AN INNOVATIVE APPROACH FOR MODELING MULTI-FACILITY LOCATION ALLOCATIONS IN EMERGENCY MEDICAL SERVICE SYSTEMS

ENAS FARES

Master of Science in Engineering Management
Qatar University
2014
AN INNOVATIVE APPROACH FOR MODELING MULTI-FACILITY LOCATION ALLOCATIONS IN EMERGENCY MEDICAL SERVICE SYSTEMS

By

ENAS FARES

Thesis Submitted to the School of Research and Graduate Studies in Partial fulfillment of the Requirements for the Degree of Master of Science in Engineering Management

July 2014
DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at Qatar University or other institutions.

Signature : -------------------------------------------

Name : Enas H. Fares

ID No. : 200550121

Date : _____________________________
Permission to Make Photocopies of Report

I, Enas H. Fares, declare that the report entitled

An Innovative Approach for Modeling Multi-Facility Location Allocations in Emergency Medical Service Systems

belongs to me. The contents of this report may be used by anyone for academic purposes of teaching, learning, and research, only. Qatar University is permitted to make photocopies of this document for the same academic purposes.

24th July 2014

Name: Enas Fares
DEDICATION

To my beloved mother, Laila Fares, who without her wisdom, patients, love, and guidance I would not be able to continue my education and achieve my ambitions.

To my lovely sisters; Elham, Eman, and Hanan, who were always giving me positive feelings and encouragement whenever I need.

To all of my relatives and friends in and outside the state of Qatar who were with me during this long and fast journey!
ACKNOWLEDGMENT

I would like to express my appreciation to several individuals and organizations who without their support the completion of this thesis was impossible.

First of all, I would like to express my gratitude and thankfulness to my beloved mother, Laila Fares, for her continuous support maternity and love. I would like to express my deepest gratitude to my supervisor, Dr. Farayi Musharavati, who continuously supported me to achieve the goal of completing this thesis and encouraged me to develop my skills as an engineering researcher. I also would like to thank Dr. Nasser Ayoub for providing me with great help on the Matlab code that I used in my analysis. Last but not least, I would like also to thank; the Department of Statistics at Qatar University, the Ministry of Municipality and Urban Planning, the Ministry of Development Planning and Statistics, and Ambulance Services at Hamad Medical Corporation for their support and willingness to provide me with the required data.
ABSTRACT

Emergency Medical Service (EMS) facility location allocation is a critical strategic planning task that can undermine the integrity of EMS providers. As such, the planning of EMS systems must be approached with utmost caution. In the operations of an EMS, inappropriate facility location and allocation decisions usually lead to dramatic consequences. For example, one additional minute can result in the loss of human life. Therefore, there is a need to ensure that EMS facilities are located in the right places so that they can provide the service at the right time. While a lot of models have been proposed for locating and allocating EMS facilities, this research investigates the prospects of integrating data mining techniques and GIS simulation modeling in developing an innovative approach for modeling EMS systems. More specifically, a fuzzy k-means method is used in conjunction with GIS to model an EMS system. In this thesis, the EMS location allocation problem is cast as a multi-facility location allocation problem. Data mining techniques are applied to extract location allocation information from clustered data. The location allocation information is then verified and validated through GIS simulation. Further GIS simulations were then used to assess the performance of the developed EMS system models. The EMS system in Qatar was used as a case study. The performance of eight model-driven strategies was evaluated against a number of performance metrics. Although no one model-driven strategy performed best in all metrics considered the overall rating of the model developed by combining a number of methods demonstrates the importance of combining data and information from a variety of sources in developing a suitable EMS model. Results also show that the fuzzy k-means method is equally capable of generating EMS clusters and providing the spatial locations of the EMS units. Therefore, combining a number of methods in developing an EMS system results in an EMS system whose overall performance is superior. These results indicate that the innovative approach is capable of recommending a more effective and more efficient location-allocation model for emergency medical services.
# TABLE OF CONTENT:

DECLARATION......................................................................................................................................................... III

PERMISSION TO MAKE PHOTOCOPIES OF REPORT ................................................................. IV

DEDICATION................................................................................................................................................................. V

ACKNOWLEDGMENT ...................................................................................................................................................... VI

ABSTRACT ................................................................................................................................................................. VII

TABLE OF CONTENT: ........................................................................................................................................ VIII

LIST OF TABLES .................................................................................................................................................. XI

LIST OF FIGURES ........................................................................................................................................ XII

CHAPTER 1 .............................................................................................................................................................. 1

INTRODUCTION............................................................................................................................................................ 1

1.1. BACKGROUND.................................................................................................................................................. 2
  1.1.1. EMS Capacity .............................................................................................................................................. 3
  1.1.2. Equity .................................................................................................................................................................. 4
  1.1.3. Response Rate and Time ............................................................................................................................ 4
  1.1.4. Survival Rate................................................................................................................................................. 4

1.2. MULTI-FACILITY LOCATION ALLOCATION .............................................................................................. 5

1.3. FACILITY LOCATION ALLOCATION MODELS .......................................................................................... 6

1.4. PROBLEM DESCRIPTION................................................................................................................................... 8

1.5. OBJECTIVES.................................................................................................................................................... 9

1.6. SCOPE .................................................................................................................................................................. 10

1.7. OUTLINE OF THIS STUDY .......................................................................................................................... 11

CHAPTER 2 ............................................................................................................................................................. 12

LITERATURE REVIEW ........................................................................................................................................... 12

2.1. EMS PERFORMANCE INDICATORS ........................................................................................................... 13

2.2. LOCATION ALLOCATION MODELS .......................................................................................................... 15
  2.2.1. Continuous and Discrete Location Models................................................................................................. 15
  2.2.2. Covering Models .......................................................................................................................................... 16
    2.2.2.1. Set Covering Problem (SCP).................................................................................................................. 16
    2.2.2.2. Maximal covering location problem (MCLP) ............................................................................................ 17
  2.2.3. P-center Models............................................................................................................................................ 17
  2.2.4. P-median Models ........................................................................................................................................ 18

2.3. LOCATION ALLOCATION MODELS RELATED TO EMS ............................................................................. 19

2.4. KNOWLEDGE DISCOVERY FROM DATA ................................................................................................. 21
CHAPTER 3 .............................................................................................................................................. 34

RESEARCH METHODOLOGY ....................................................................................................................... 34

3.1. OVERVIEW OF THE RESEARCH PROCESS .................................................................................... 35
3.2. TOOLS USED ...................................................................................................................................... 36
3.3. CASE STUDY ...................................................................................................................................... 37
   3.3.1. Study Area ................................................................................................................................. 37
   3.3.2. Case Study EMS ......................................................................................................................... 40
   3.3.3. Data Collection ........................................................................................................................ 44
3.4. METHODS USED TO ACHIEVE THE OBJECTIVES ........................................................................ 45
   3.4.1 Data Analysis ............................................................................................................................ 45
   3.4.1.1. GIS Data Mapping ................................................................................................................ 45
   3.4.1.2. Statistical Analysis ........................................................................................................... 45
   3.4.2. GIS Network Simulation .......................................................................................................... 47
   3.4.2.1. Service Area Analysis ........................................................................................................ 48
   3.4.2.2. Closest Facility Location ..................................................................................................... 48
   3.4.3. Integrated use of Data Mining and GIS Methods ...................................................................... 49
   3.4.3.1. Data Clustering ................................................................................................................... 50
   3.4.3.2. Fuzzy Clustering Method .................................................................................................. 50
   3.4.3.3. Fuzzy k-Means Algorithms ............................................................................................... 52
   3.4.3.4. Constraints and Assumptions ............................................................................................ 55
   3.4.3.5. Verification and Validation of Clustering Techniques ......................................................... 56
   3.4.3.6. GIS Method ....................................................................................................................... 62
   3.4.3.7. Hybrid Clustering and GIS Simulation ............................................................................... 63

CHAPTER 4 .................................................................................................................................................... 66

RESULTS AND DISCUSSIONS ...................................................................................................................... 66

4.1. DATA ANALYSIS ............................................................................................................................... 66
   4.1.1. Response Times .......................................................................................................................... 67
   4.1.2. Pattern of Incidents .................................................................................................................... 68
   4.1.3. EMS Availability ....................................................................................................................... 69
   4.1.4. Incidents-Zone Analysis ........................................................................................................... 69
4.2. ANALYSIS OF ALTERNATIVE EMS SYSTEMS ............................................................................. 76
   4.2.1. Existing EMS System ............................................................................................................... 76
   4.2.1.1. Service Area Analysis ......................................................................................................... 76
   4.2.1.2. Closest Facility Analysis .................................................................................................... 78
   4.2.2. Proposed GIS Based EMS System (Updated Existing System) ................................................ 80
   4.2.2.1. Service Area Analysis ......................................................................................................... 82
   4.2.2.2. Closest Facility Analysis .................................................................................................... 84
   4.2.3. New System Based on Fuzzy k-means .................................................................................... 85
List of Tables

Table 1: General Clustering Analysis Methods ................................................................. 23
Table 2: Summary of Research Papers .............................................................................. 30
Table 3: Resources Summary of EMS System at State of Qatar ........................................ 43
Table 4: EMS Facility and Time Required to Reach a Patient ............................................. 52
Table 5: Distance and Percent Difference for Each Clustering Algorithm ...................... 59
Table 6: Zones That Have More Than 1% Occurrence of Incidents ................................. 72
Table 7: Summary of Service Area Analysis – Existing System ..................................... 78
Table 8: Closest Facility Results of the Existing System (minutes) .................................... 80
Table 9: Summary of Service Area Analysis – Updated Existing System ....................... 83
Table 10: Closest Facility Results of the Updated Existing System (minutes) .................... 84
Table 11: Summary of Service Area Analysis – fuzzy k-means Clustering System ......... 89
Table 12: Closest Facility Results of the Fuzzy k-means Clustering System (minutes) .......... 90
Table 13: Closest Facility Results for the Existing, Updated, and Fuzzy k-means Clustering Systems- Population ................................................................. 94
Table 14: Closest Facility Results for the Existing, Updated, and Fuzzy k-means Clustering Systems- Incidents .............................................................................. 94
Table 15: Description of Model Strategies Simulated in the GIS Environment .............. 97
Table 16: Time Execution of Matlab Code ......................................................................... 98
Table 17: Number of Family Members in Qatar ................................................................. 99
Table 18: Uncovered Population- (Indirect-NoH&S-Incident) and Comprehensive Models .... 106
Table 19: Closest Facility Results - Population ................................................................. 108
Table 20: Closest Facility Results - Incidents .................................................................... 108
Table 21: Closest Facility Results in Rural Areas- Population .......................................... 111
Table 22: Closest Facility Results in Rural Areas - Incidents ............................................ 111
Table 23: Models Ranking ............................................................................................... 113
List of Figures

