity this case adds, it requires preparing the new facilities, such as storing supplies and communication equipment, which explain the increased cost of y.

On the other hand, if a team is allocated from one of the original stations to a hospital where another team is settled already, its marginal cost is less than the first one's since most logistical aspects are taken care of. Thus it's assumed that x is the expense of operating a team anywhere and y is a fixed cost of starting operations in a new facility. The fixed cost F remains constant in any discussed scenario.

As the first stage's goal is to make use of the current resources, the 23 vehicles, the operating cost is fixed and only y matters, being the added cost of a solution the number of new stations chosen times y.

Although it's harder to quantify, a solution's cost may be surpassed by its operational savings by requiring less travelling to reach victims, saving fuel and time.

4.3 Models

As discussed previously, the models utilized were simple mainly because there was a lack of accurate input data and for the availability of a simulation tool which was expected to fill the weaknesses of the mathematical optimization.

The major developments overviewed in FLP applied to EMS fall into reassuring multiple coverage and accounting for stochasticity in demand or operations. Both of these were tackled rather with the computer simulation model, by monitoring the utilization of stations in the first case, and including random inputs in the later.

The mathematical optimization results were then mainly utilized as a starting points for iterations in the built computer environment.

4.3.1 Inputs and outputs

In Figure 4.5 it is presented the data entering and leaving the models solved.

The ultimate output of the models is the allocation of ambulances to the deployment stations, however, a secondary result is the facility that serves each node, in the "p" formulations, and the covered demand points, for the coverage problems.

The input data changes from model to model, though the time matrix, the demand of each point and the number of vehicles are common to all.

The time standards are inputs required only by the covering models (*) and percentage value is part of the DSM model only (**).

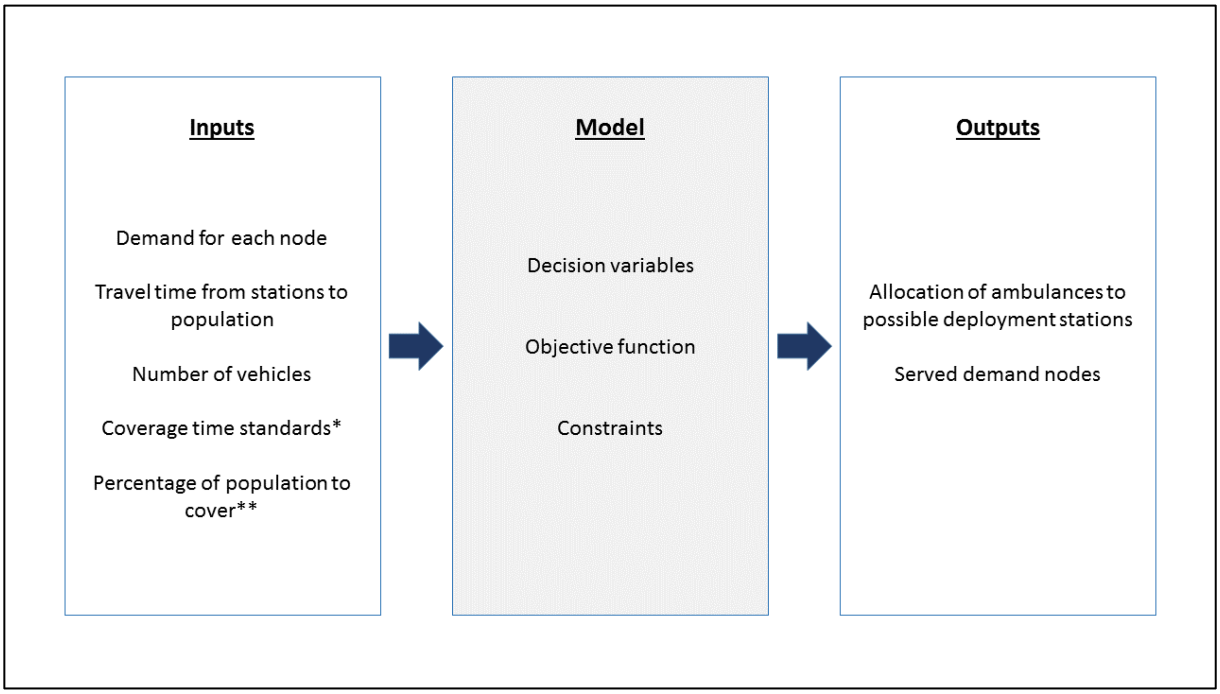

Figure 4.5 - Inputs and outputs of the mathematical models

4.3.2 Variables

V	Set of demand nodes
W	Set of possible ambulance stations
W_i	Set of possible ambulance stations within coverage time standard of demand node i
W_i^1	Set of possible ambulance stations within the shortest coverage time standard of demand node i
W_i^2	Set of possible ambulance stations within the longest coverage time standard of demand node i
t_{ij}	Travel time from station j to demand point i
d_i	Demand of node i
y_i	Equal to 1 if node i is covered, 0 otherwise
y_{ij}	Equal to 1 if node i is supplied by station j, 0 otherwise
y_i^1	Equal to 1 if node i is covered within the shortest time range by 1 vehicle, 0 otherwise
y_i^2	Equal to 1 if node i is covered within the shortest time range by 2 vehicles, 0 otherwise
x_j	Number of ambulances placed at station j
Q	Longest travel time between any demand node and its closest station

4.3.3 P-median model

Objective function:

$$Minimize \sum_{j \in W} \sum_{i \in V} d_i t_{ij} y_{ij} \tag{4.1}$$

$$\sum_{j \in W} y_{ij} = 1 \ (i \in V) \tag{4.2}$$

$$y_{ij} \leq x_j \ (i \in V, j \in W) \tag{4.3}$$

$$\sum_{j \in W} x_j = p \tag{4.4}$$

$$x_j \in \{0,1\} \ (j \in W) \tag{4.5}$$

$$y_{ij} \in \{0,1\} \ (i \in V, j \in W) \tag{4.6}$$

The objective function minimizes the sum of the travel time between each demand node and the facility that serves it, weighting the demand for each link. The Boolean variable y_{ij} takes the value 1 only when the facility *j* is the supplier of node *i* thus only these “utilized” times are summed in the equation. It is considered that a demand point is served by a single station since there aren’t restrictions to the demand each can serve and there’s only one that is the closest.

Equation (4.2) states that each population node must be served by one supply site, while (4.3) sets the decision variable x_j by ensuring that, if y_{ij} states that *j* is in use, then an ambulance must exist in site *j*.

The number of vehicles available is expressed by (4.4) and equations (4.5) and (4.6) define the decision variables Boolean.

4.3.4 P-centre model

Objective function:

$$Minimize\ Q \tag{4.7}$$

Constraints:

$$\sum_{j \in W} y_{ij} = 1 \ (i \in V) \tag{4.8}$$

$$y_{ij} \leq x_j \ (i \in V, j \in W) \tag{4.9}$$

$$\sum_{j \in W} x_j = p \tag{4.10}$$

$$\sum_{j \in W} t_{ij} y_{ij} \leq Q \ (i \in V) \tag{4.11}$$

$$x_j \in \{0,1\} \ (j \in W) \tag{4.12}$$

$$y_{ij} \in \{0,1\} \ (i \in V, j \in W) \tag{4.13}$$

The goal of this model, equation (4.7), is to minimize the variable Q, which is forced by (4.11) to be greater or equal to any used travel time to the demand locations, this results in Q being the longest one.

The remaining constraints are identical to the P-median model.

4.3.5 Maximal covering location model (MCLP)

Objective function:

$$Maximize \sum_{i \in V} d_i y_i \tag{4.14}$$

Constraints:

$$\sum_{j \in W_i} x_j \geq y_i \; (i \in V) \tag{4.15}$$

$$\sum_{j \epsilon W} x_j = p \tag{4.16}$$

$$x_j \in \{0,1\} \; (j \in W) \tag{4.17}$$

$$y_i \in \{0,1\} \; (i \in V) \tag{4.18}$$

The aim of the objective function is to serve as much demand as possible, the variable y_i states if node *i* is covered.

Equation (4.15) attributes ambulances to each site *j* by stating that the sum of vehicles within coverage time of node *i* must be greater or equal to the value of y_i. If *i* is covered, it's value is 1 thus there must exist at least one ambulance within range.

The resources constraint is set by (4.16) and the remaining equations define the variables.

The coverage time standard used for this model was 15 minutes.

4.3.6 Double standard coverage (DSM)

Objective function:

$$Maximize \sum_{i \in V} d_i y_i^2 \tag{4.19}$$

Constraints:

$$\sum_{j \in W_i^2} x_j \geq 1 \; (i \in V) \tag{4.20}$$

$$\sum_{j \in V} d_i y_i^1 \geq \alpha \sum_{i \in V} d_i \tag{4.21}$$

$$\sum_{j \in W_i^1} x_j \geq y_i^1 + y_i^2 \; (i \in V) \tag{4.22}$$

$$y_i^2 \leq y_i^1 \; (i \in V) \tag{4.23}$$

$$\sum_{j \in W} x_j = p \tag{4.24}$$

$$y_i^1, y_i^2 \in \{0,1\} \; (i \in V) \tag{4.25}$$

$$x_j \; integer \; (j \in W) \tag{4.26}$$

The goal of this model is to maximize the demand covered twice within the shortest time standard.

Equation (4.20) entails that all population must have at least one ambulance within the longest time standard. Only a percentage α of it is required to have a vehicle within the shortest one, this is guaranteed by (4.21).

The variable x_j is set by equation (4.22) which states that the number of ambulances allocated to site *j*, within the shortest time standard of *i*, must be greater or equal to the number of vehicles node *i* has covering it. If one is covered twice, then it is implicitly covered once, this is assured by (4.23).

The following equations are meant to limit the number of vehicles and define the decision variables.

The shortest time standard utilized for solving this model was the same as in MCLP – 15 minutes. The longest, however, can only take values greater than 61 minutes since there's a condition requiring the whole population to lie within it and, as it can be observed in the chart from section 4.1.4, there's a demand node which is that far apart from the its closest station.

The percentage α was given the value of 0.75 since this was the maximum possible population to cover within 15 minutes.

4.3.7 Cost constraint

In order to soften the costs and make use of the currently operated facilities, all presented models were solved a second time with an added constraint implying that the first 5 stations (the ones owned) had to be allocated at least one ambulance.

Constraint:

$$x_j \geq 1 \, , j \in \{1,5\} \tag{4.27}$$

In total, 8 models were solved and all the optimal objective functions were reached, these may be checked in Appendix C.

5 Computer simulation

This chapter describes the second part of the solution – the computer simulation. The goal was to create a virtual system where the different designs could be tested in and analysed to decide the most suited for Hanoi EMS.