Figure 1: Spanish EMS facilities ................................................................. 3
Figure 2: EMS Multi-facilities .................................................................. 6
Figure 3: The survival rate versus the time for cardiac arrest patients ....... 14
Figure 4: Knowledge Discovery from Data Steps ................................. 22
Figure 5: Main Stakeholders of the EMS System in State of Qatar ....... 35
Figure 6: Research Process ..................................................................... 36
Figure 7: Location and Distribution of Population and Road Network in Qatar 38
Figure 8: Distribution of emergency incidents in Qatar for the year 2013/2014 ... 39
Figure 9: Summer Response Locations – 2013 ......................................... 40
Figure 10: Winter Response Locations – 2013 ......................................... 41
Figure 11: EMS Facilities and Staffs in State of Qatar ............................. 42
Figure 12: GIS Layer Required/Generated for/from the Research .......... 46
Figure 13: Statistical Method Flowchart .................................................. 47
Figure 14: Fuzzy k-means Clustering Algorithm ...................................... 54
Figure 15: Verification and Validation ...................................................... 56
Figure 16: Validation of Fuzzy k-means Clustering ................................. 58
Figure 17: Validating Number of Clusters - In Doha ............................... 61
Figure 18: Validating Number of Clusters - Out Doha ............................ 61
Figure 19: Flowchart for Implementing the Hybrid Clustering GIS Simulation 65
Figure 20: Box Plot for Response Time .................................................... 67
Figure 21: Incidents Over a Year .............................................................. 68
Figure 22: Number of Incidents in Hourly Bases ..................................... 69
Figure 23: Available EMS Units Every Hour .......................................... 70
Figure 24: Zone Incidents Density ............................................................ 71
Figure 25: Number of Incidents s. Zones ................................................ 73
Figure 26: Number of Incidents For Each Response Priority - 2012/2013 .... 74
Figure 27: Total Number of Cases for Each Class - 2012/2013 .............. 74
Figure 28: EMS Usage with Respect to Case Classification ................. 75
Figure 29: EMS Usage with Respect to Response Priority .................... 75
Figure 30: EMS Locations of Existing System ....................................... 77
Figure 31: Service Areas of the Existing System .................................... 79
Figure 32: Location of EMS units based on p-median and maximize coverage models 81
Figure 33: Network Analyst Message ..................................................... 82
Figure 34: Service Areas of the Updated EMS Units of the Existing System 83
Figure 35: Matlab Graph of the Various Clusters and Center of Clusters for the EMS System Proposed by the Fuzzy k-means Algorithm ..................... 87
Figure 36: EMS Locations Based on Integrated Use of Fuzzy k-means Clustering and GIS ...... 88
Figure 37: Service Areas of The Fuzzy k-means Clustering System .......... 91
Figure 38: Percentage of Covered Population for the Three Alternative EMS Systems 92
Figure 39: Percentage of Covered Districts - Existing, Updated, and Fuzzy k-means Systems .. 92
Figure 40: Percentage of Covered Populated Districts - Existing, Updated, and Fuzzy k-means Systems ......................................................................................................................................... 93
Figure 41: Proposed Systems............................................................................................................................................................................ 96
Figure 42: EMS Locations of Existing and Proposed Systems ................................................................. 101
Figure 43: Districts in and Out Doha .............................................................................................................. 102
Figure 44: Service Area of Proposed and Existing Models ................................................................. 104
Figure 45: Percentage of Uncovered Districts .......................................................................................... 105
Figure 46: Number of Uncovered Population .......................................................................................... 105
Figure 47: Number of Uncovered Population-Filtered ........................................................................... 106
Figure 48: Percentage of Covered Populated Districts .............................................................................. 107
Figure 49: Results of Kruskal-Wallis Test ............................................................................................................. 110
CHAPTER 1

INTRODUCTION

Facility location allocation models provide an optimal location and allocation of facilities in order to supply a set of demand points. The literature in this field is vast since facility location allocation is considered a critical element in strategic planning for both private and public sectors. In many cases, inappropriate facility location or wrong allocation decisions lead to dramatic consequences particularly for those systems that are related to human life such as emergency medical service (EMS) systems. There is, therefore, no room for mistakes in the planning and operations of EMS systems.

The focus of this thesis revolves around the effective and efficient planning of EMS systems at strategic level. This planning involves determining the locations of EMS facilities (these can be permanent, temporary or both) as well as allocating EMS facilities to demand points. The challenge in this undertaking lies in that the success of a proposed EMS system is usually revealed at the operational stage. While it is true that wrong decisions at the operational level can result in poor performance, it is also true that a poorly planned EMS system will similarly affect operational issues. Therefore, the need for implementing an EMS that is planned for optimality can never be over emphasized.

In addition to the discussion in the previous paragraph, the decision making process for EMS systems in the public sector is compounded by a number of factors. Firstly, the strategic nature of the decision making process will force operations to deal with, say, a poorly planned EMS system for a long time before a decision to review is due. In cases where there is rapid population growth, rapid expansions of infrastructure (residential or industrial), and random locations/re-locations of residential, commercial and industrial areas the “current” EMS system may be overwhelmed. In such and similar cases simulation and predictive tools can be used to appraise the “current” system. Therefore simulation is a very important tool in the strategic planning and decision making process of an EMS system. Secondly, for public EMS systems, budget availability or unavailability may result in undesirable situation. If the EMS budget is adequate, there is a tendency to over design the EMS in order to account for uncertainties. On the
other hand, the EMS may be under designed due to budget constraints. Above all, the challenges of planning for an optimal EMS system are made even more difficult by the uncertain nature of EMS system dynamics in a particular environment. Whatever the case maybe, EMS demand will continue to increase. Therefore, EMS planners have to review and improve the EMS system continuously. This leads to a number of themes that will be investigated in this thesis.

Elsewhere, it has been claimed that one way of improving an existing EMS system is to find an optimal re-location of EMS facilities. Since this re-location depends on the original design of the EMS in a specific environment and under different conditions and constraints, this thesis will investigate the effects of re-locating EMS facilities on the performance of an existing EMS system. In addition, this thesis will argue that with effective and efficient planning it is possible to achieve better EMS performance without increasing costs (i.e. without increasing the number of EMS units). Since different, methods, models, tools and techniques can be used to strategically determine an optimal location and allocation of EMS units, this thesis will also assess the effects of applying different methods, models, tools and techniques on the performance of the resulting EMS system. More details about EMS systems and location sciences are provided in the following sub-sections.

1.1. Background

Emergency Medical Services (EMS) systems provide the best possible means of pre-hospital or out-of-hospital medical care to a variety of patience. Once a sudden life threatening incident occurs, EMS providers will receive a call regarding that incident. In response to such a call, EMS providers initiate a series of actions which includes the dispatching of service facilities, triage, emergency medical response, and transportation by the EMS vehicle (ambulance or helicopter) to the nearest hospital/medical facility (Committee on the Future of Emergency Care in the United States Health System Board on Health Care Services, 2006). This process is critical since the life of the patient in need of EMS maybe lost.
Historically, EMS can be traced back to the Greek and Roman eras, where injured soldiers were taken-out from the battlefield to a safe place where they received treatment. At that time EMS was only responsible for transporting the injured soldiers. This concept has evolved since then. Nowadays, EMS is capable of providing the required treatment for seriously injured people while they are being moved to a hospital or medical facility.

It is worth mentioning that EMS strategies vary from one country to another. For instance, in most of low-income and developing countries, the EMS system functions only as a transportation facility where it moves the patients to a health care facility without providing any medical services. In developed counties a basic or even an advanced life support is provided during the transportation (Jamison, et al., 2006; Wikipedia, 2013).

Based on the discussion in the previous paragraph, EMS vehicles can be classified into four main types; namely, (i) the non-assistance ambulances used only to transport patients (these are not specially equipped to provide medical assistance), (ii) the basic life support (BLS) assistance ambulances which are equipped with basic life support equipment to help in increasing the survival rate, (iii) the advanced life support (ALS) ambulances (which are provided with equipment which allow the staff to even perform surgery while moving the patient), and (iv) the health emergency helicopters (air vehicles used to reach to the places that are far away from the hospitals and they are also used for fast transfer of patience with serious cases). The type of transportation deployed usually depends on the nature of the request for EMS. EMS response to requests is affected by; (a) the availability of EMS facilities, (b) current locations of the EMS facilities, (c) policies regarding deployment and dispatch of the facilities, and (d) decisions on how to allocate the available facilities in a satisfactory manner.

EMS plays an important role in saving human life. Its purpose is to reduce the death and disabilities that usually result from consequences of life threatening incidents. Prominent examples of life threatening incidents include trauma and cardiac arrest. Therefore, there is a need to ensure that quality care is provided to emergency patients.
As stated by the Institute of Medicine in USA; safe, effective, patient centered, timely, efficient, and equitable care service are all considered as dimensions of quality care (Committee on the Future of Emergency Care in the United States Health System Board on Health Care Services, 2006). To achieve a high level in these dimensions, EMS service providers have to optimally design the EMS system by providing the required facilities, equipment, and personnel. This ideal system design is not easy to realize in real life because of the conflicting objectives between the quality care dimensions and the EMS design of operations (El-Sayed, 2012). For instance, to respond efficiently to all calls, an EMS provider has to have enough number of equipped ambulance vehicles which will increase the designing cost of the EMS facility. Such conflicting objectives calls for a trade-off among the objectives, thus rendering EMS operational issues to optimization approaches in seeking solutions to operational problems. The important dimensions/performance indicators of EMS for this research will be discussed in more details in later sections.

One of the good examples of EMS systems is the Spanish EMS system. The system consists of five types of vehicles as follows (Fogue, et al., 2013):

1. Non-assistance ambulance: used in non-emergency needs when the patient needs only to be evacuated. It is not equipped for providing assistance. The crew consists of one or two drivers.
2. Basic life support (BLS) assistance ambulance: it is used to provide first aid for the patients with minor injuries. It helps in reducing the risk of death. The staff consists of two and sometimes three trained drivers.
3. Advanced life support (ALS) ambulance: this kind of ambulances is capable of providing advanced life support and the practice of surgery to the patients. It is equipped to treat the patients with serious injuries. The staff consists of one physician, one nurse and a driver.
4. Fast intervention vehicles (FIV): large cars equipped with the same equipment of ALS. They are used to attend to areas with rough terrain, under adverse weather condition or in cases where special services are carried out. The staff consists of a doctor, a nurse and a paramedic.
5. Health emergency helicopter (HEH): are used in serious cases to provide emergency care for the injured people within few minutes. The staff consists of a pilot, a mechanic, a doctor, a nurse and sometimes a medical technician.

An example of facilities used in the Spanish EMS is shown in

It is worthy to mention that the multi-facility location allocation problem is more flexible than the single-type facility location problem and therefore it is expected to have better solutions and better results. Taking the EMS example, there are many objectives which an EMS provider seeks to meet. For instance, EMS provider may seek to have enough capacity to cover all demand points in different areas of the country. This objective cannot be achieved with single-type facility. On the other hand, by having transportation facilities such as ambulances located in different areas in the country, the coverage area of EMS will increase. In other words, it will increase the capacity of EMS as well as reduce the response time and increase the survival rate since the ambulances will provide the required pre-hospital care which will reduce the disabilities and death.

1.2. Facility Location Allocation Models

For the past decades, facility location allocation models have been applied to solve EMS problems. Facility location allocation problems can be classified in accordance with the objective function, constraints, solution methods, and demand patterns. In what follows, different classes of facility location allocation models will be discussed based on various solution approaches.

Solution methods for facility location allocation models can be classified into classical and non-classical or contemporary approaches. The classical approach is simply the use of mathematical programming and enumeration methods to optimally solve particular problems. As mentioned by Vanegas, et al. (2008) the generated solutions from this kind of models are a tradeoff between different objectives so as to achieve the contiguity and compactness of the requirements. These kinds of models are mostly used as an optimality reference to validate the results generated from
the non-classical approaches. Some famous examples of classical models are the linear and integer programming. It is worthy to mention that these kinds of models are not capable of solving complex, large size real problems.

On the other hand, the non-classical and contemporary approaches are used to solve complex, large size problems. They are usually used to search among extremely large spaces to generate a solution which consists of only one point or small set of points (Lindeskov, 2002; Vanegas, et al., 2008). However, there is no guarantee that the generated solution is optimal but it is usually considered to be good or near optimal. Non-classical approaches are iterative such that at each successive stage an enhanced pattern of the solution is achieved. The search continues until a solution with a desired degree is reached. Some examples of non-classical and contemporary approaches are genetic algorithm, simulated annealing, and data mining. Recently, Geographic Information Systems have been proposed as crucial additions and enhancements to contemporary approaches to the solution of location allocation problems. It is worthy to mention that nowadays, many facility location allocation models combine more than one tool in order to generate more advanced and more practical solutions.

Although there is a lot of research accomplished in the location allocation field, relatively few of these researches have addressed the multi-facility location allocation problems. This means that the exact nature of the multi-facility location allocation problems is still not well understood even more so if a different context is considered. This research provides a unique contribution to location-allocation modelling by proposing an innovative approach to modeling multi-facility location allocations in emergency medical service systems.

1.3. Problem Description

EMS design and operational issues differ from country to country. In the State of Qatar, for example, the population is continuously increasing. As a result, the demand for emergency medical services will continue to rise. This requires an effective and efficient emergency medical service system that is capable of providing the EMS service proficiently. Albeit, it is sometimes
necessary to evaluate and re-evaluate the performance of an existing EMS in order to avoid operating a sub-optimal EMS system. Evaluation and re-evaluation of the EMS system may be necessary because of; (a) rapid population growth, (b) rapid expansions in infrastructures such as new residential areas and new cities, which must be covered by the existing EMS, and (c) changes in both residential, commercial or industrial locations and/or re-locations. As such, evaluations and re-evaluations of EMS systems are crucial inputs required for strategic decisions and strategic planning for an effective EMS system.

In the state of Qatar, the EMS system incorporates more than one type of facility, namely; (i) static facilities, which are ambulance hubs and, (ii) transportation facilities, which are mainly represented as ambulance vehicles. As such, the EMS system naturally lends itself as a multi-facility location allocation problem. This represents a challenging problem because two types of facilities are considered and locations of transportation facilities depend on locations of static facilities and demand points or vice-versa (GU, et al., 2009). As such, an appropriate mixture of facilities is required to serve a given population. One way to address this problem is to determine optimal locations for both static and transportation facilities in such a way that cost are minimized. However, static and transportation facility location allocation problems are relatively new in facility location allocation sciences (GU, et al., 2009). Despite the growing interest in the field of location allocation, very few researches have addressed the multi-facility location problems i.e. addressing both static and transportation facilities. Therefore, the importance of extending location allocation models to multiple-facility location-allocation problems can never be overemphasized.

In some attempts at solving the multi-facility location allocation problems, geographical information systems (GIS) have recently evolved as contenders for addressing these problems. As such, GIS is a very promising tool which offers multiple functions that could in principle be used for solving multiple facility location allocation problems. Therefore, this research focuses on integrating GIS and location-allocation models in order to solve location-allocation problems for the EMS system in the state of Qatar.
1.4. Objectives

The goal of this research is to develop an innovative approach for modeling, simulating and evaluating multi-facility location allocation problems in emergency medical service systems. This goal is achieved through an integrated use of; a fuzzy k-means algorithm, data mining techniques and GIS simulation. These tools are used concurrently to analyze and evaluate the performance of a designed EMS system. Since a lot of data will be used in this analysis, data mining techniques will be employed to extract valuable information that can be used for strategic decision making or for planning for an effective and efficient EMS system. Data mining applications to facility location allocation problems is in itself a relatively new area of research whose strengths will be exploited in this research. The ambulance service system in the State of Qatar will be used as a case study.

In order to meet the goal of this research, the following objectives will be addressed:

1. To assess and evaluate the performance of the EMS system in the State of Qatar
2. To compare the effectiveness of the existing and a proposed EMS system through service area analysis and closest facility performance metrics
3. To develop a multi-facility location allocation model for an EMS system through integrated use of fuzzy k-means, data mining techniques and GIS methods
4. To simulate and compare the effectiveness of a variety of model-driven strategies for developing EMS systems

1.5. Scope

In this research, a mathematical model that addresses both static and transportation facilities will be constructed. The constructed model will be implemented through fuzzy k-means and the solution will be visualized, analyzed and simulated using Arch/GIS. Both tools are combined to take advantage of the strengths of location allocation modeling techniques and modern GIS techniques. The merits and demerits for using single and combined tools will be analyzed.
through the case study on the emergency medical service system in Qatar. For the case study, data was collected from; the operations of the emergency medical service system in the State of Qatar, Ministry of Development Planning and Statistics, and Ministry of Municipality and Urban Planning.

In Qatar, if there is a need for an ambulance, a person dials (999) to connect to the call center. In response to the 999 call, an ambulance dispatch system is initiated. The efficiency and effectiveness of the dispatch system mainly depends on the location and allocation of EMS units. This research focuses only on finding the best location and allocation of EMS units. Analysis of the dispatch system is outside the scope of the research. Moreover, this study does not put into consideration different types of ambulance vehicles. Thus, it does not differentiate between BLS and ALS vehicles and it does not include the dispatching policies and procedures of these vehicles. In other words, the study will focus on the optimal location of EMS units regardless of vehicle type and capability. The quoted experimental response times assume that issues of receiving 999 calls and dispatch of the EMS units consume minimum possible times without any delays.

The results of this research will contribute to a better understanding of the multi-facility location allocation problems and the merits and demerits of the various strategies, options, procedures and techniques for determining the optimal solutions. In addition, the case study will provide valuable information that can be used to improve emergency medical service systems in Qatar.

1.6. Outline of This Study

This research is organized in the following manner: In chapter 2 a review of the literature on EMS systems, location-allocation sciences, geographical information systems and their relation to each other is presented. Chapter 3 describes the methodology adopted and the various methods used to conduct this research. This includes a detailed discussion on the research process, tools used, and data gathering pre-processing and post processing of the collected data. Chapter 4 gives a report of the results and findings of this thesis including data analysis and a discussion of
results. Finally, Chapter 5 gives a summary of the findings, recommendations and a brief description of the future work related to this study.
CHAPTER 2

LITERATURE REVIEW

The main purpose of EMS is to minimize the effect of emergency incidents on human lives and health. Besides these objectives, minimization of cost, and maximization of equity, back-up coverage and service level to the uncovered zones are other additional objectives of EMS systems (Araz, et al., 2007). As stated in Disease Control Priorities Project “Emergency health incidents could typically occur through a sudden insult to the body or mind, often through injury, infection, obstetric complications, or chemical imbalance. They may also occur as the result of persistent neglect of chronic conditions” (Jamison, et al., 2006). Therefore, to prevent disabilities and to control morbidity and mortality, which result from health incidents, EMS has to provide rapid and appropriate pre-hospital care, and efficiently transport the patient to the nearest healthcare facility. Furthermore, Kobusingye, et al. (2005) showed that with efficient planning, EMS systems in developing and low income countries can implement better outcomes without increase in costs. Therefore, systematic analysis is one major area of concerns in EMS systems (Setzler, 2007).

For the EMS location allocation problems, intensive studies have been done in order to find the best location of EMS facilities especially ambulance (GU, et al., 2009; Comber, et al., 2011; Knight, et al., 2012). The purpose of these studies was to reduce the response time and/or to improve the survival rate. Comber, et al. (2011) showed that by relocating the ambulances in Niigata city-Japan, one can get better results in terms of distance and response time. They found that the average response time is reduced from 5.35 minutes to 4.12 minutes which is a great improvement since several studies shows that the survival improved when the time is less than 4 to 5 minutes.

In analyzing a vehicle location problem for EMS, Araz, et al. (2007) used different methods in order to meet multiple objectives such as; maximizing the population covered by each vehicle, maximizing the backup coverage, and increasing the service level. It was found that fuzzy goal programming (FGP) generates better results than other methods. Even though the percentage of
first covering was reduced by about 3% than the original setup, it is substituted in the backup coverage and it is found that the backup coverage was increased by more than 100%.

In an effort to locate facilities and vehicles with respect to minimizing the response time, Lightner, et al. (2006) used a MOFLEET mixed integer programming model. It was observed that the coverage area was improved and the response time was reduced. They also found that adding one vehicle to an existing EMS system may increase the coverage by 1.5% which means it will respond to 230 calls more.

The research reported in this document will focus on the location allocation problems in EMS systems. As mentioned earlier on, EMS systems are considered to be complex and large. Therefore, solution approaches typically favor the use of non-classical approaches or a combination of both classical and non-classical approaches in the search for a desired solution. Section 2.2 will address in more details different models used by other researches. In the following section the most critical EMS performance indicators will be discussed.

2.1. EMS Performance Indicators

EMS decision makers, administrators are service providers are always seeking to enhance the system’s performance. EMS performance is mainly measured by equity, response time and survival rate. Equity nowadays becomes an important performance measure in many services including EMS systems. Consequently, many studies have addressed such an important factor in their models (Beraldi, et al., 2004; Sorensen & Church, 2010; Chanta, et al., 2011). There is no standard threshold for these indicators. However, a lot of safety/health associations and departments have carried out a number of studies that focus on finding the best thresholds that will ultimately increase health outcomes. It is worthy to mention that in many studies, it was observed that there is a correlation between the survival rate and response time especially for specific medical conditions such as cardiac arrest which has high mortality risk but it accounts only for about 1% of the total emergency cases (Ludwig, 2004; Wilde, 2009; Emergency Medical Services Authority, 2013). Pell, et al. (2001) stated that the potential survival rate could increase from 6% to 8% if the target improved from 14 minutes to 8 minutes. Moreover, if 90%
of calls have been responded within 5 minutes, the survivor rate would increase by 10% to 11%. Vukamir also reached similar conclusions in his study. He found that the survival of cardiac arrest patient is increased when the response time decreases from 6.81 minutes to 5.52 minutes for BLS and from 9.49 minutes to 7.29 minutes for ALS (Vukmir, 2006).