In the next pages, the modelling of the different aspects that compose the emergency operations' environment is explained, such as the demand generation, deployment decisions and performance indicators to monitor. It follows the actual results obtained for the 8 distinct layouts obtained in the previous section and their analysis, which led to further tinkering with the ambulance allocations to search for better designs.

The software utilized for the computer simulation was AnyLogic (version 7.2.0) since it has GIS tools and allows the three different kinds of models: agent based, discrete events and system dynamics.

5.1 Description of the model

The model is mostly agent based, with some state charts controlling their behaviour. In the main agent, it was placed a GIS interface in order to display and use the map of Hanoi. The parameters and variables used for each agent are presented in Table 5.1.

Table 5.1 – Parameters and variables of each agent present in the model

Agents	Parameters	Variables
Main	-	-
Victim	*parGoToHospital* *parNearestStation*	-
Help	*parVictim*	-
Station	*parCode* *parName* *parCapacity*	*varAvailableAmbulances* *varRescues* *varBusyTimes* *varGoToHospitalTimes*
Ambulance	*parStation*	*varHelp*

5.1.1 Facilities distribution

The first task to build the model was to draw the network of the EMS in the GIS map of the simulation environment.

It was then created a type of agent called *Station* and a population containing agents of the mentioned type. This population is initially empty, and it was programmed to be loaded from a database sheet and added to the model in the beginning of each simulation run. This contains fields for latitude, longitude, name, capacity (number of ambulances) and code for the 39 sites.

The GIS coordinates were connected to the latitude and longitude of the table thus placing the facilities in the map for visual representation. They are represented by an icon of a building and it was added a text label which was set to show the station's code, capacity and available vehicles in real time during the simulation run.

Moreover, the parameters name, capacity and code were added to the agent *Station* and also linked to the database values.

In terms of variables, this agent contains one for the available ambulances, equal to the number of vehicles idle. Three additional variables capture the utilization and proximity metrics of each stations: “varRescues”, “varBusyTimes” and “varGoToHospitalTimes”.

Lastly, each station has a pool of resources, which are agents of type *Ambulance* and their quantity is set from the parameter capacity.

5.1.2 Demand modelling

In order to create the demand on the simulation, a population map from the online provider Esri was analysed with the GIS tool: ArcGIS. As it is shown in Figure 5.1, the demand can roughly be divided in three different zones: the central one with a very large population density, followed by the suburban areas with an intermediate concentration of people and, lastly, the outer, rural locations with the least number of inhabitants per km^2.

The goal was to draw the shapes of these areas, and not to rely on the population values provided since these were available from good sources at district level as seen before.

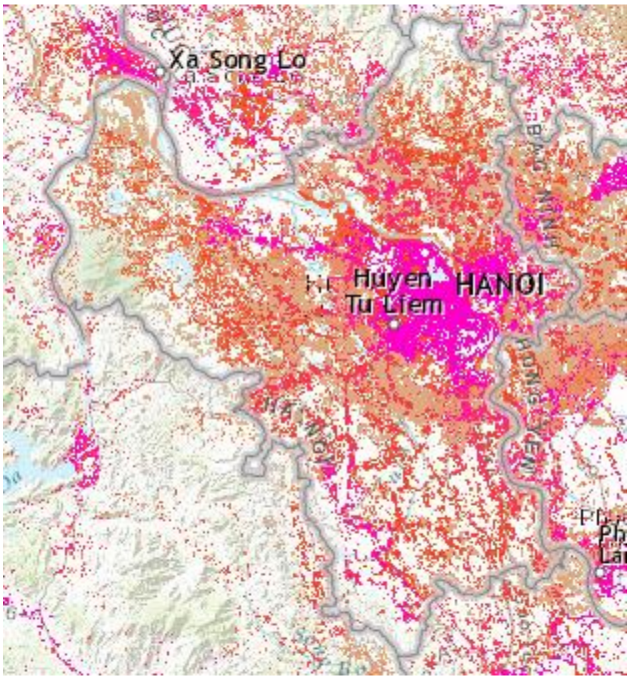

Figure 5.1 - Population map of Hanoi from Esri (2015)

These three areas were placed over the model’s base map as presented in Figure 5.2. It was left to select the demand of each zone. For this, the province of Hanoi was analysed in order to decide which districts belonged to each demand area. The population of each set of districts was summed, representing the population of the zone, as can be followed in Table 5.2.

Table 5.2 - Population in each of the three demand zones

Districts	Population	central	medium	outer
Hoàn Kiếm	147334	1		
Thanh Xuân	223694	1		
Ba Đình	225910	1		
Hai Bà Trưng	370726	1		
Đống Đa	410117	1		
Cầu Giấy	260643	1		
Tây Hồ	130639	1		
Nam Từ Liêm	232894	1		
Hoàng Mai	380509	1		
Bắc Từ Liêm	320414	1		
Hà Đông	260136	1		
Long Biên	271913	1		
Thanh Trì	198706	1		
Đan Phượng	142480		1	
Hoài Đức	191106		1	
Phúc Thọ	159484		1	
Sơn Tây	125749		1	
Gia Lâm	251735		1	
Thường Tín	219248		1	
Thanh Oai	167250		1	
Mê Linh	191490		1	
Quốc Oai	160190			1
Phú Xuyên	181388			1
Đông Anh	333337		1	
Ứng Hòa	182008			1
Thạch Thất	177545			1
Mỹ Đức	169999			1
Chương Mỹ	286359			1
Sóc Sơn	282536		1	
Ba Vì	246120			1
	total population:	3433635	2064415	1403609
	relative population:	49.75%	29.91%	20.33%

Considering the information given by the Hanoi EMS, the total demand should approximate the 100 calls per day (in real emergencies). Thus the inner area should contain roughly 50% of it (50 victims), the suburban zone 30% (representing 30 people) and the outer locations around 20% (20 calls per day).

One flowchart element "source" was added to the *Main* agent for each zone, creating demand – agents of the type *Victim*. These created the persons randomly in time and within the containing space, according to the rate per day established for each area.

This type of agent was added two parameters: "parGoToHospital", to store facility nearest to the victim's, which it's where it should be taken, and "parNearestStation", the station responsible to make the rescue. These may coincide.

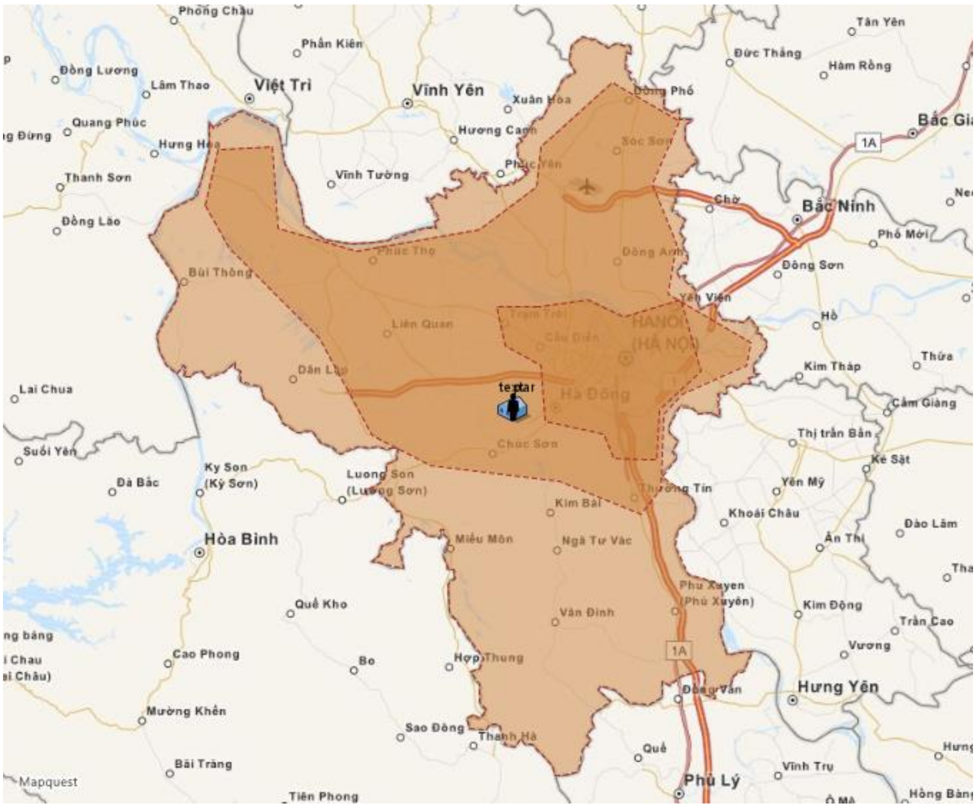

Figure 5.2 - Demand zones, with 50%, 30% and 20% of the province's population, from the inside to the outside

5.1.3 Model logic

The actions in the model were programed in Java which is the language of the utilized software – AnyLogic. The necessary connections between agents were made in order to be possible to move flowchart agents and messages between them, thus coordinating the system.

Main agent

The action in the main agent happens through logic blocks, starting in the mentioned three sources in parallel, as it is shown in Figure 5.3. Before continuing down the chain, a command implemented in this block's exit searches for the nearest facility to the newly created victim, by real road distance from the GIS map, and attributes its parameter "parGoToHospital" the selected station.

To set the parameter "parNearestStation", a code finds the set of sites with the variable "varAvailableAmbulances" greater than zero, which are the ones with immediate availability, and selects the nearest one by GIS path. In case of every existing resource being busy, it is chosen the nearest facility from the set of those with capacity different from zero, meaning it has vehicles in its fleet.

Still in the same block, a command checks if the chosen station for the rescue is the nearest one, from the ones in use. If not, the variable "varBusyTimes" of this later facility is added one unit since it didn't have vehicles available when needed. A last command adds one unit also to the variable "varGoToHospital" of the station chosen as the destination to take the victim to, meaning it was the nearest site (either in use or not).

It was decided this way, instead of nearest vehicle, since ambulances are required to return to a station for resupplying after each rescue, as it was imposed by the Hanoi organization.

The next block is called “timeStart” and it initiates the clock on the agent *Victim* that went through it.

Afterwards, there is a time delay, corresponding to the time the person is waiting for an ambulance. It’s exit is only triggered by calling the function “stopDelay()” and it is invoked when the a vehicle reaches the victim. The time up to this point, the RT, is stored by placing in the chain a time measure block.

The next delay represents the time of patient’s stabilization on-spot and the transportation to the hospital, being interrupted only by a stop function. The RTT is marked at this stage by a time measure block and, lastly, the agent victim sinks (leaves the simulation).