It was found that in cardiac emergency the survival rate is significantly associated with the response time only for critical cases (Jena & Adibabu, 2009). The chart in Figure 3 represents the survival rate versus the time for cardiac arrest patients as represented by Sund (2012).

Despite most of the aforementioned researches, Wilde found that reducing the response time will increase the survival rates not only for cardiac arrest but also for other types of emergency conditions (Wilde, 2009). Based on the cost benefit analysis, Wilde argued that the anticipated benefits of reducing the response time exceed the costs. He also stated that the mortality rate increases by 8% to 17%, for every minute increases in response time. In addition to that, Blackwell concluded that the mortality is increased by 1.58% for patients whose response time
was exceeding 5 minutes and 0.51% for those whose response time is less than 5 minutes (Blackwell & Kaufman, 2002). Recall that the American Heart Association stated that the brain and permanent death occur after 4 to 6 minutes in the case of a cardiac arrest. Therefore, the suitable response time should fall within this range in order to satisfy cardiac arrest patients.

2.2. Location Allocation Models

The purpose of location allocation models is to locate an optimal number of facilities in an area of interest such that it satisfies the customer demand and minimizes the corresponding cost (Azarmand & Jami, 2009). As such, location allocation models are considered as crucial tools for strategic decision making which will help in matching facilities to demand points (Simchi-Levi, et al., 1997).

Facility location modeling takes more interests when used to locate healthcare facilities since the effect of poor location systems will go beyond cost and customer satisfaction (Daskin & Dean, 2005). Thus, it will also increase the mortality and morbidity of customers.

2.2.1. Continuous and Discrete Location Models

Continuous location models occur in a plane which tends to look to a problem from a macro level perspective. The locations are determined by coordinates, e.g. x, y, (Eiselt & Sandblom, 2010). On the other hand, the discrete location models occur almost in the nodes of a transportation network. Thus, discrete location models tend to look to the problem in a micro level. In this respect, the locations are determined by a point (node). Therefore, continuous location models come under the linear or nonlinear optimization, while discrete location models considered as integer programming problems (Eiselt & Sandblom, 2010).

Since the introduction of location allocation models, many models have been proposed (Azarmand & Jami, 2009). Such models can be divided based on their objectives, namely: covering problem, P-median problem, and P-centered problem. These models are the main models used in location problems in healthcare (Daskin & Dean, 2005). They will be discussed in more details in the following sections.
2.2.2. Covering Models

Covering models are popular and have been extensively used in research because they are applicable in many real-life problems, especially in emergency systems (Eiselt & Sandblom, 2010; Farahani, et al., 2012). The purpose behind this kind of models is to cover the demand points that are located within a predefined distance from a facility (Daskin & Dean, 2005; Fallah, et al., 2009; Eiselt & Sandblom, 2010; Farahani, et al., 2012). This predefined distance is called coverage distance or coverage radius. Consequently, a demand point is generally said to be covered by a facility if the distance between the demand point and the facility is less than or equal to the coverage distance (Daskin & Dean, 2005). As mentioned by Farahani, et al. (2012), the notion of this kind of models is related to a satisfactory method rather than a best possible one since the purpose behind them is to satisfy as much demand points as possible. For example, an increase in the time to respond to an area for fire protection will mean that the fire has greater chance to spread and it may decrease the chance to save life (Eiselt & Sandblom, 2010).

Covering models have many applications such as: designing of switching circuits, data retrieving, assembly line balancing, airline staff scheduling, locating defend networks (at war), distributing products, and warehouse locating (Fallah, et al., 2009). Application of covering models is also found in emergency facilities such as fire stations, ambulances, police cruisers, or any other facilities (Eiselt & Sandblom, 2010). There are mainly two objectives from covering models which are either to cover all demand points with minimum number of facilities or to cover as many demand points as possible with a predefined set of facilities (Eiselt & Sandblom, 2010). These kinds of models are widely used in set covering problems (SCP) and maximal covering location problems (MCLP), respectively.

2.2.2.1. Set Covering Problem (SCP)

The set covering problem is one of the first models which was developed in early1970s to locate emergency facilities (Eiselt & Sandblom, 2010). As mentioned earlier, the objective of set covering problem is to minimize the cost of facilities by locating the smallest number of required facilities to cover all demand points (Daskin & Dean, 2005; Eiselt & Sandblom, 2010; Farahani,
et al., 2012). The demand could be covered by one or more facilities if its distance does not exceed the predefined coverage distance (Eiselt & Sandblom, 2010). There are several variant forms of set covering problems, including; Location SCP, capacitated SCP, quadratics SCP, multiple optimal SCP, covering tour problem, path covering problem, probabilistic SCP, Stochastic SCP, Fuzzy SCP, multiple coverage SCP, backup coverage SCP, and multi-criteria SCP (Farahani, et al., 2012).

2.2.2. Maximal covering location problem (MCLP)

In real world, using the set covering model is associated with a number of issues. For example, covering all demand points is usually not practical due to the budget and resource constraints (Fallah, et al., 2009; Farahani, et al., 2012). Moreover, SCP considers all demand points to be the same regardless of their demand volume (Daskin & Dean, 2005; Fallah, et al., 2009). Consequently, the MCLP was developed in order to take into consideration the issues discussed above (Farahani, et al., 2012). This model tries to cover as much demand points as possible by sufficiently locating a given number of facilities. The demand points have to be within the predefined distance. There are many types of variants of MCLP such as: MCLP implicit and explicit, planar maximal covering, capacitated MCLP, MCLP with a criticality index analysis metric, MCLP with mandatory closeness constraints, probabilistic MCLP, MALP I, MALP II, Partial coverage problem, and gradual coverage.

2.2.3. P-center Models

In certain circumstances, we may be restricted to work within a certain budget and at the same time we may be restricted to locate P numbers of facilities. One solution is to relax the total coverage requirement. Another strategy to solve this problem is by using the P-center concept which focuses on minimizing the maximum distance between the demand point and the facility (Daskin & Dean, 2005; Biazaran & SeyediNezhad, 2009; Eiselt & Sandblom, 2010). P-center model is different from the covering problem in sense that in this model instead of minimizing the number of facilities within a given coverage distance, we minimize the coverage distance using a predefined set of facilities such that all demand points are covered (Biazaran & SeyediNezhad, 2009). Therefore, unlike the coverage problem, the coverage distance
requirement in P-center problem is relaxed. Thus, the objective of this model is to locate P facilities in such a way that all demand points are covered while minimizing the maximum distance between a demand point and the nearest facility. Accordingly, the focus here is to make the worst case as good as possible (Biazaran & SeyediNezhad, 2009). This concept was inspired from Rawls’s theory of justice in which the quality of a solution depends on the least well-served entity (Eiselt & Sandblom, 2010). It is worthy to mention that each demand point could have a weight which could represent the time, cost, or any other weight’s figure per unit distance (Daskin, 1995). As mentioned by Eiselt & Sandblom (2010) one of the limitations of P-center model is its exclusive focus on the farthest demand point which could lead to undesirable outcomes.

The P-center model can be applied in hospital emergency services, fire stations, warehouses, parks, hotels and location of post boxes and bus stops (Biazaran & SeyediNezhad, 2009). It is worthy to mention that the facilities in discrete models could be established either in any place in the network or could be placed only on nodes. These problems are called absolute and vertex models respectively.

### 2.2.4. P-median Models

P-median models are used to find the location of P facilities in order to minimize the total coverage distance subject to a requirement that all demand points have to be covered (Daskin & Dean, 2005). Thus, the objective is to minimize the average distance between demand points and facilities, so that the sum of costs can be minimized through this target (Jamshidi, 2009). In this way, the fixed cost for a facility disappears from the objective function. Unlike coverage problem and P-center problem, which treat service as binary (a demand point is either covered or not covered), the P-median problem focuses on minimizing the average/total weighted distance that a provider or a customer has to travel to get a service. It is suited mostly to the public sector such as: schools, hospitals, firefighting, ambulance, technical audit stations of cars, and etc. As mentioned previously, it focuses on the average distance rather than the cost of candidate locations (Chan, 2005).
2.3. Location Allocation Models Related to EMS

Three main location models used in healthcare are: set covering model, maximal covering model, and P-median model (Daskin and Dean, 2005; Shuib & Zaharudin, 2010). These models can be classified into deterministic or probabilistic models. Deterministic models ignore stochastic conditions regarding the availability of healthcare facility such as ambulances. On the other hand, probabilistic models reflect the fact that ambulances could be busy at a point of time and cannot always answer a call (Brotcorne, et al., 2003).

It is worthy to mention that covering models (including both the SCP and MCLP) are the most widespread location models in the emergency location-allocation problem (Jia, 2006; Farahani, et al., 2012). One of the earliest models used was location set covering model by Toregas in 1971, where the location of ambulances that covers all demand points with respect to a pre-specified distance was determined (Toregas, et al., 1971). Later on, the MCLP was proposed in order to maximize the demand covered with respect to one condition, which is minimizing the number of ambulance allocated. MCLP was used by Eaton, et al. (1985) in order to reorganize the EMS system in Austin, Texas. It was found that this new model saved millions of dollars in construction and operating costs.

Sadigh, et al. (2010) used contemporary edge covering model, an extension of covering model, where the partial covering of an edge through vertices is permitted. Since this problem is considered as an NP-Hard problem, tabu search was used to find a solution. The authors claimed that this type of model can be used in locating emergency facilities including EMS facilities. It was found that the proposed mathematical model could solve up to 40 vertexes with 456 edges optimally and find high quality solutions for situations with larger scale. For EMS systems, it is worthy to mention that if there are multiple types of vehicles to be dispatched to a scene, both SCP and MCLP models cannot recognize them (Brotcorne, et al., 2003). Therefore, a number of models were proposed to address issues related to the types and availability of ambulance vehicles. For EMS systems, the coverage models are considered deterministic since they do not consider the busyness of an ambulance vehicle (Brotcorne, et al., 2003; Shuib & Zaharudin, 2010). Accordingly, new models which put into consideration the availability of ambulances
have been developed by Daskin & Stern (1981) and Hogan & Revelle (1986). Fujiwara, et al. (1987) used the model developed by Daskin and Stern and concluded that reducing the number of ambulances from 21 to 15 will have similar expected covering and response time. Araz, et al. (2007) used multiple solution approaches such as lexicographic multi-objective linear programming and fuzzy goal programing (FGP) to solve backup coverage objective and service level in addition to the first covering objective in order to locate EMS vehicles. It is found that FGP model is an effective tool than others for generating a set of more realistic and flexible optimal solution.

Another model for consideration is the ambulance allocation capacity model (AACM) which is a combine of location set covering model and probability measures. The main reason behind AACM was to provide equity to all residence (Knight, et al., 2012). Convex path, convex hull, and potential sites algorithms were used to solve this problem. The results show that by adding one ambulance, the coverage increased to 90% instead of 49.3%.

All of the above studies assumed that there is fixed capacity for the facility. Yin & Mu (2012) claimed that the capacity of a facility varies according to the number of vehicles available on it. For example, the capacity of an ambulance station can be changed based on the number of ambulance vehicles available in the station. Therefore, Yin & Mu (2012) proposed a new MCLP model; namely modular capacity maximal covering location problem, which takes the capacity into consideration. Besides the modular capacity, it considers the allocation of all demand points, and the proximity of the uncovered demand points. The model input files were constructed in visual basic for application (VBA) program of ArcObjects in ArcGIS. Where then they have been solved using a branch-and-cut technique in CPLEX to find the optimal solution.

Few other researches have used P-median models. For instance, GU, et al. (2009) proposed a new heuristic method called Static Transportation Facilities Location Searching Algorithms (STFLS) to solve both static and transportation facility location problem. The purpose of their research was to minimize the average weighted traveling distance. The solution using STFLS was compared with the solution generated from a GIS software. It was found that the solution
using STFLS has better performance in average and maximum traveling distance but it is time consuming.

In EMS systems, very large amount of data and information are received on a daily bases. This data must be analyzed in order to assess the effectiveness and efficiency of a given EMS system. Moreover, EMS decisions, more specifically location-allocation decisions, involves many complex factors such as distribution of population and their characteristics, types of facilities and their capacities and resources, and road network and buildings’ distributions. This means that EMS systems are data intensive systems since different types of information and data are used to make informed decisions regarding the state of an EMS system. As such, it is important to extract valuable information for decision making. Data mining and GIS are two powerful contemporary tools that have been successfully used to handle data intensive systems. The capabilities of GIS and Data Mining techniques will be used in this research for handling large amounts of data required to assess the performance of an EMS system.

2.4. Knowledge Discovery from Data

As discussed earlier on, EMS systems of today are data intensive, i.e. data from various sources is available for use in EMS systems. Since data is from many sources, the data is complex and may not be useful for decision making and planning purposes. Therefore, it is important for decision makers to be provided with the necessary information extracted from available sources for the purpose of planning EMS systems. In this regard data mining techniques have been used to extract information from data (Witten & Frank, 2005). (Han & Pei, 2011). Figure 4, shows the necessary steps in knowledge discovery from data. Data mining can be used in four patterns, namely; characterization and discrimination, association and correlation, classification and regression, and clustering.
In EMS systems, there are no clear class labels in which instances as well as people could be grouped in. Moreover, the main focus for EMS systems is to locate each facility in such a way to serve properly a group of people who are living at a particular area. Thus, each EMS facility has to satisfy as much as possible the area that it is assigned to with the ignorance of other areas. As such it is deemed necessary in this research to investigate the impact and the extent to which clustering techniques can be used in the planning of an EMS system.

2.4.1. Clustering

Clustering can be defined as grouping set of data points into multiple groups, called clusters, where data points inside the cluster are very similar to each other and dissimilar to others that are belonging to other clusters. Unlike other data mining patterns, the grouping or classification criteria are unknown and need to be discovered by the user. The relation between points is determined using an attribute value which describes the point and it is usually represented as distance (Han & Pei, 2011). Clusters can be fully separated from each other, they can overlap or be continuously connected to each other (Abonyi & Feil, 2000). Applications using cluster analysis includes; business intelligence, image pattern recognition, Web search, biology, and
security (Han & Pei, 2011). It is worthy to mention that unlike other data mining methods, clustering is learning by observation not by example (Han & Pei, 2011). Thus, it does not need to specify class label. Therefore, it is known as unsupervised learning. Moreover, clustering can be used for outlier detection.

There are mainly four different clustering methods; partitioning methods, hierarchical methods, density-based method and grid-space methods (Han & Pei, 2011). Using the same dataset, different clustering methods leads to different clusters and results. This is because clustering methods are varied in the level of partitioning; is it hierarchical or not, level of clusters’ separation; whether they are mutually exclusive or not, the similarity measures used; is it based on distance or density or other measures, and the clustering space used; whether the entire data space used or it is sub-spaced. Table 1 summarizes the difference between the four clustering methods (Han & Pei, 2011).

Table 1: General Clustering Analysis Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Partitioning</strong></td>
<td>- Find mutually exclusive clusters of spherical shape</td>
</tr>
<tr>
<td>Methods</td>
<td>- Distance-based</td>
</tr>
<tr>
<td></td>
<td>- Effective for small- to medium-size data sets</td>
</tr>
<tr>
<td><strong>Hierarchical</strong></td>
<td>- Clustering is represented in multiple levels</td>
</tr>
<tr>
<td>Methods</td>
<td>- Cannot correct erroneous merges or splits</td>
</tr>
<tr>
<td></td>
<td>- May incorporate other techniques like micro clustering</td>
</tr>
<tr>
<td><strong>Density-based</strong></td>
<td>- Find arbitrary shaped clusters</td>
</tr>
<tr>
<td>Methods</td>
<td>- Clusters are dense regions of objects in space that are separated by low-density regions</td>
</tr>
<tr>
<td></td>
<td>- May filter out outliers.</td>
</tr>
<tr>
<td><strong>Grid-based</strong></td>
<td>- Use multi-resolution grid data structure</td>
</tr>
<tr>
<td>Methods</td>
<td>- Fast processing time</td>
</tr>
</tbody>
</table>

The distribution of EMS facilities heavily depends on the distance between the facility and incidents. Moreover, there is no specific classification for people in different zones to help in
locating EMS facilities. Therefore, the best method suitable for the current research requirements
is partitioning. The following section will discuss in details partitioning method; more
specifically k-means clustering.

2.4.2. Partitioning

Partitioning is clustering, in which a set of objects (points) are partitioned into exclusive groups
called clusters. The objective of partitioning is to maximize the intra-cluster similarities and
minimize the inter-cluster similarities (Han & Pei, 2011). It is worth mentioning that achieving
global optimum in partitioning clustering requires extremely high computational cost and
exhaustive enumeration of all the possible partitions. Therefore, heuristic methods, such as k-
means algorithms, are used to achieve an optimum by gradually improving clustering quality.
The algorithms work well in small- to medium-size problems. In this research, only k-means
clustering method will be explained in details in later sections.

k-means clustering results in assigning each object to only one cluster. This rigid cluster
assignment rule does not work with some real life applications, where an object could be related
to more than one cluster. Therefore, a more advanced k-means clustering method has been
demonstrated to deal with such situations by assigning a probability of belonging (Abonyi &
Feil, 2000; Han & Pei, 2011). This method is called “fuzzy clustering”.

Fuzzy clustering is also called soft clustering since it is flexible and allows an object to belong to
more than one cluster at the same time (Abonyi & Feil, 2000; Esnaf & Kucukdeniz, 2009). Esnaf
and Kucukdeniz (2009) used a fuzzy clustering-based method to assign customers to plants. It
consisted of two sequential phases: fuzzy clustering of multi-facility location problem and then
they used center of gravity method to find the optimal location for each facility at each cluster. It
was a pioneer paper which combined both the clustering algorithms and center of gravity
methods. It is found that this hybrid method helped in decreasing the transportation cost.

Kaundinya, et al. (2013) used a GIS-based k-medoid data mining algorithm to select suitable
locations to install biomass power plants in rural regions in Tumkur, India. This approach helped
in finding the optimal solution by minimizing the cost of power system installation, biomass
transportation and power transmission. A hybrid spatial clustering method for the selection of customer service location has been used by Fan (Fan, 2009). The proposed method considered huge number of factors which cannot be handled by conventional mathematical models. Such factors were: the spatial restrictions, constraints, traffic, terrain, cost, and environmental factors.

2.5. Geographical Information Systems (GIS)

During the past thirty years, GIS has become a popular application which is used in many diverse areas such as businesses, universities, and governments (Environmentaal systems Research Institute. Inc., 1990). The reason of using such system is that GIS is capable of analyzing both spatial and non-spatial data (Cheng, et al., 2007; Zhang, et al., 2011). It is also a powerful and efficient instrument to (Harvey, 2008): i) make location decisions specially for dynamic and complicated conditions (Vafaeinezhad, et al., 2009), ii) find the best location, iii) find the best way to get to that location, and iv) optimize the use of the available resources (Albrecht, 2007).

GIS can be defined as “an organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information to support geographical decision making” (Environmentaal systems Research Institute. Inc., 1990; Murray, 2010). Geographic information and maps are representations that follow a number of principles and conventions that help deal with the complexity of the world and guide choices that improve communication (Harvey, 2008). The skeleton of a GIS is the use of a location referencing system where the data of a specific location could be represented and analyzed with respect to other locations (Church, 2002). This powerful tool is used particularly in many research facilities and offices in order to analyze and manage resources (Harvey, 2008). It is worthy to mention that location science have been advanced by the use of developed mathematical models that combine the spatial problem and by the usage of innovative optimization techniques (Murray, 2010). Also, general purpose optimization software has been used to support advancement in model development and application. Moreover, heuristic solutions techniques have been emerged which leads to solve more complex systems in a reasonable processing time. However, all of the above methods are
not able to handle spatial data (Zhang, et al., 2011), and moreover, nowadays more real world interactions are integrated into the mathematical optimization problems which increases its complexity (Murray, 2010). Therefore, there is a need to use more advanced tools such as GIS to accurately solve such complex problems.

As stated by Murray (2010) “much of the model advancement in location science can be directly or indirectly linked with the maturation of GIS”. In his paper, Murray (2010) discussed how GIS develops the analysis and modeling of location science. He detailed that GIS is used in location allocation sciences mainly in three functions: i) model input; where GIS has supported location sciences by extracting location coordinates and its attributes to be in location models, ii) visualization; where a better understanding of the model, its objective, geographic spaces, and even the model result is reached and thus determine whether the solution makes sense or not (Murray, 2005), and (iii) GIS helps problem solution; where a location problem could be solved through the usage of GIS alone or through the combination of other modeling approaches.

As mentioned previously, due to the increasing complexity of real-life systems and accordingly the location modeling, there is a need to use advanced tools and methods which are capable of solving such systems. Therefore, the usage of GIS has been growing in recent years since it has been proved that combining GIS with other methods is very effective in decision support systems for selecting locations (Zhang, et al., 2011). For instance, Vafaeinezhad, et al. (2009) modeled task allocation to persons in a totally dynamic and complicated environment where they used GIS to simulate data, and to generate and evaluate the results of the tasks for two groups of life-detectors and rubble-removers of earthquake rescue teams. Thus, GIS used here as both to generate inputs and to find problem solution.