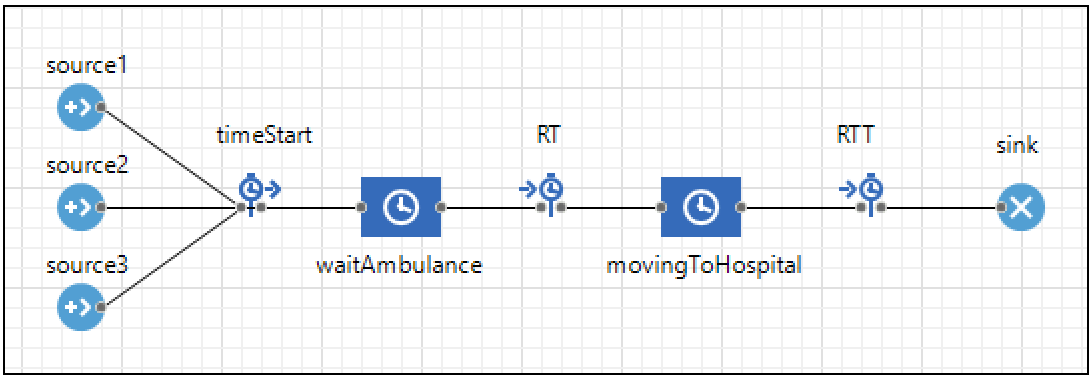

Figure 5.3 - Block diagram of Main agent

Victim agent

When appearing in the map, besides the chosen person icon, it was added a text label to the victim which was programmed to show the parameter “parNearestStation” so that the user was able to visually follow which facility was rescuing it.

The behaviour of the victims is controlled by a state chart as in Figure 5.4. After being created and its parameters set as explained previously, the first state is calling for help. An entry action was programed to create a new agent of the type *Help* and set its only parameter, which indicates its requester, to be itself, the victim. Still in the same sequence, the *Help* agent is sent to the station indicated as the nearest available, in order to be used in its block chain.

The transition to the next state, called “stMovingToHospital” is triggered by receiving a message in form of the string “helped”, which comes from the agent *Ambulance*. At entering this state, the person icon is commanded to disappear, visually simulating the rescue team assistance and pick-up, and the function to stop the waiting for ambulance delay in the *Main* agent is called.

Similarly, the transition to the final state occurs when the message “arrived” is received and an action follows calling the function to end the delay block “MovingToHospital” in the *Main* agent.

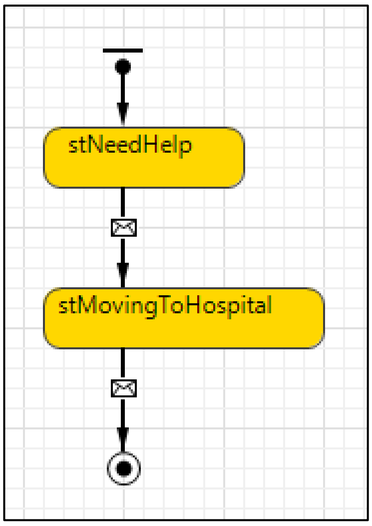

Figure 5.4 - State chart of agent Victim

Station agent

After receiving the *Request* agent sent by a victim, this enters a logic blocks chain (Figure 5.5) within the agent *Station*. The first element is of the type “Seize” and is used to select a vehicle of the pool of resources. Here the variable indicating the available ambulances is updated (reduced by one) and the parameter “parHelp” of the seized vehicle is set to be the agent *Help* going through the block diagram, indicating the help request it is attending.

Afterwards, a delay was placed, which ends only when a stop function is called (when the ambulance is back at base), and it follows a block to release the resource, making it available and updating the variable of available vehicles.

Lastly, the variable of rescues made by the station is added one unit.

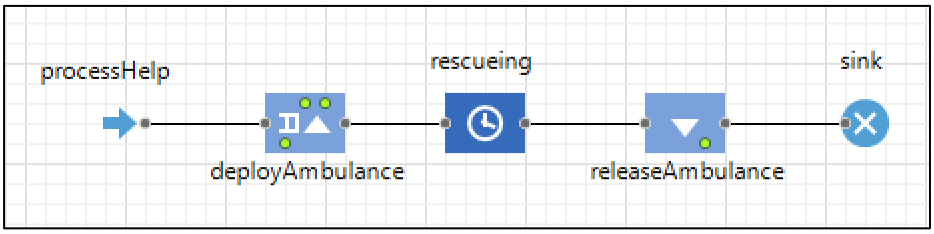

Figure 5.5 – Block diagram of the agent Station

Ambulance agent

The vehicles’ behaviour is imposed by state charts (Figure 5.6), being the starting one named “atStation”, which represents it being idle.

It transitions to the state “Preparing” when it is attributed a help request by the station. Its exit is triggered by a timeout with a triangular distribution ranging from 0.5 to 4 minutes, with mode of 2. This translates the time, in real operations, that it takes for the team to leave the station after the emergency call is received.

After that, the ambulance is programed to move to the victim, selected from the parameter “parVictim” of the *Help* agent, by the fastest path using real roads in the map and this travel can be followed in the GIS display during the simulation.

The travel speed of the vehicles is given by a triangular distribution with vertexes in 35, 60 and 90 km/h.

Another transition occurs when the person is reached, at which point the string “helped” is sent to that victim as a message. The state “helping” is left when a timeout (triangular distribution with 2, 5 and 10 minutes) happens, simulating the duration of assisting the injured on-site.

The next action is to move the vehicle to the destination hospital, that is retrieved from the victim’s parameter “parGoToHospital”. When it has arrived, another timeout (with a duration distributed triangularly with values 0.1, 0.5 and 1) mimics the duration of unloading the patient and the message “arrived” is sent to the agent *Victim*. Then, if that hospital is the vehicle’s home station, it returns to the starting idle state, otherwise it returns to its original facility, and waits for the next emergency. In either case, when the ambulance finishes the cycle, a function is called to stop the delay “rescuing” in the home station’s flowchart.

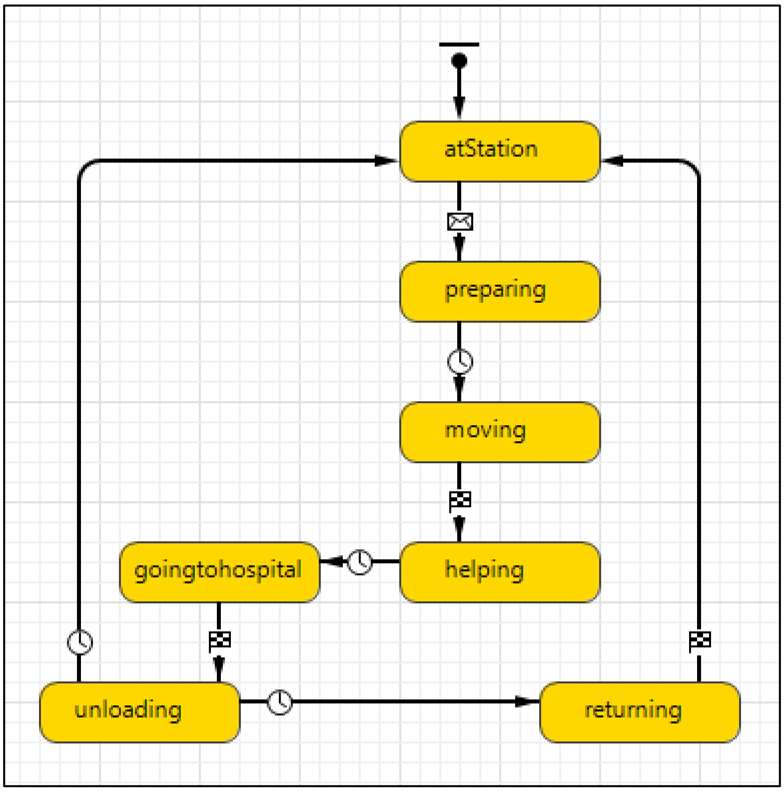

Figure 5.6 - State chart of the agent Ambulance

The reason to consider all time delays distributions lies on their random nature, it doesn’t last the same to assist a victim every time, or to travel a certain distance due to traffic conditions. Although triangular distributions aren’t the most accurate, they were used for there was a lack of great amounts of data to shape curves from and for their simplicity. All minimum, maximum and median values were estimated by the Hanoi EMS personnel based on their experience. They also validated the logic and sequence of tasks of the model.

5.1.4 Control panel

In the pursuance of providing the EMS of Hanoi a simple and easy to work with tool, yet still complete and insightful, the simulation was needed to be made available and capable of testing different designs in varying scenarios of demand and time delays, without requiring the proprietary software AnyLogic or any difficult programming.

A control panel, presented in Figure 5.7, was developed for this effect and the model was exported as a Java Applet. The main desired function was that of controlling the capacity (number of ambulances) of the stations. Thus “edit boxes” were added to the main simulation panel, one for each station, which allowed to do so.

For testing different demand scenarios, it was made possible to tune the rates of each zone, either maintaining the current ratio or not. It isn’t achievable, however, for a user of the applet to change the shapes of the population areas without working in the building software.

Finally, the average speed and time delays were also changeable so that they could be updated to real data shifts or to witness the gains for the overall system by improving the performance in those tasks. The times referred include the deployment, the victims’ stabilizing and the hospital drop-off.