GIS has also been used by Rodrigues, et al. (2012) as source of input data and to present model results. Similarly, Bender, et al. (2002) and Murawski & Church (2009) used GIS to export attribute and coordinate information to be used in location modeling problems. It is worthy to mention that the allocation of task manpower in GIS was not implemented i.e. the dynamics of different elements such as tasks, environment and unpredicted events were not taken into account. Therefore, the research of Vafaeinezhad, et al. (2009) is worthwhile since it addresses
such issues. In Bozakaya, et al (2010), a Tabu search heuristic algorithm was coupled with GIS software to solve a vehicle routing problem. It was also used to store, analyze and visualize the location routing solutions. On the other hand, Murray (2005) formulated a set covering problem (SCP) using GIS to reduce inconvenient spatial effects when it is represented using only the SCP model. Based on specific criteria and assumptions, Zhang, et al. (2011) used GIS in order to minimize the set of candidate locations for pulpwood-to-biofuel conversion facilities. Similarly, Vega, et al. (2011) used GIS to improve the traditional solution methods by discarding any location which is less than the specified thresholds or which does not meet specific criteria. Cheng, et al. (2007) took the advantage of GIS to find the best location for a super shopping mall in order to minimize distance, maximize demand coverage, and to maximize average monthly income coverage. It was found that GIS easily supported all of the aforementioned enquiries.

In a competitive sitting, Vega, et al (2012) used GIS to find the best location for a retailer, where mainly two tools were used. The first tool, called Market Share Calculation (MSC), calculates the projected market share of each new facility for each point within the feasible region. The second tool, called Weighted Market Share (WMS), was used to identify the optima for the objectives and then evaluate the trade-off between all of them to find a final solution to the problem.

GIS has also been used in location problems related to health and emergency sectors. For instance, GIS was used to assess the efficiency of fire station systems (Lui, et al., 2006). Moreover, Ant-colony algorithm was applied within using the GIS to locate new fire stations in order to decrease the response time to be within 5 minutes instead of 8 minutes. Sasaki, et al. (2010) implemented GIS to count the emergency cases in each census area under study and to measure distance between the ambulance units and those census areas. The purpose of Sasaki’s study was to find the optimal location of ambulances by expecting future EMS cases and to create future EMS management strategies by expecting the demand. To improve the EMS services, GIS was implemented by E Estochen, et al. (1998) to identify existing EMS service areas and compare these areas with traffic crash densities. Moreover, the response time were estimated and compared to actual response time. Based on the results generated, improvements were suggested.
2.6. Summary of Gaps in the Literature Review

The literature review has shown that there are many models, methods, tools and techniques for designing and developing EMS systems. It has also been shown that different approaches will result in different designs of EMS systems. While most researches prefer to focus on one method, this thesis will focus on the integrated use of two methods, i.e. data mining and GIS. In particular applications of data mining are relatively a new area of study in multi-facility location allocation problems. As such, this thesis will put an effort in investigating how some data mining techniques can be utilized in the multi-facility location allocation problem in EMS systems. More specifically, the effectiveness of the fuzzy k-means algorithm will be tested and evaluated in this thesis.

Although the idea of combining a number of methods is not entirely new, this approach is deemed worthy to pursue in this thesis in order to assess the merits and demerits of combining methods in seeking better solutions. Thus both location-allocation and GIS methods will be combined with the hope of finding a better solution within an acceptable execution time. For example, Comber, et al. (2011) used a modified grouping genetic algorithm in order to optimally select best ambulance locations to minimize the population-weighted average distance. The results showed an improvement by 1.23 minutes on average in the EMS response time.

Table 2 summarizes the various location allocation models used in the public literature. Literature review has also shown that with respect to the p-canter and the p-median, the set covering problems are more widely used in EMS systems. However, Carson & Batta (1990) used the P-median in order to find the dynamic ambulance location strategy to serve emergency cases in Campus of the State University of New York at Buffalo. The ambulances were relocated in different scenarios in order to minimize the average response time to the service calls. As mentioned in Jia (2006) priority dispatching using P-median model was also used by Mandell in 1998 to optimally locate two types of ambulances ALS and BLS. GIS is a contemporary tool that is now widely used for location allocation planning. In this thesis GIS will be used in conjunction with fuzzy k-means to develop an EMS system model. GIS will also be used for analyzing the performance of various EMS models that will be created in the various
investigations of this thesis. This thesis derives inspiration from the fact that if the p-median problem is configured without an impedance cut-off, the resulting location allocation model will provide more equitability that most models (ArcGIS version 10). Since the fuzzy k-means model can be considered as an advanced p-median model, it implementation and the results thereof will contribute significantly to the importance of the p-median model in the location allocation of EMS systems.

The literature review has shown that very little research has been done to study both static and transportation facilities. However, GU, et al. (2009) used both the optimal method and GIS to solve such problems. In their study, GU et al. (2009) found that the optimal method gave better solution in comparison to GIS. However, the GIS method was faster in reaching the solution when compared to the optimal method. Based on GU, et al. (2009) and since the area of static and transportation facility still new, this thesis attempts to address both static and transportation facilities in order to improve the service level of EMS in Qatar. In other words, this study aims at improving the survival rates by reducing the response time. This objective will be accomplished by combining GIS and data mining techniques, in search for a more effective solution to EMS system in Qatar.
Table 2: Summary of Research Papers

<table>
<thead>
<tr>
<th></th>
<th>Name of the paper</th>
<th>Name of the author</th>
<th>Model</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Allocation of Emergency Ambulances to Fire Stations</td>
<td>James A. Fitzsimmons</td>
<td>p-median</td>
<td>1973</td>
</tr>
<tr>
<td>4</td>
<td>Application of an expected covering model to emergency medical service system design</td>
<td>Mark S. Daskin</td>
<td>Covering problem</td>
<td>1982</td>
</tr>
<tr>
<td>5</td>
<td>Determining Emergency Medical Service Vehicle Deployment in Austin, Texas</td>
<td>David J. Eaton, Mark S. Daskin, Dennis Simmons, Bill Bulloch and Glen Jansma</td>
<td>Covering problem</td>
<td>1985</td>
</tr>
<tr>
<td>6</td>
<td>A Hybrid Fleet Model For Emergency Medical Service System Design</td>
<td>Geoffrey Bianchi, and Richard L. Church</td>
<td>Covering problem</td>
<td>1988</td>
</tr>
<tr>
<td>7</td>
<td>Locating an ambulance on the Amherst campus of the State University of New York</td>
<td>Yolanda M Carson; Rajan Batta</td>
<td>P-median</td>
<td>1990</td>
</tr>
<tr>
<td>#</td>
<td>Title</td>
<td>Authors</td>
<td>Problem</td>
<td>Year</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>----------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>8</td>
<td>Implementing a mathematical model for locating EMS vehicles in Fayetteville, NC</td>
<td>Asad Tavakoli, Constance Lightner</td>
<td>Covering problem (FLEET)</td>
<td>2004</td>
</tr>
<tr>
<td>9</td>
<td>Designing robust emergency medical service via stochastic programming</td>
<td>P. Beraldi, M.E. Bruni, D. Conforti</td>
<td>P-median</td>
<td>2004</td>
</tr>
<tr>
<td>10</td>
<td>Developing A Mathematical Model For Locating Facilities And Vehicles To Minimize Response Time</td>
<td>Constance Lightner, Asad Tavakoli, Yahya Fathi.</td>
<td>Covering problem (FLEET)</td>
<td>2006</td>
</tr>
<tr>
<td>11</td>
<td>Extensions to emergency vehicle location models</td>
<td>Othman Ibraheem Alsallouma, Graham K. Rand,</td>
<td>Covering problem</td>
<td>2006</td>
</tr>
<tr>
<td>12</td>
<td>A fuzzy multi-objective covering-based vehicle location model for emergency services</td>
<td>Ceyhun Araz, Hasan Selim, Irem Ozkarahan</td>
<td>Covering problem (MCLP)</td>
<td>2007</td>
</tr>
<tr>
<td>13</td>
<td>Heuristic Solutions for Locating Health Resources.</td>
<td>J. Pacheco; S. Casado; J.F Alegre</td>
<td>Covering problem</td>
<td>2008</td>
</tr>
<tr>
<td>14</td>
<td>A multi-period set covering location model for dynamic redeployment of ambulances</td>
<td>Hari K. Rajagopalan, Cem Saydam, Jing Xiao</td>
<td>Covering problem</td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>Authors</td>
<td>Method</td>
<td>Year</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------</td>
<td>---------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>15</td>
<td>GIS-FLSolution: A Spatial Analysis Platform for Static and Transportation Facility Location Allocation Problem</td>
<td>Wei Gu, Xin Wang, and Liqiang Geng</td>
<td>mix of P-median and P-center</td>
<td>2009</td>
</tr>
<tr>
<td>17</td>
<td>A mixed integer linear program and tabu search approach for the complementary edge covering problem</td>
<td>Ali Naimi Sadigh, Marzieh Mozafari, Ali Husseinzadeh Kashan</td>
<td>Covering problem</td>
<td>2010</td>
</tr>
<tr>
<td>18</td>
<td>A modified grouping genetic algorithm to select ambulance site locations</td>
<td>Alexis J Comber, Satoshi Sasaki, Hiroshi Suzuki, Chris Brunsdon</td>
<td>P-median</td>
<td>2010</td>
</tr>
<tr>
<td>19</td>
<td>Integrating expected coverage and local reliability for emergency medical services location problems</td>
<td>Paul Sorensen, Richard Church</td>
<td>Covering problem</td>
<td>2010</td>
</tr>
<tr>
<td>20</td>
<td>A modified grouping genetic algorithm to select ambulance site locations</td>
<td>Alexis J. Combera, Satoshi Sasakib, Hiroshi Suzuki and Chris Brunsdona</td>
<td>p-median</td>
<td>2011</td>
</tr>
<tr>
<td>21</td>
<td>Improving emergency service in rural areas: a bi-objective covering location model for EMS systems</td>
<td>Sunarin Chanta, Maria E. Mayorga, Laura A. McLay</td>
<td>Covering problem</td>
<td>2011</td>
</tr>
<tr>
<td>No.</td>
<td>Title</td>
<td>Authors</td>
<td>Problem Type</td>
<td>Year</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>22</td>
<td>A bi-objective model for emergency services location-allocation problem with maximum</td>
<td>Mansoureh Haj Mohammad Hosseinia, Mohammad Saeed Jabal Ameli</td>
<td>Covering, p-center, p-median</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>distance constraint</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>A multi-period double coverage approach for locating the emergency medical service</td>
<td>Ayfer Basar; Bulemt Çatay; Tonguc Ünlüyurt</td>
<td>Covering problem</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>stations in Istanbul</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Ambulance allocation for maximal survival with heterogeneous outcome measures</td>
<td>Vincent Knight; Paul Harper; Leanne Smith</td>
<td>Covering problem (Set covering )</td>
<td>2012</td>
</tr>
<tr>
<td>25</td>
<td>Modular capacitated maximal covering location problems for the optimal siting of</td>
<td>Ping Yin; Lan Mu</td>
<td>Covering problem (MCLP)</td>
<td>2012</td>
</tr>
<tr>
<td></td>
<td>emergency vehicles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Joint location and dispatching decisions for Emergency Medical Services</td>
<td>Hector Toro-Diaz, Maria E. Mayorga, Sunarin Chanta, Laura A. McLay</td>
<td>Covering problem</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>approach</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 3

RESEARCH METHODOLOGY

Functions of an emergency medical service (EMS) system include; (a) to preserve life, (b) to prevent further injury, and (c) to promote recovery. While these functions can be achieved through a variety of methods, it is important to ensure optimality in providing these functions. Although operational issues such as EMS units dispatch are critical to the success of an EMS system they can be enhanced and/or enabled if the EMS system is strategically planned for optimality. Such strategic planning is usually based on sound theoretical and technical principles that are formulated based on achieving desired targets in operations.