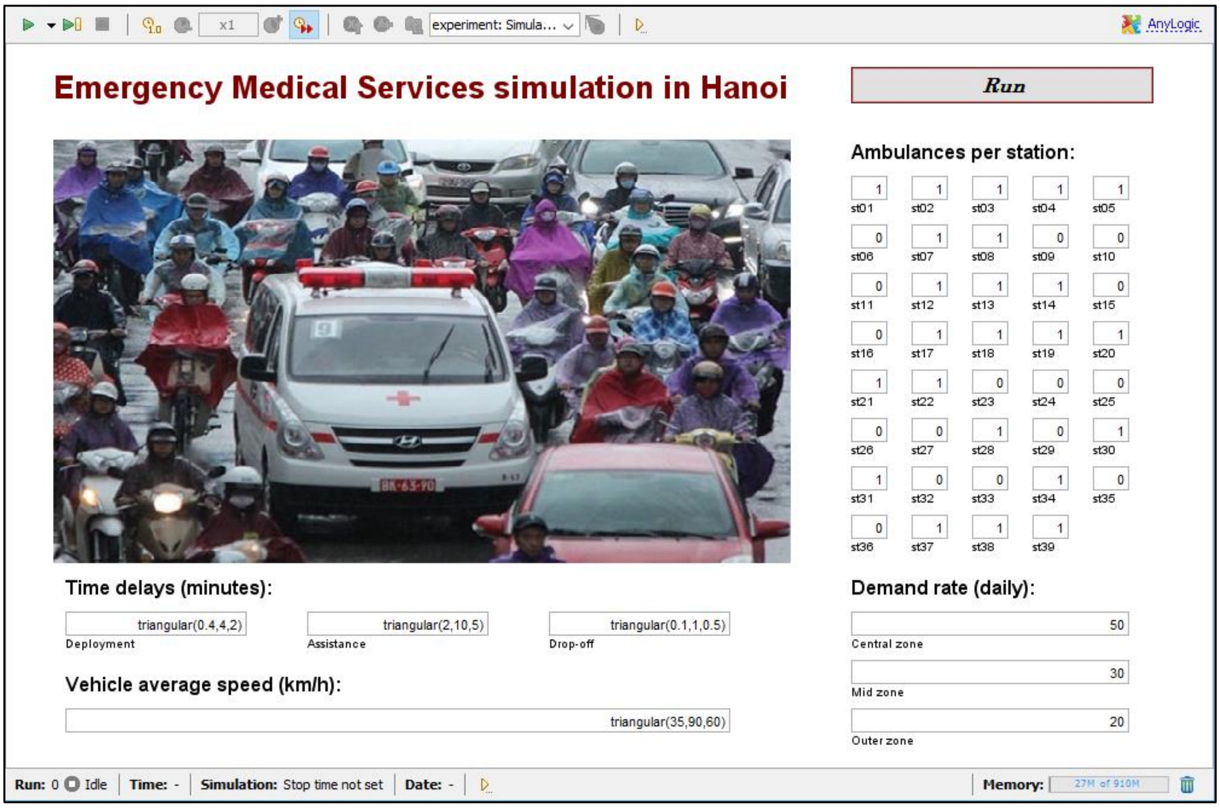

Figure 5.7 - Control panel for the simulation

5.2 Performance metrics

The key indicators to evaluate a design on are the Response and Round-trip times, for these are the most strongly correlated with the victims’ survival. During the simulation, it is presented in real time the distribution of both metrics as well as their average and maximum values, as displayed in Figure 5.9.

Other aspect to monitor is the utilization of each facility: number of rescues made and number of requests denied due to the unavailability of resources. This was a path to try relocations in the designs. By reviewing which stations lacked vehicles when needed and which were very scarcely utilized, ambulances were moved to where they seemed more necessary in the pursuance of reaching a better allocation plan than the original layouts.

The number of instances that each station (utilized or not) was the nearest to a victim was also captured with the intent of determining which sites not in use should be allocated vehicles.

The whole data is exported to a spreadsheet in the end of each run, in which further analysis can be completed. Besides building histograms and collecting maximum and average values on the performance times, the values on the percentiles 25%, 50%, 75%, 90%, 95% and 99% were exposed, for each iteration done, to have a better view of the population coverage.

The summary of the described functioning of the simulation model is presented in Figure 5.8.

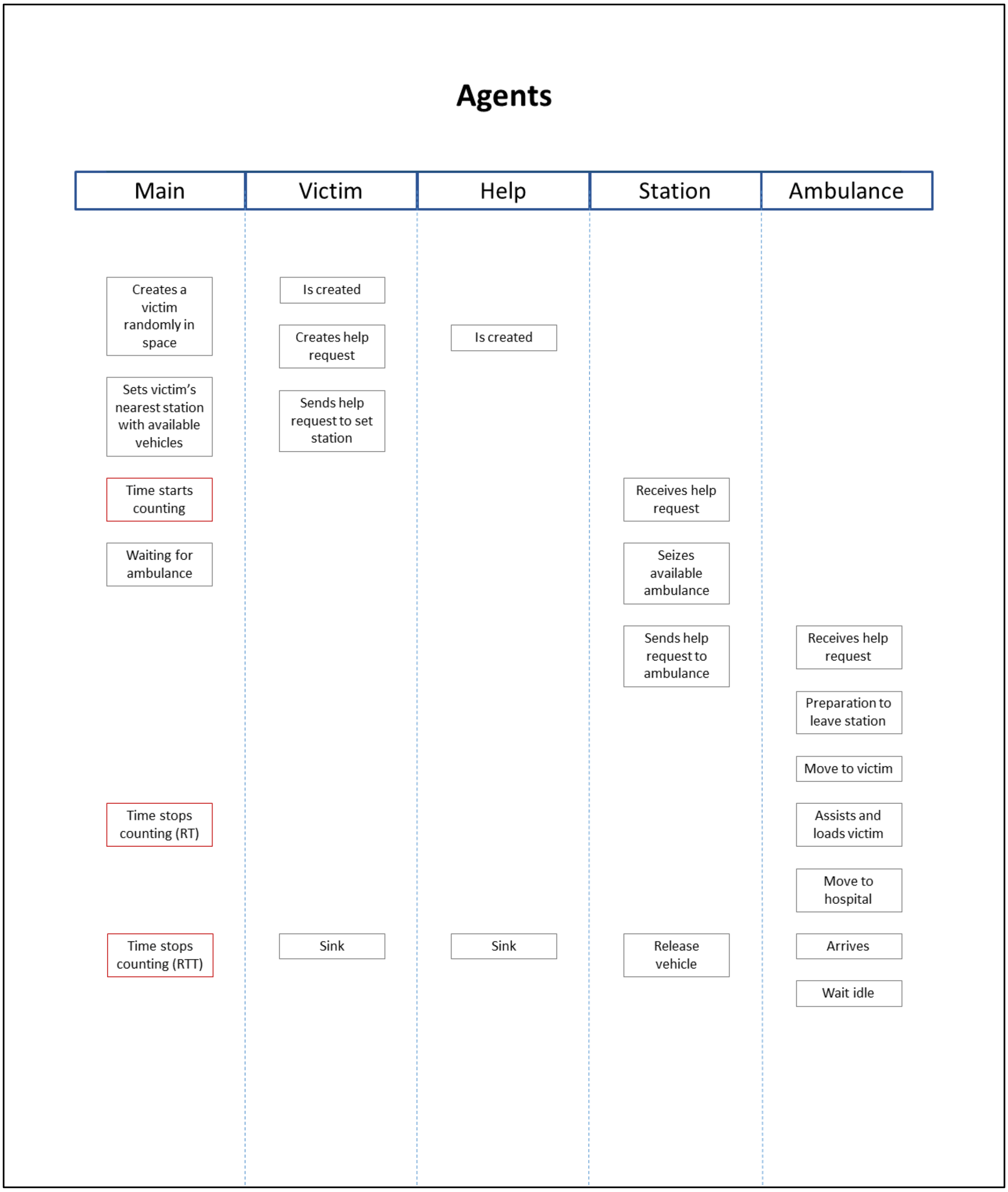

Figure 5.8 - Summary of the simulation model main tasks

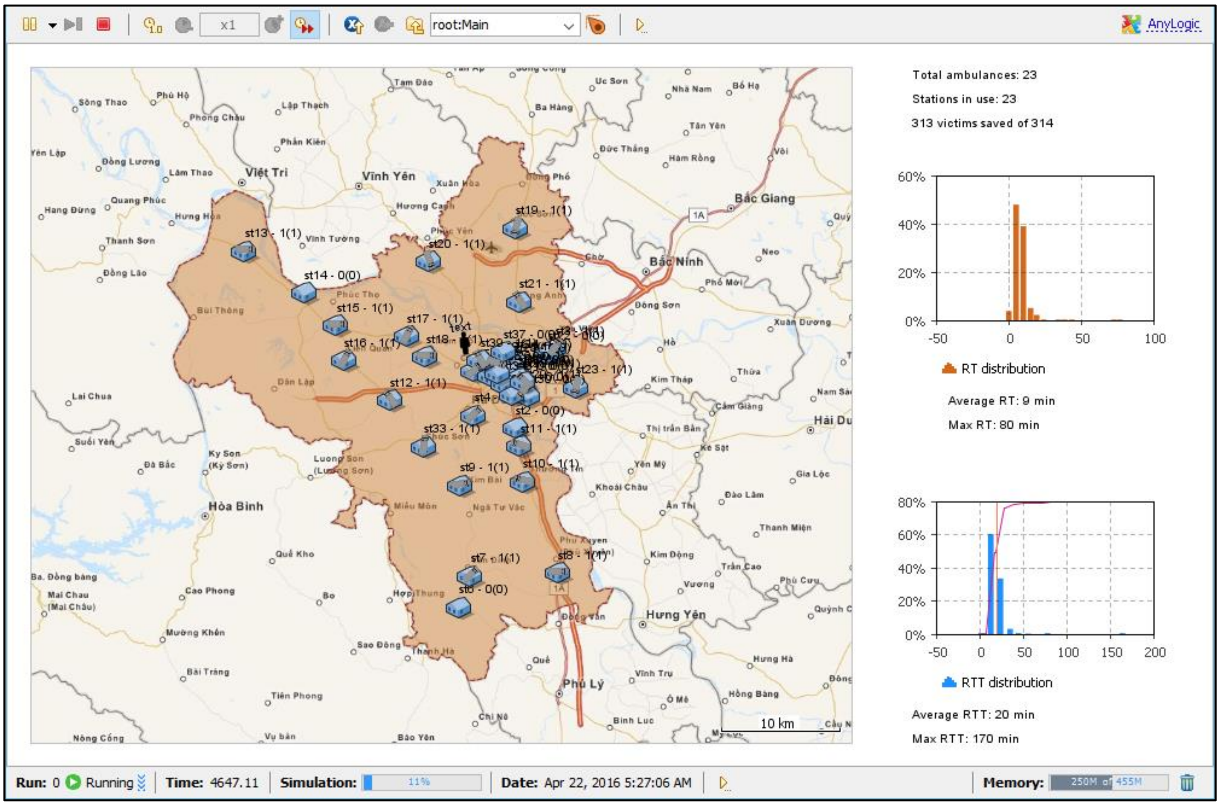

Figure 5.9 – Display panel when running a simulation, showing the KPIs in real time

5.3 Simulation run length

The length that a simulation run should have in order to be statistically precise is a very difficult parameter to estimate. Mahajan and Ingalls (2004) argue that it is preferable to have a longer run than several shorter replications in order to have less deviation from the expected mean results. When data is available it is a good practice to adapt the length to the range of the sample used. However, this is not the case of this research and, as to allow for several iterations in practical time, it was decided that each run would be of one virtual month, originating around 3000 calls.

5.4 First results

The first layout to test in the simulation was the currently used (with the 5 original stations) in order to validate the model and have a starting point for comparisons.

The 8 scenarios obtained with the mathematical modelling followed and the performance of all is summed in Table 5.3 for the RTT and in Table 5.4 for the RT, in the different metrics discussed.

In terms of notation, it was adopted to call each model by their designation and add a “5” to it when referring to the formulation with the added restriction of utilizing the five original deployment stations. The model “current” specifies the current design of the Hanoi EMS.

The RT will be the main focus and the best judge of a design’s values since it is the most strongly linked with survival and RTT is dependent on it. It is also the only parameter observed in most literature cases. Where the metric (RT or RTT) is not explicit, the results that are being talked about are the RTs.

Table 5.3 - Simulation results for RTT (minutes)

			Percentile					
Model	**avg**	**max**	**0.25**	**0.50**	**0.75**	**0.90**	**0.95**	**0.99**
current	32	105	17	26	44	61	70	95
p-median	25	124	15	20	32	45	54	76
p-median5	23	161	15	19	25	39	53	76
p-centre	28	110	17	23	36	53	62	90
p-centre5	26	145	15	21	33	48	56	80
MCLP	24	126	14	20	29	40	52	92
MCLP5	29	160	16	22	33	55	82	106
DSM	26	90	18	23	31	46	55	73
DSM5	26	106	15	20	30	45	63	101

Table 5.4 - Simulation results for RT (minutes)

Model	avg	max	0.25	0.50	0.75	0.90	0.95	0.99
current	20	82	8	14	29	43	49	65
p-median	12	62	6	9	14	25	32	39
p-median5	10	87	6	8	12	20	27	41
p-centre	15	66	7	11	22	34	37	52
p-centre5	14	85	6	9	18	30	35	55
MCLP	11	70	6	8	15	20	27	46
MCLP5	14	90	7	10	16	27	41	52
DSM	13	58	7	10	15	27	35	43
DSM5	12	58	6	9	14	30	34	49

As to easily rank the overall best designs, these were ordered from best performance (shortest time) to worst, top to bottom, in Table 5.5.

Table 5.5 - Models ordered from best performance (on top) to worst in each metric for RT

avg	max	0.25	0.50	0.75	0.90	0.95	0.99
pmedian5	dsm	pmedian5	pmedian5	pmedian5	mclp	pmedian5	pmedian
mclp	dsm5	mclp	mclp	dsm5	pmedian5	mclp	pmedian5
pmedian	pmedian	dsm5	dsm5	pmedian	pmedian	pmedian	dsm
dsm5	pcentre	pmedian	pmedian	mclp	dsm	dsm5	mclp
Dsm	mclp	pcentre5	pcentre5	dsm	mclp5	dsm	dsm5
pcentre5	current	mclp5	mclp5	mclp5	dsm5	pcentre5	mclp5
mclp5	pcentre5	pcentre	dsm	pcentre5	pcentre5	pcentre	pcentre
pcentre	pmedian5	dsm	pcentre	pcentre	pcentre	mclp5	pcentre5
current	mclp5	current	current	current	current	current	current

It is straightforward to choose the p-median5 as the best candidate and the MCLP following it by observing the rankings. The field "maximum RT" isn't very indicative of one's performance because there's randomness in the system and, as such, it is prone to outliers appearing in the extreme range of values. These also pushes the percentile 0.99 up, which is a good explanation for the weaker performance of both designs in this metric, even though ranking best in the others.

In order to have a perspective of the improvements these two selected layouts mean facing the original scenario, Table 5.6 presents the fraction of time metrics obtained with those, compared to the later. Moreover, a histogram was built for each set of results, with identical axis and scales, to allow a visual comparison.

Table 5.6 - MCLP and p-median5 results compared with current design's (model/current)

Model	avg	max	0.25	0.50	0.75	0.90	0.95	0.99
MCLP	56.60 %	85.22%	74.61%	59.06%	50.44%	45.56%	54.62%	71.74%
p-median5	52.62%	106.24%	72.58%	55.20%	40.28%	46.61%	53.97%	63.64%

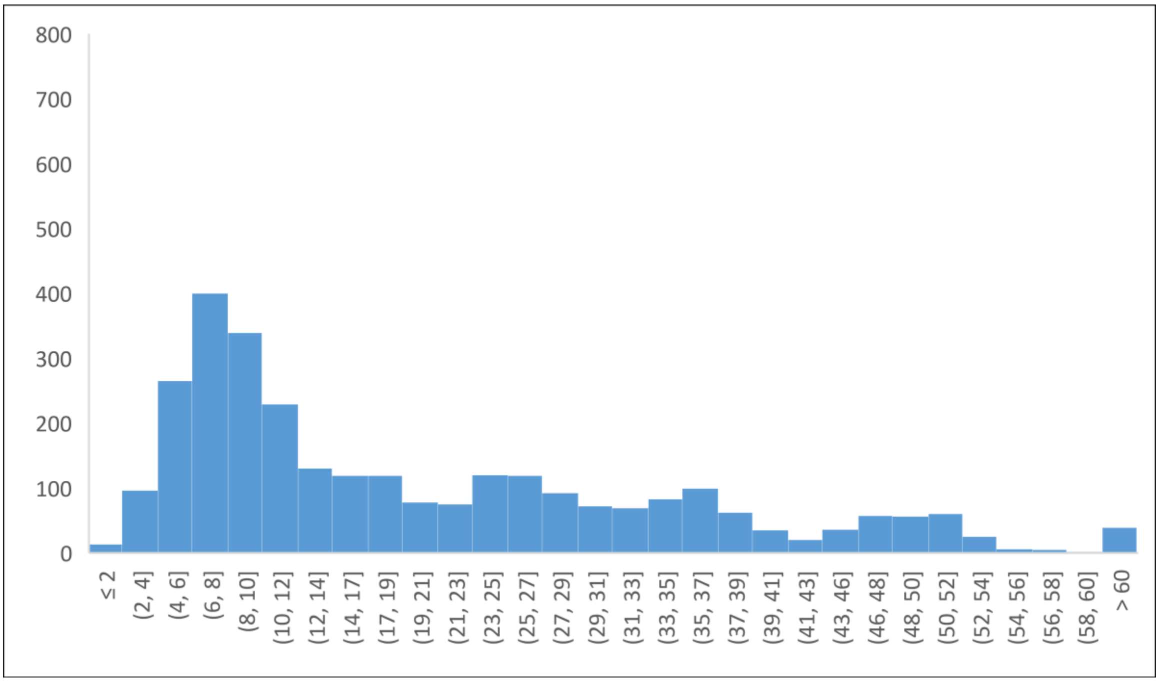

Figure 5.10 - Distribution of RT with current design (minutes)

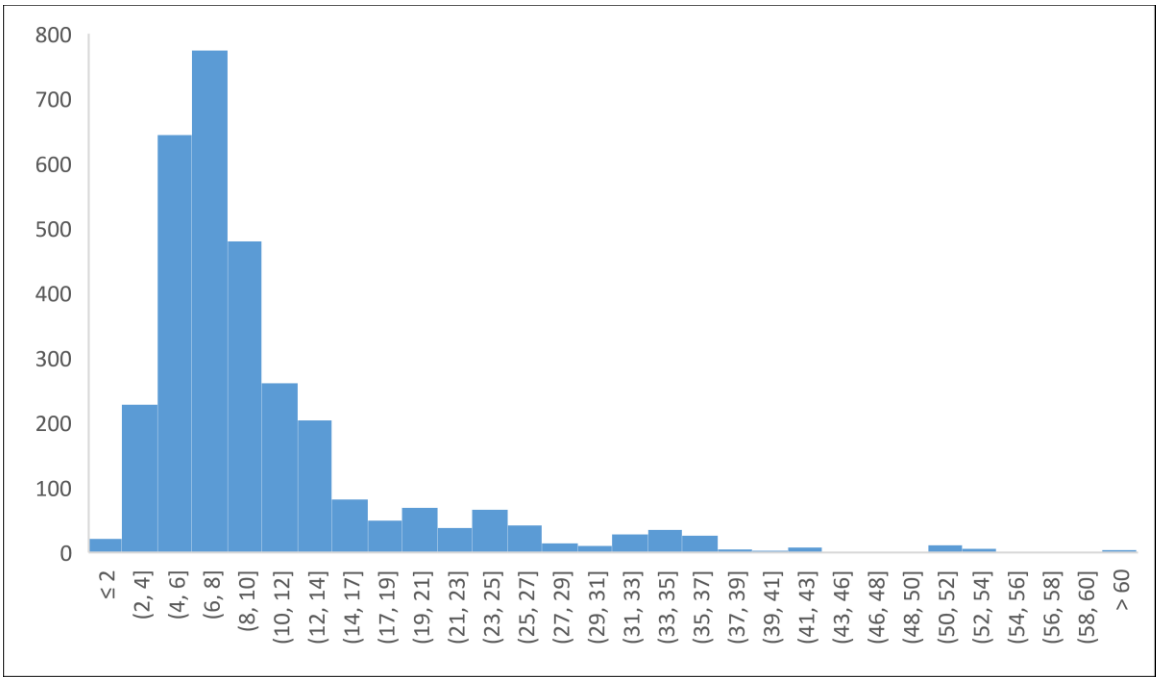

Figure 5.11 - Distribution of RT with p-median5 design (minutes)

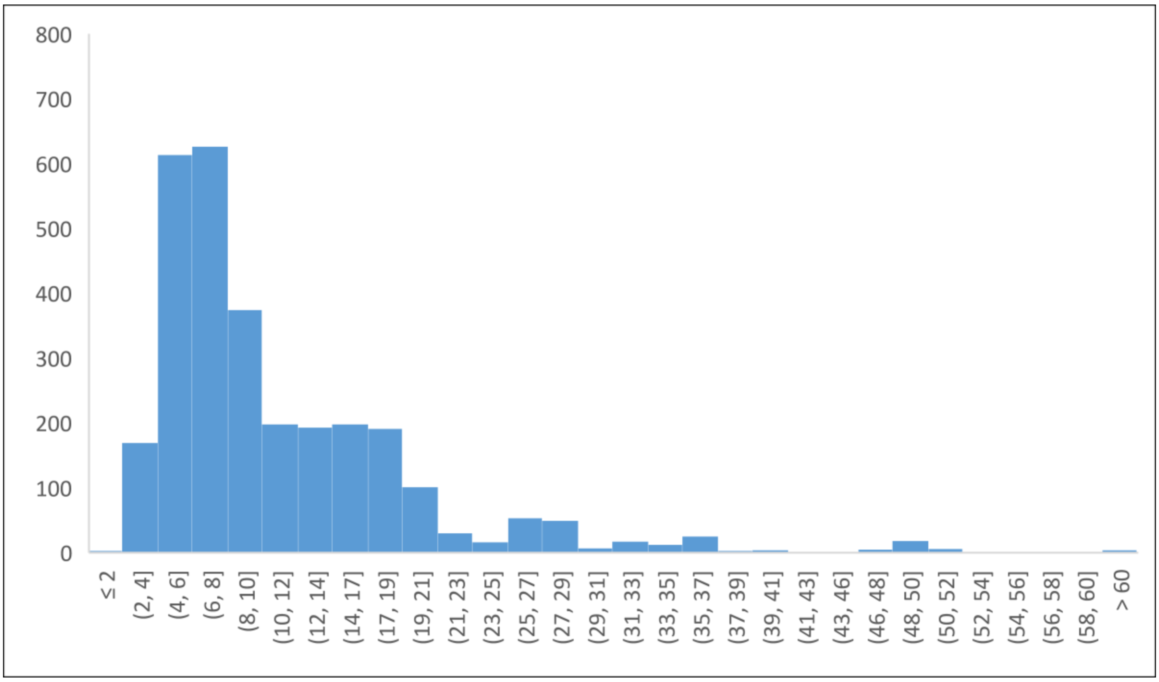

Figure 5.12 - Distribution of RT with MCLP design (minutes)

From the analysis of the histograms presented, it’s possible to notice their resemblance with a Poisson distribution. It makes sense considering that the events (emergency calls) appear independently in space and occur at a given average rate, while the ambulances attempt to uniformly cover that area.