In the process of designing and developing an optimized location-allocation plan for an EMS system this research employs a mix of qualitative and quantitative methods. The qualitative method was used to investigate how EMS systems (e.g. the EMS in the State of Qatar) are designed and why EMS systems are operated in a particular manner. This method was envisaged to help in understating; (a) the dynamics of EMS systems, and (b) the various generic problems in EMS systems. Such an inquiry-based analysis can be used to derive relevant information required for developing innovative approaches and unique analytical methods that can be used to develop new, improved and futuristic EMS systems. The quantitative method was then used to seek empirical support for the proposed approach to modeling EMS systems. In this thesis, the EMS system in Qatar was used as a case study.

Data collection for the quantitative approach was carried out by; (a) studying and analyzing available data sets, and (b) conducting structured and semi-structured interviews and meetings with a number of stakeholders affiliated to the EMS system in the State of Qatar. These interviews were conducted through phone calls and/or planned meetings. In cases of an in-depth understanding of complex ideas, techniques and
procedures, face-to-face meetings with key stakeholders were preferred. Interviews were conducted with key personnel from the following organizations; (a) Ambulance Services in Qatar, (b) Qatar 2022 Supreme Committee, (c) Ministry of Development Planning and Statistics, and (d) GIS Center under the Ministry of Municipality and Urban Planning. Collected information and data were used in the empirical analysis of EMS systems. Figure 5 summarizes the key stakeholders consulted in this research. A number of tools were used to analyze and process data into useful information from which findings were determined and conclusions made.

![Figure 5: Main Stakeholders of the EMS System in State of Qatar](image)

3.1. **Overview of the Research Process**

Like any other engineering research problem, the approach taken in this thesis was based on the engineering design method. Available solution methods for similar and related work were extracted from the public literature, studied and their applicability investigated. New solutions and ideas were proposed and implemented and the effectiveness of these solutions was explored. Results were then analyzed and
accordingly conclusions were drawn. The overall research process is presented in the flowchart in Figure 6.

![Figure 6: Research Process](image)

### 3.2. Tools Used

Since the EMS system in Qatar was used as a case study, the core data required for this research is related to the population demographics, incidents, emergency medical services locations, as well as the road networks in the state of Qatar. Relevant mathematical models, tools, and software were identified. Two contemporary methods (i.e. Data mining and GIS location science) were used to optimally locate and relocate EMS
facilities. Data mining and GIS location sciences were used concurrently to implement optimal location/re-location of EMS units. Closest facility analysis, service area analysis, and normality tests were used to statistically analyze the results of this study. Recommendations were suggested based on the analysis of the obtained results.

A number of tools were used in this research. The main tools used were; ArcGIS software, Matlab software, Excel and Minitab software. Minitab and excel softwares were used for statistical analysis of the various data sets used in the investigation.

Matlab is a software that is capable of analyzing data, developing algorithms, plot functions and data, as well as create models and applications (The MathWorks, Inc., 2013). In this research, Matlab was used to develop and implement a fuzzy clustering algorithm. In this respect, Matlab was used to cluster the large amount of population/incidents demographics data collected from databases provided by stakeholders. Obtained data was then processed into useful information which was used in the design and development of a more effective EMS system.

Arc/GIS is a comprehensive system that allows users to collect, organize, manage, analyze, communicate, and distribute geographic information (Economic and Social Research Institute, 1995). Arc/GIS was used to; (a) visualize the existing EMS system, (b) visualize proposed EMS models developed through a variety of procedures and techniques, (c) conduct a network analysis of the existing and proposed EMS systems, and (d) simulate the performance of proposed EMS models.

3.3. Case Study

3.3.1. Study Area

The study area is the State of Qatar, a peninsula in the gulf region. The state of Qatar lies halfway along the west coast of the Arabian Gulf. Its total area is 11,606.8 km² with a width of 85km and length of 185km (Ministry of Development Planning and Statistics, 2010). The population of State of Qatar is estimated to be 1,551,821 according to the
2010 census. More than 90% of them live in Doha and its suburbs. Local planners of emergency medical services (EMS) in Qatar are often faced with the need to assess and re-assess the performance of the current EMS system in order to avoid sub-optimal operations. This is because the performance of the EMS system is affected by unexpected population growth, rapid expansion of cities and industrial areas as well as random relocations of residential, commercial and industrial facilities. Figure 7 shows the spatial positions of the population and road network in the State of Qatar, while Figure 8 shows the distribution of EMS incidents and the road network.

Figure 7: Location and Distribution of Population and Road Network in Qatar
Figure 8: Distribution of emergency incidents in Qatar for the year 2013/2014
3.3.2. Case Study EMS

The case study EMS was designed in GIS and based on the gravity-based two-step floating catchment area method (Luo & Qi, 2009; McGrail & Humphreys, 2009; Ngui & Apparicio, 2011). This EMS operates on two modes, namely the winter model and the summer model. The EMS units are divided into hubs (ambulance stations) and spokes (ambulance vehicles). Figure 9 and Figure 10 show the positions of the EMS units and the hub catchment areas for the summer and winter models respectively for 2013. The development of the current location-allocation model for the EMS system in Qatar was based on historical data regarding the incident calls.

Figure 9: Summer Response Locations – 2013
In the winter EMS system, the ambulance units are spread more around the hubs than the summer. Thus, the system in winter is regarded (by ambulance service providers) to be more ideal than that in summer. The reason behind having two different modes is that most of the people travel to other countries for the summer holiday since most companies have either a shutdown or temporary closure during this period. Consequently, the number of incidents is expected to be less. There are seven hubs in summer while there are 6 in winter. These hubs cover the whole country as shown in Figure 9 and Figure 10. The colored polygons (hub catchment areas) in Figure 9 and Figure 10 represent the coverage of each hub while the circles represent the spokes locations.

Figure 10: Winter Response Locations – 2013
The EMS system in Qatar consist of ambulance hubs and ambulance spokes (i.e. vehicles), all of which will be referred to as EMS units. For the existing system the total number of hubs is seven and the total number of spokes is 49. It is important to mention that hubs are more advanced locations than spokes in sense that they have more facilities for staffs such as offices for supervisors, and spare equipment and ambulances. Three main ambulance vehicles are; Alpha, Delta, and Charlie as shown in Figure 11.

![Figure 11: EMS Facilities and Staffs in State of Qatar](image)

The Alpha unit has the highest number and is used almost in all emergency incidents. It is used to provide prehospital treatment and care for any emergency case and it is usually operated by two ambulance paramedics. The Delta unit is a non-clinical supervisory unit used to facilitate the process of EMS by managing resources and communicating with different agencies involved in various emergency cases. It consists of only one operational supervisor or an ambulance paramedic. The Charlie unit is used in critical
cases and when the Alpha units need advanced assistance. There are few Charlie units and these units consist of critical care paramedic and ambulance paramedic. Recently, the life flight unit has been used for incidents outside Doha city and for any critical case that need fast response. It consists of one critical care paramedic and one ambulance paramedic. Table 3 summarizes the resources available for each EMS type.

Table 3: Resources Summary of EMS System at State of Qatar

<table>
<thead>
<tr>
<th>Unit Type</th>
<th>Maximum Number Available</th>
<th>Man power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>62</td>
<td>✓</td>
</tr>
<tr>
<td>Delta</td>
<td>10</td>
<td>✓</td>
</tr>
<tr>
<td>Charlie</td>
<td>9</td>
<td>✓</td>
</tr>
<tr>
<td>Life flight</td>
<td>1</td>
<td>✓</td>
</tr>
</tbody>
</table>

The current EMS prioritizes the response into three main categories; (a) “life threat” scenarios, which are characterized by light and sirens driving and has the highest priority; (b) “urgent” scenarios which are characterized by normal driving and has priority of 2, and; (c) “Transport” which means the response for scheduled patient or inter-facility transfer and has the lowest priority.

Besides the response classification, there is also the patient classification. They classify the patient into: (i) Trauma: if injury caused by external force, (ii) Medical: which refers to any kind of illnesses such as diabetic collapse, and (iii) Transport: which refers to scheduled patients or inter-facility transfer. Most emergency calls fall under life threatening and urgent cases.

One of the mostly used performance measures is the response time, i.e., the time spent from dispatching and EMS unit until it reaches the scene of the incident. The targets in the current EMS system in Qatar are as follows: (a) in urban areas the target is to serve
75% of the cases within 10 minutes, and 95% within 15 minutes, (b) in rural areas the target is to serve 75% within 15 minutes and 95% within 20 minutes. It is obvious from the aforementioned targets that there is a difference in response times between rural and urban area. Thus, there is preference for the urban areas over the rural areas i.e. non-equitability). However, human life is important regardless of location. Consequently, this research investigation will consider equitability as an important measure of performance of the EMS system in Qatar regardless of the location in Qatar.

3.3.3. Data Collection

This section provides a discussion about the data available for this research and how it was used. Collected data included; population data, the number and type of incidents, as well as details of the road-network in the State of Qatar. Details of collected data are provided in appendices A and B.

Population data included all population figures in each district in Qatar. The available data was divided into; (a) male and female, and (b) groups based on age such as; 0 to 4, 5 to 9, 10 to 14, 15 to 19, 20 to 59, and 60+.