The flat section of the MCLP distribution between the 10 and 19 minutes explains why this layout underperformed in covering 75% of the population.

Up to this point, the p-median5 and the MCLP designs were outpointed as the strongest ones, thus, further improvements were based on them, by analysing more simulation data.

5.5 Iterations maintaining the number of resources

In this sub-section, it is intended to relocate existing ambulances in the designs, in the pursuance of achieving better results. The choice of which to move is supported by the data gathered during the simulation runs on the utilization of resources and their positioning to demand.

It was decided that the first vehicle to relocate would be from the station with the least rescues made. This would be placed in the station in greatest need of resources, the highest number of times busy, if this was greater than the number of rescues. An ambulance was considered to be more valuable where it would've saved more lives. Furthermore, it was tested the allocation of that same vehicle to the station, previously unutilized, with the highest number of times being the closest to an incident. In these cases, this station would have been the one deploying if it had resources.

5.5.1 Relocation in MCLP design

As observed in Table 5.7, the ambulance more suitable to be relocated is of station 12, since it showed the lowest count of rescues. However, this number is still greater than the value the vehicle would bring by filling the 47 times that station 28 could not respond to victims. Therefore, the only relocation adopted in the MCLP design was to place the mentioned resource in the station 37, which would have been the first to respond to 196 emergencies if it had resources available.

Table 5.7 - Number of rescues each station made and number of times it was busy to respond (ordered by value)

Station	Rescues		Station	Times busy
28	337		28	47
29	196		15	35
15	193		7	27
3	186		20	25
13	169		3	24
19	168		19	18
38	166		29	15
21	151		18	13
20	130		21	12
7	123		13	9
4	112		16	9
39	111		33	9
18	109		23	6
16	101		10	5
11	98		38	4
10	91		12	4
23	91		39	3
5	86		4	2
33	78		5	2
9	74		8	2
17	68		11	1
8	59		9	0
12	49		17	0

Table 5.8 - Number of times each station not in use was the closest to victim

Station	Times being closest
37	196
14	157
6	147
2	95
30	95
25	90
22	82
34	54
26	42
35	40
36	39
27	38
1	0
24	0
31	0
32	0

The results for the RT of the MCLP design with an ambulance relocated from station 12 to station 37 are presented in Table 5.9 and compared with the original layout in Table 5.10.

Besides the metrics of maximum time and percentile 0.99, which aren't very indicative of the overall performance, the new design performed poorly. This is explained by observing that station 12, from which a vehicle was removed, lies in a rural area, where stations are more spread apart. Thus, when victims appear near it, and station 12 can't respond, the nearest resource takes a long time to reach the emergency site. Whereas the station 37 is located in the central urban area, with lots of demand, explaining the high number of nearby incidents, but, when it happens to be busy, the second or third nearest helps are very close still, due to the high density of stations there.

This relocation attempted, thus, results in more RTs in the upper values added to the set, while the time gains in the city aren't justified.

Table 5.9 - Results for RT with MCLP design with ambulance relocation to station 37

Model	avg	max	0.25	0.50	0.75	0.90	0.95	0.99
MCLP_37	12	64	7	10	16	24	31	41

Table 5.10 - Comparison with original MCLP (new/original)

Model	avg	max	0.25	0.50	0.75	0.90	0.95	0.99
MCLP_37	110.8%	91.5%	109.9%	118.9%	110.3%	125.1%	113.8%	87.7%

5.5.2 Relocation in p-median5 design

The same method was followed for the p-median5 layout, though in this case there was a station which was busy more times than the number of rescues that the facility 8 did, as seen in Table 5.11. Therefore, two relocations of the resource of this last station were tested in simulation: to the station 13 and to site 22, which was unutilized and was the nearest facility for 182 times.

Table 5.11 - Number of rescues each station made and number of times it was busy to respond (ordered)

Station	Rescues		Station	Times busy
4	278		13	30
5	239		4	29
3	207		3	29
1	199		15	27
19	195		19	17
18	194		7	15
15	192		5	12
2	170		2	12
7	163		29	12
13	161		20	12
10	142		18	10
29	128		1	9
12	124		12	8
17	110		10	7
21	97		17	6
20	96		21	3
33	74		33	1
9	73		9	0
23	73		23	0
34	62		34	0
36	60		36	0
35	47		35	0
8	29		8	0

As observed in the comparison of Table 5.14, the results of the new iterations did not translate improvements, except in the top range of the RT set. Once again, a vehicle was removed from a station in a remote area, and placed in an urban zone, which hindered the rescues in the former location.

Table 5.12 - Number of times each station not in use was the closest to victim

Station	Times being closest
22	182
25	141
6	131
14	125
30	88
11	55
37	37
39	36
27	35
24	26
16	20
26	0
28	0
31	0
32	0
38	0

Table 5.13 - Results for RT with p-median5 designs with ambulance relocation to station 13 or 22

Model	avg	max	0.25	0.50	0.75	0.90	0.95	0.99
p-median5_13	12	61	6	10	16	24	27	39
p-median5_22	14	57	7	10	16	27	36	51

Table 5.14 - Comparison with original p-median5 (new/original)

Model	avg	max	0.25	0.50	0.75	0.90	0.95	0.99
p-median5_13	119.0%	71.0%	111.1%	128.7%	133.0%	119.2%	102.3%	94.5%
p-median5_22	130.5%	65.3%	123.3%	120.5%	136.9%	132.9%	132.8%	123.6%

The best allocation of ambulances reached with the resources the Hanoi EMS currently owns was thus the original p-median5 result. It was then the base for further improvements attempted, which suggested the addition of new vehicles to the fleet.

5.6 Iterations increasing fleet

As the starting point, the p-median5 design, is already performing well for the majority of the population – observed in the average and coverage values up to 75% of demand – it is a higher priority to improve the RT in the upper range.

To accomplish this, it was decided to try the addition of new vehicles in the more remote areas. The procedure was to examine the layout of facilities in the map, determine the most isolated one and select it to be allocated the new ambulance, to improve the coverage of the whole province.

This was done up to five new vehicles and the order of stations opened were 14, 16, 6, 11 and 37, though this later site is in the central demand zone already.

The results are displayed in Table 5.15 - Results for RT with increasing fleet and the evolution of the metrics can be followed in Figure 5.13.

Table 5.15 - Results for RT with increasing fleet

Design	avg	max	0.25	0.50	0.75	0.90	0.95	0.99
p-median5	10	87	6	8	12	20	27	41
+1	9	73	7	9	12	20	25	38
+2	9	71	6	8	12	18	24	36
+3	9	68	6	8	12	18	22	35
+4	9	57	6	8	12	17	21	32
+5	9	55	5	8	11	16	19	31
Improvement	90%	63%	83%	100%	92%	80%	70%	78%

As it can be observed, the average value and percentiles up to 75% remained fairly steady while the maximum and higher percentiles suffered great improvements. This is explained by the low demand that the newly added ambulances supply, for being placed in rural areas. While the responses to those are improved significantly, they are a very small portion of the population, thus not weighting much on the average but eliminating the long travel times previously necessary to reach those locations.

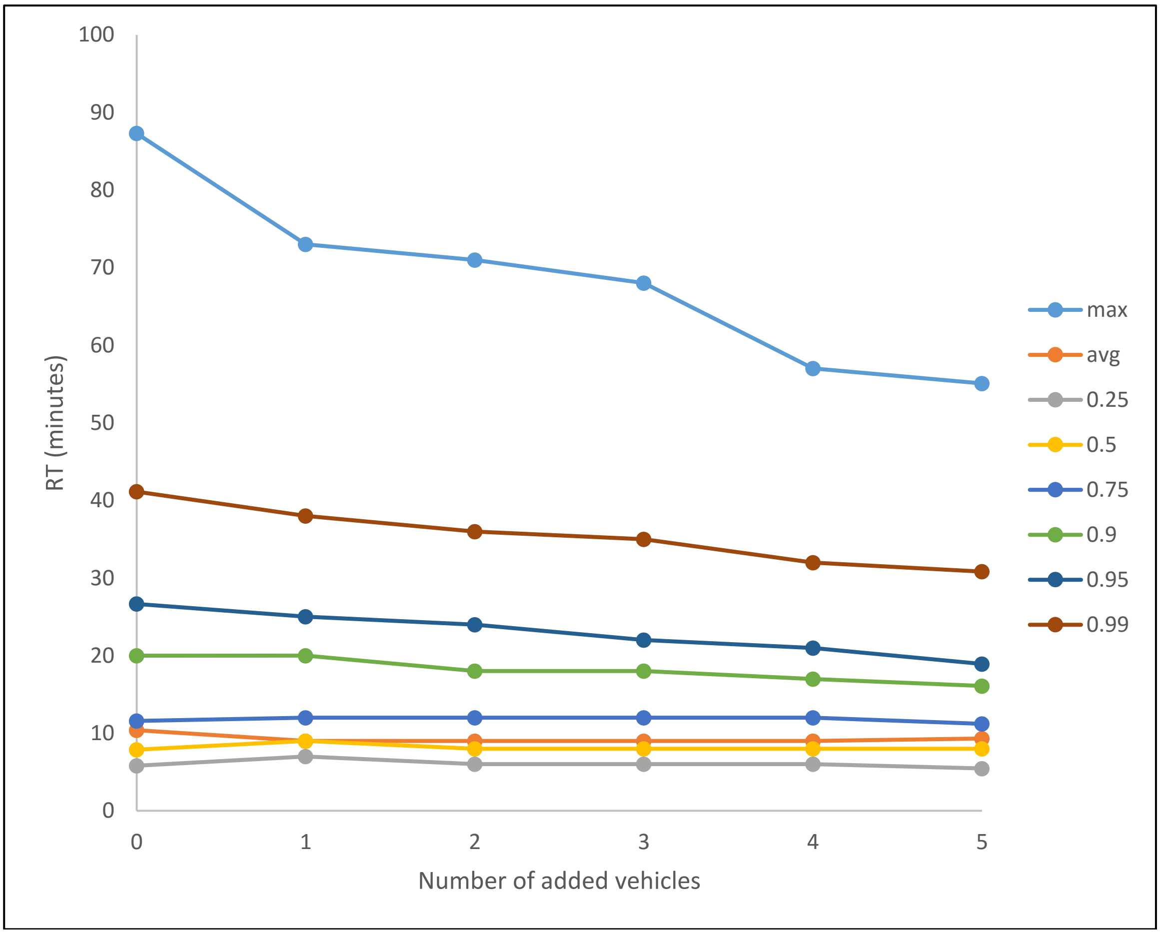

Figure 5.13 - Evolution of the different metrics with increasing fleet

5.7 Demand scenarios

As the demand limit was supported by information of the current EMS in Hanoi, it is prone to increase due to its elasticity or simply because of population growth.

Therefore, the p-median5 design and the one with the five extra ambulances were tested in a simulation producing twice as demand as originally.

As seen in Table 5.16, the performance of both layouts dropped with more demand to supply, as expected. However, they still revealed robustness by not underperforming by more than 30% in any metric, even though the number of calls increased by 100% (doubled).

Table 5.16 - Results for RT with p-median5 design the fleet increase design with double demand

Added vehicles	Demand	avg	max	0.25	0.5	0.75	0.9	0.95	0.99
+0	double	12	89	6	10	15	22	28	44
	original	10	87	6	8	12	20	27	41
Comparison:		120%	102%	100%	125%	125%	110%	104%	107%
+5	double	11	61	6	9	14	21	24	40
	original	9	55	5	8	11	16	19	31
Comparison:		122%	111%	120%	113%	127%	131%	126%	129%

5.8 Discussion of results

After many iterations, the design resulting from the formulation p-median with the constraint of utilizing the 5 original stations proved to be the best with 23 vehicles. It's possible to visualize the mentioned layout in Figure 5.14 - Stations used in the p-median5 design (in red), in which the stations in red are the ones with vehicles.

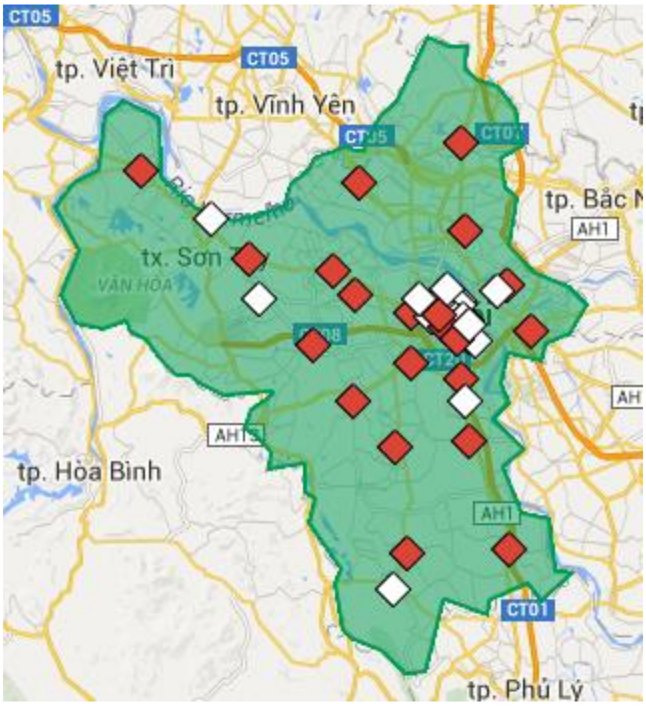

Figure 5.14 - Stations used in the p-median5 design (in red)

Due to the iterative process utilized, and the sources of stochasticity and error, this solution isn't known to be the optimal. It is, however, a feasible one which proved significantly better than the starting point scenario.

As an experiment to witness the value added by the acquisition of new vehicles, it was tested the introduction of up to 5 ambulances. These were intended to tackle demand in rural areas, where help is usually further away than in cities, and proved their worth by shrinking the times distribution. Meaning that they reduced the higher RTs and provided that a greater portion of the population is within a shorter reach of emergency teams.

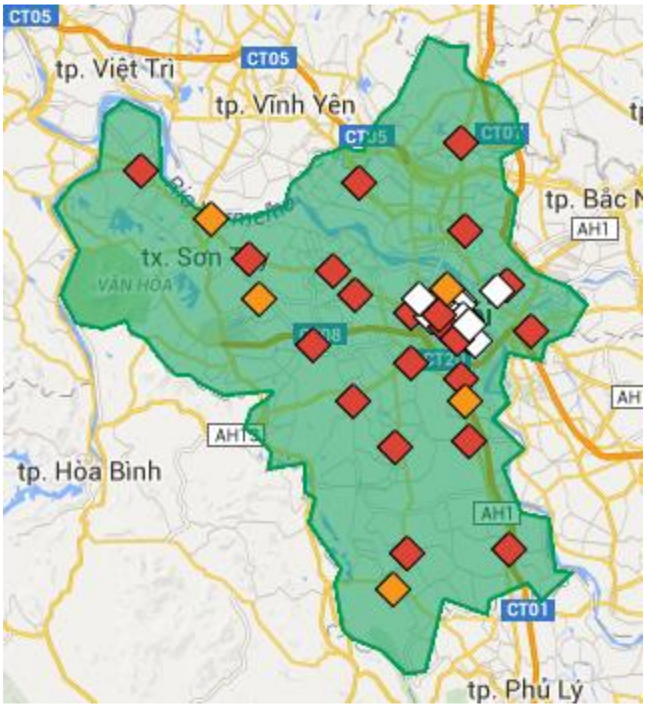

Figure 5.15 - P-median5 design with the added 5 stations (in orange)

6 Conclusions and recommendations

The objective of finding a better allocation for the 23 ambulances in the province of Hanoi in order to improve the service level of the EMS organization was reached. The advantages of expanding the network, in the sense of increasing the number of deployment stations, were proven, and in a short-term and low-capital manner.

It was demonstrated that, by adopting a design as the p-median5, rather than being restricted to the five EMS stations, the average time that the population would wait for an emergency response is 10 minutes instead of 20, and 90% of the people would be served in under 20 minutes, rather than 43, which is less than half.

As to the acquisition of new fleet, the added ambulances in more remote locations proved valuable in providing quicker rescues to the whole population. As in the case of adding five vehicles, 90% of the population became covered in less than 16 minutes, an improvement of 20%.

These scenarios were also exposed to an increase of 100% in demand, in the computer simulation, and proved robust still, though more studies must be conducted if a higher rate of calls is to be expected in the future.

It is urged that the EMS of Hanoi starts storing and analysing demand data, in order to model it and forecast future trends. This was a missing piece of this research and the more accurate the input data, the most valuable and applicable the outcome is. Thus it is of utmost importance that the organization in study improves their understanding about the population needs and builds a better service from there.

The cooperation with other existing facilities is highlighted as a very important and inexpensive step towards improving the current network.

As a starting point, it is recommended the use of the simulation tool produced to test iterations, in order to collect insights and help to support the decision making process in planning activities.

References

Aboueljinane, L., Sahin, E., Jemai, Z. (2013). A review on simulation models applied to emergency medical service operations. Ecole Centrale Paris, Laboratoire Génie Industriel, Grande Voie des Vignes, 92295 Châtenay-Malabry, France.

Andersson, T., & Värband, P. (2007). Decision support tools for ambulance dispatch and relocation. Journal of the Operational Research Society, 58(2), 195–201.

Berlin, G., ReVelle, C. and Elzinga, J. (1976). Determining ambulance-hospital locations for on-scene and hospital services. Environment and Planning A, 8, pp. 553-561.

Brandeau, M.L., Chiu, S.S., Kumar, S. and Grossman, T.A. (1995). Objective in location problems. In: Drezer, Z. (ed.), Facility Location: A Survey of Applications and Methods, Springer-Verlag, New York.

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

Carbone, R. (1974). Public facility location under stochastic demand, INFOR, 12.

Carson, Y. and Batta, R. (1990). Locating an ambulance on the Amherst campus of the State University of New York at Buffalo. Interfaces, 20, pp.43-49.

Church, R. and ReVelle, C. (1974). The maximal covering location problem. Papers of the Regional Science Association, 32, pp. 101-118.

Current, J., Daskin, M. and Schilling, D. (1995). Discrete network location models. In: Drezner, Z. and Hamacher, H.W. (ed.), Facility Location: Applications and Theory, Springer-Verlag, New York.

Daskin, M. (2008). What You Should Know About Location Modeling. Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208.

Daskin, M.S. (2000). A new approach to solving the vertex P-center problem to optimality: Algorithm and computational results. Communications of the Japanese OR Society, 9, pp. 428-436.

Daskin, M.S., 1983. A maximum expected location model: Formulation, properties and heuristic solution. Transportation Science 7, 48–70.

Fitzsimmons, J. A. (1971). An emergency medical system simulation model. In Proceedings of the 1971 winter simulation conference (pp. 18–25). New York, NY, USA: ACM.

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

Garfinkel, R. S., Neebe, A. W. and Rao, M. R. (1977). The m-center problem: Minimax facility location. Management Science, 23, pp. 1133-1142.

Gendreau, M., Laporte, G., Semet, F., 1997. Solving an ambulance location model by Tabu search. Locatio 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.

Goldberg, J., Dietrich, R., Chen, J. M., Mitwas, M., Valenzuela, T., & Criss, E. (1990). A simulation model for evaluating a set of emergency vehicle base locations: Development, validation, and usage. Socio-Economic Planning Sciences, 24(2), 125–141.

Goldberg, J., Dietrich, R., Chen, J.M., Mitwasi, M.G., 1990b. Validating and applying a model for locating emergency medical services in Tucson, AZ. European Journal of Operational Research 49, 308–324.

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

Henderson, S. G., & Mason, A. J. (2005). Ambulance service planning: simulation and data visualisation. In Brandeau, M.L., Sainfort, F., & Pierskalla, W.P. (Eds.),

Hogan, K. and ReVelle, C. (1986). Concepts and applications of backup coverage. Management Science, 32, pp.1434-1444.

Ingalls, R G, Rossetti, M D, Smith, J S and Peters, B A (Eds.), Winter Simulation Conference. Washington D.C.

Ingolfsson, A., Erkut, E., & Budge, S. (2003). Simulation of single start station for Edmonton EMS. The Journal of the Operational Research Society, 54(7), 736–746.

Iskander, W. H. (1989). Simulation modeling for emergency medical service systems. In Proceedings of the 1989 winter simulation conference (pp. 1107– 1111).

Laroque, J. Himmelspach, R. Pasupathy, O. Rose, & A.M. Uhrmacher (Eds.), Proceedings of the 2012 winter simulation conference.

Lubicz, M., & Mielczarek, B. (1987). Simulation modelling of emergency medical services. European Journal of Operational Research, 29(2), 178–185.

Mahajan P S and Ingalls R G (2004) Evaluation of methods used to detect warm-up period in steady state simulation. In

Mandell, M.B. (1998). A P-median approach to locating basic life support and advanced life support units. Presented at the CORS/INFORMS National Meeting, Montreal, April, 1998.

Operations research and health care (pp. 77–102). Boston: Kluwer Academic Publishers.

Paluzzi, M. (2004). Testing a heuristic P-median location allocation model for siting emergency service facilities. Paper Presented at the Annual Meeting of Association of American Geographers, Philadelphia, PA.

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. and Hogan, K. (1989b). The maximum reliability location problem and a-reliable p-center problem: Derivatives of the probabilistic location set covering problem. Annals of Operations Research, 18, pp. 155-174.

ReVelle, C. and Swain, R.W. (1970). Central facilities location. Geographical Analysis, 2, pp. 30-42.

ReVelle, C.S., Hogan, K., 1989. The maximum availability location problem. Transportation Science 23, 192–200.

Savas, E. S. (1969). Simulation and cost-effectiveness analysis of New York’s emergency ambulance service. Management Science, 15(12), 608–627.

Schilling, D., Elzinga, D., Cohon, J., Church, R. and ReVelle, C. (1979). The TEAM/FLEET models for simultaneous facility and equipment siting. Transportation Science, 13, pp. 163 175.

Sylvester, J. (1857). A Question in the Geometry of Situation. Quarterly Journal of Pure and Applied Mathematics, 1, 1857.

Talmar, M. (2002). Location of rescue helicopters in South Tyrol. Paper Presented at 37th Annual ORSNZ Conference, Auckland, New Zealand.

Toregas, C., Swain, R., ReVelle, C. and Bergman, L. (1971). The location of emergency service facility. Operations Research, 19, pp. 1363-1373.

Van Buuren, M., Aardal, K., Van der Mei, R., & Post, H. (2012). Evaluating dynamic dispatch strategies for emergency medical services: Tifar simulation tool. In C.

Wears, R. L., & Winton, C. N. (1993). Simulation modeling of prehospital trauma care. In G. W. Evans, M. Mollaghasemi, E.C. Russel, & W.E. Biles (Eds.), Proceedings of the 1993 winter simulation conference.

Appendix A: Network nodes

Table 6.1 - Possible deployment stations and their coordinates

Code	Station / facility name	Coordinates
1	Trung Tâm Cấp cứu 115	21.0222965, 105.8567074
2	Bệnh viện Thanh Trì	20.9486239, 105.8464602
3	Trung tâm Y tế Quận Long Biên	21.06865, 105.9104165
4	Trung Tâm Y Tế Hà Đông	20.9679609, 105.779196599999
5	Bệnh viện Đa khoa Y học cổ truyền	21.0342569, 105.779242999999
6	Bệnh viện đa khoa huyện Mỹ Đức	20.678369, 105.7547394
7	Bệnh viện Vân Đình	20.7253476, 105.7739645
8	Bệnh viện huyện Phú Xuyên	20.730711, 105.9147299
9	Bệnh viện huyện Thanh Oai	20.8621361999999, 105.757306999999
10	Bệnh viện Đa khoa Thường Tín	20.8680150999999, 105.8589256
11	Bệnh viện Da Khoa Nông Nghiệp	20.9211639, 105.8534329
12	Bệnh Viện Đa Khoa Huyện Quốc Oai	20.9910523, 105.6441396
13	bệnh viện đa khoa ba vì	21.2154603, 105.409355199999
14	Bệnh viện Sơn Tây	21.1527429, 105.5055714
15	Phuc Tho District Hospital	21.1036791, 105.5570161
16	Bệnh viện đa khoa Thạch Thất	21.0509625, 105.5701972
17	Bệnh viện Đa khoa Đan Phượng	21.0871188, 105.6709467
18	Bệnh viện huyện Hoài Đức	21.0564757, 105.701929
19	Bệnh viện Đa khoa Sóc Sơn	21.2499075, 105.847194199999
20	Bệnh viện Đa Khoa khu vực Mê Linh	21.1996982, 105.706879299999
21	Bệnh viện Đông Anh	21.1392295, 105.8530659
22	Bệnh viện Đa khoa Đức Giang	21.0611986, 105.8983943
23	Bệnh viện Đa khoa Gia Lâm	21.0095778, 105.9441008
24	Bệnh viện đa khoa Medlatec	21.0485391, 105.8461052
25	BỆNH VIỆN ĐA KHOA HỒNG NGỌC	21.0424817, 105.8441528
26	Saint Paul Municipal Hospital	21.0311944, 105.8350844
27	Bệnh viện Phụ sản Hà Nội	21.026959, 105.8071981
28	Bệnh viện Đa khoa Quốc tế Thu Cúc	21.0450336, 105.814457299999
29	Bệnh viện Thanh Nhàn	21.0036379, 105.8591429
30	Vinmec International Hospital	20.9961247, 105.8668284
31	National Hospital of Traditional Medicine	21.015892, 105.848636599999
32	Vietnam Cuba Hospital	21.0246076, 105.8507093
33	Bệnh viện huyện Chương Mỹ	20.9209337, 105.6982622
34	Bệnh viện Đống Đa	21.0156617, 105.827011599999
35	Bach Mai Hospital	20.999138, 105.841168899999
36	Family Medical Practice Hanoi	21.031104, 105.818295
37	International SOS Medical and Dental Clinic	21.0637718, 105.8273497
38	Military Hospital 108	21.0186139, 105.859758899999
39	E Hospital	21.0504128, 105.7892809

Table 6.2 - Number of nodes used per district and their demand

District	Area (km^2)	Population	Density (pop/ km^2)	Nodes	Demand/node
Hoàn Kiếm	5.29	147334	27851	1	147334
Thanh Xuân	9.11	223694	24555	1	223694
Ba Đình	9.22	225910	24502	1	225910
Hai Bà Trưng	9.60	370726	38617	1	370726
Đống Đa	9.96	410117	41176	1	410117
Cầu Giấy	12.04	260643	21648	1	260643
Tây Hồ	24.00	130639	5443	1	130639
Nam Từ Liêm	32.27	232894	7217	2	116447
Hoàng Mai	41.04	380509	9272	2	190255
Bắc Từ Liêm	43.35	320414	7391	2	160207
Hà Đông	47.91	260136	5430	2	130068
Long Biên	60.38	271913	4503	3	90638
Thanh Trì	68.22	198706	2913	3	66235
Đan Phượng	76.80	142480	1855	4	35620
Hoài Đức	95.30	191106	2005	4	47777
Phúc Thọ	113.20	159484	1409	5	31897
Sơn Tây	113.47	125749	1108	5	25150
Gia Lâm	114.00	251735	2208	5	50347
Thường Tín	127.70	219248	1717	6	36541
Thanh Oai	129.60	167250	1291	6	27875
Mê Linh	141.26	191490	1356	6	31915
Quốc Oai	147.00	160190	1090	6	26698
Phú Xuyên	171.10	181388	1060	7	25913
Đông Anh	182.30	333337	1829	8	41667
Ứng Hòa	183.72	182008	991	8	22751
Thạch Thất	202.50	177545	877	9	19727
Mỹ Đức	230.00	169999	739	10	17000
Chương Mỹ	232.90	286359	1230	10	28636
Sóc Sơn	306.74	282536	921	13	21734
Ba Vì	428.00	246120	575	18	13673

Appendix B: Python script

The following script, written in Python, was used to create the duration matrix with the Google API Distance Matrix. The coordinates of both origins and destinations were imported from two separate text files.

```
import googlemaps

with open('C:/Users/Rui/Desktop/origins.txt') as f:
    lista_origens = f.read().splitlines()

with open('C:/Users/Rui/Desktop/destinations.txt') as f:
    lista_destinos = f.read().splitlines()

gmaps = googlemaps.Client(key='AIzaSyBqOeR_O-thK65ATyli9fd_OCaGGm-GmzE')

vector = []

for i in range(0, 151):
    vector = []
    for origem in lista_origens:
        vector.append(gmaps.distance_matrix(origem,lista_destinos[i])['rows'][0]['elements'][
    0]['duration']['value'])
    print(vector)
```

Appendix C: Mathematical modelling results

Table 1 - Ambulance allocation results for each model

station code	p-median	p-median5	p-centre	p-centre5	MCLP	MCLP5	DSM	DSM5
1	0	1	0	1	0	1	0	1
2	1	1	0	1	0	1	0	1
3	0	1	0	1	1	1	1	1
4	1	1	0	1	1	1	1	1
5	1	1	0	1	1	1	2	1
6	0	0	0	1	0	0	0	0
7	1	1	0	0	1	1	1	1
8	1	1	0	0	1	1	1	1
9	1	1	1	0	1	1	1	1
10	1	1	0	0	1	1	1	1
11	0	0	1	1	1	0	1	0
12	1	1	1	1	1	1	1	1
13	1	1	0	1	1	1	1	1
14	0	0	0	0	0	0	0	0
15	1	1	1	0	1	1	1	1
16	0	0	1	0	1	1	0	0
17	1	1	1	0	1	1	1	1
18	1	1	0	0	1	1	1	1
19	1	1	0	1	1	1	1	1
20	1	1	0	0	1	1	1	1
21	1	1	0	1	1	1	1	1
22	1	0	0	1	0	0	0	0
23	1	1	1	1	1	1	1	1
24	0	0	1	1	0	0	0	0
25	0	0	1	1	0	0	0	0
26	0	0	1	1	0	0	0	0
27	0	0	1	1	0	0	0	0
28	0	0	1	0	1	1	1	1
29	1	1	1	1	1	1	1	1
30	0	0	1	1	0	0	0	0
31	0	0	1	1	0	0	1	0
32	1	0	1	1	0	0	0	0
33	1	1	1	0	1	1	1	1
34	1	1	1	1	0	0	0	0
35	1	1	1	0	0	0	0	1
36	1	1	1	0	0	0	0	0
37	0	0	1	0	0	0	0	0
38	0	0	1	0	1	0	0	0
39	0	0	1	1	1	1	1	1

Table 2 - Objective function result for each model

p-median	p-median5	p-centre	p-centre5	MCLP	MCLP5	DSM	DSM5
11.5*	11.5*	61.5	61.5	5224357	5224357	2372515	2242002

*The values presented for the p-median problems are the objective function (sum of all distances) divided by the demand in order to show the average distance.

The units of time are minutes while the covering models' functions represent population.