Call incidents data was provided by the ambulance services. This data was used to identify patterns about the emergency cases with respect to the time over the day and the month of year. Available incidents data was for a period of one year starting from October, 2012 till September, 2013. This data included many details such as the date and time of the incident, the zone number, location name, case classification, response priority, call time, dispatch time, response time, and time to reach to hospital, and even the destination or in other words the name of the hospital. Road network data is important in this research since it helps in calculating the distance and, accordingly, the time required to reach to an incident. The available road network data did not take into account the speed of the street, the traffic lights and any other specifications of the street that might affect the performance of the EMS system.
3.4. Methods Used to Achieve the Objectives

Four principal methods were used to achieve the objectives of this thesis. These methods are: (i) statistical data analysis, (ii) GIS network simulation, (iii) integrated use of data mining techniques and GIS methods, and (iv) hybrid clustering and GIS simulation. Details of these methods are provided in the following sub-sections.

3.4.1 Data Analysis

One of the objectives of thesis is to assess and evaluate the performance of an EMS system. In order to assess and evaluate the performance of the EMS system in Qatar, collected data was statistically analyzed. Statistical analysis was carried out in Minitab software, while descriptive statistics were analyzed through Microsoft Excel. Data analysis aimed at classifying incidents, determining response priorities and response times for the existing EMS case study. Microsoft Excel was used to analyze patterns in the historical incidents EMS calls. Such patterns included; variation and scatter of incidence during the course of a year and incidents zone analysis.

3.4.1.1 GIS Data Mapping

GIS mapping was used to develop district population density maps and zone incidents density maps. The various GIS layers, objects and attributes used in this research are shown in Figure 12.

3.4.1.2 Statistical Analysis

Statistical analysis was used to study and explore the pattern of large amount of data generated and collected in this study. In this research, it is also used to compare among different models and distributions to study whether there is a statistical significance or not. Statistical analysis is used because simple measures such as average or median may
not be enough for decision making and hence may not be suitable for drawing meaningful conclusions (Zomaya, 2005).

Figure 12: GIS Layer Required/Generated for/from the Research

Usually t-tests or ANOVA are used to test the statistical significance of data points if the distribution of these points is normal. There are several statistical methods used depending on the type and number of distributions available. In order to use these methods, first we need to test the normality of the data points and then depending on the results, a suitable method is used to test the significance of data provided. Figure 13 shows a flow chart that was used to select suitable methods for statistical analysis.
3.4.2. GIS Network Simulation

Another objective of this thesis is to compare the effectiveness of the existing (case study EMS) and a proposed EMS system through service area analysis and closest facility performance metrics. In order to achieve this objective, the Network Analyst extension tools in ArcGIS were used in simulation of the operations of the EMS systems.

3.4.2.1. Service Area Analysis

A network service area is a region that includes all accessible streets. Therefore service area analysis helps to evaluate accessibility of EMS units. In ArcGIS, concentric service areas show how accessibility varies with impedance. Service areas can be used to identify how many people can be served within the neighborhood or region. The steps for service area analysis can be summarized as follows:

1. Click on Network Analyst > New Service Area. A service area analysis layer will be created which shows six networks analysis classes—Facilities, Lines, Polygons, Point Barriers, Line Barriers, and Polygon Barriers.
2. Add all facilities that you need to find their service area under the facilities class. Specify the field that represents their names.
3. In the Layer Properties dialog box, modify the analysis parameters as following:
   a. In the analysis setting tab, there is what is called Impedance: it reflects the cost attribute. It could be time, length, or anything which could be considered as cost.
b. Add resection if there is any.
c. Modify “the default breaks” which is the value of the cost.
d. In the polygon generation tab, choose the appropriate polygon setting. We commonly use “Generalized” as a polygon type and overlapping under the multiple facilities options.

4. After that, click solve in the network analyst toolbar.

3.4.2.2. Closest Facility Location

The closest facility solver measures the cost of traveling between incidents and EMS facilities and determines which are nearest to one other. The steps for the closest facility analysis can be summarized as follows:

1. Click on Network Analyst > New Closest Facility. A closest facility analysis layer will be created which shows six network analysis classes: Facilities, Incidents, Routes, Point Barriers, Line Barriers, and Polygon Barriers.
2. Add all data points which are considered as facilities. Specify the filed that represents their names.
3. Add all data points which are considered as incidents. Specify the filed that represents their names.
4. In the Layer Properties dialog box, modify the analysis parameters as following:
   a. In the analysis setting tab, there is what is called Impedance: it reflects the cost attribute. It could be time, length, or anything which could be considered as cost. Choose the appropriate attribute which you seek to minimize.
   b. Specify (if you need) the default cutoff value, in which GIS will stop searching for a facility whenever that value is exceeded.
   c. Specify the number of facilities that need to be found for each incident under the “Facilities to find” parameter.
   d. Specify the travel form and add any restriction if there is any.
5. After that, click solve in the network analyst toolbar.
3.4.3. Integrated use of Data Mining and GIS Methods

Another objective of this thesis is to develop a multi-facility location allocation model for an EMS system through integrated use of fuzzy k-means, data mining techniques and GIS methods. As discussed in the literature review chapter of this thesis, an EMS system can be modeled as a location-allocation problem. Models for modelling EMS systems are usually classified as classical methods, non-classical methods, and contemporary methods. This research employs both non-classical (fuzzy k-means clustering) and contemporary methods (GIS) in the modelling of EMS systems. In this integrated methodology, solution procedures for general location science problems are used in conjunction with GIS modeling and simulation. This solution procedure is envisaged to enable the determination of optimal facility locations as well as the allocation of demand to those facilities. The mathematical basis of such procedures is built on the application of various location models.

As discussed in section 2.3, relatively very little research has been done to implement p-median models in EMS systems. Therefore, the impact of p-median models and their effectiveness and efficiency in modelling EMS systems is less known in comparison to other models. This research contributes to knowledge by implementing a modified p-median method in the form of a fuzzy clustering algorithm. In addition, this research further studies the usefulness of this model in EMS systems by integrating the fuzzy k-means algorithm (developed in Matlab software) with GIS simulation (developed in Arc/GIS software). For the purposes of this research, a comparison will be made between a GIS based location allocation model and the fuzzy k-means-based model. Various components of this integrated methodology are discussed in the following sub-sections.

3.4.2.1. Data Clustering

Large volumes of data sets were used in the quantitative analysis of the EMS system in Qatar. In addition, large volumes of data were also used in the design and development of
the proposed EMS system in Qatar. In order to organize the collected data and extract relevant information for strategic planning and decision making data mining methods and techniques were employed. Data mining, a non-classical method, was employed to; (a) classify data, and (b) extract useful information from the generated data. Thus, data mining techniques and heuristic based algorithms were used to formulate the EMS system as a location allocation problem. Data clustering was implemented through the fuzzy clustering method.

3.4.2.2 Fuzzy Clustering Method

Based on the discussions in the literature review chapter, a fuzzy clustering model was used to represent an EMS system. The logic behind the choice of the fuzzy k-mean approach is that since an incident can be served by more than one facility depending on; its nearness, the road network, EMS facility availability, and any other circumstances, the fuzzy k-means gives a better representation of the EMS problem. Thus, given a set of objects $X = \{x_1, x_2, ..., x_n\}$, a fuzzy set $C_j$ is a sub-set of $X$ that assign for each object at $X$ a membership degree that ranges from 0 to 1. Fuzzy clustering of $k$ fuzzy clusters can be represented by a partition matrix $U = [u_{ij}]$, where $u_{ij}$ is the membership degree of object $x_i$ in fuzzy cluster $C_j$, $1 \leq i \leq n$ and, $1 \leq j \leq k$. The partition matrix should satisfy the following requirements (Abonyi & Feil, 2000; Han & Pei, 2011):

(i) $0 \leq u_{ij} \leq 1$
(ii) For each object $x_i$, $\sum_{j=1}^{k} u_{ij} = 1$, this is to ensure that every object participates in the clustering equivalently.
(iii) For each cluster $C_j$, $0 < \sum_{i=1}^{n} u_{ij} < n$, this to ensure that for every cluster, there is at least one object for which the membership value is nonzero.

The logic in the concept of partitioning in fuzzy clustering can be explained as follows: consider an EMS example in which the letters A, B, C, and D in Table 4 represent EMS facilities and time column represents the time required by each EMS facility to reach a
customer. The following formula \((Z)\) will be used to represent the degree of nearness of EMS facility to the customer.

\[
Z = \begin{cases} 
1, & \text{if time} \leq 10 \text{ min} \\
\frac{10}{t}, & \text{if time} > 10 \text{ min}
\end{cases}
\]  

(eq.1)

Based on the relationship in equation 1 above, the degree of nearness for facilities A and B (Table 4) is 1, whereas the degree of nearness for facility C = 0.5 while that for facility D = 0.25. Therefore, facilities A and B are better than the rest to serve the patient.

<table>
<thead>
<tr>
<th>EMS Facility</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
</tr>
<tr>
<td>D</td>
<td>40</td>
</tr>
</tbody>
</table>

Thus, based on specific criteria (which in this case is the distance), the fuzzy clustering algorithm was used to group residents and incidents in the state of Qatar into clusters. The middle of the cluster will represent the location of the EMS hubs or spokes as named by Qatar ambulance services.

### 3.4.2.3. Fuzzy k-Means Algorithms

The objective of the implemented fuzzy clustering algorithm is to minimize the following function:
\[ J_m = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^m \|x_i - c_j\|^2, \quad \text{(eq.2)} \]

where

\[ 1 \leq m < \infty \]

is a weighting exponent which determines the fuzziness of the resulting clusters.

\[ U = [u_{ij}] \]

is a fuzzy partition matrix of X,

\[ c_j = [c_1, c_2, \ldots, c_k], c_j \in \mathbb{R}^n \]

is a matrix of cluster centers, which have to be determined.

The logic behind implementing Fuzzy k-means clustering is that it can be configured to locate EMS facilities by finding a meaningful grouping for a given data objects (i.e. population, residential houses and incidents in this research) based on the distance between these objects. Fuzzy k-means model is a variation from p-median model with squared Euclidean distances (Eislet, 2011). It is based on the minimization of the objective function shown in Eq. 2 iteratively by recognizing and updating \( u_{ij} \) and \( c_j \), such that:

\[ u_{ij} = \frac{1}{\sum_{j=1}^{k} \left( \frac{\|x_i - c_j\|}{\|x_i - c_j\|^m} \right)^{m-1}}, \quad c_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m} \]
The iteration will stop when the termination criterion $\varepsilon$ is met. Thus, the iterations will be terminated when:

$$\|U^{(y+1)} - U^{(y)}\| < \varepsilon$$

The notations and parameters used above can be defined as follows:

- $i$: set of houses/incidents, $i \in \{1,2,3,\ldots,n\}$
- $j$: set of EMS clusters, $j \in \{1,2,3,\ldots,k = 49\}$
- $y$: iteration steps, $y \in \{1,2,3,\ldots\}$
- $m$: A weighting exponent which determines the fuzziness of the resulting clusters, $1 \leq m < \infty$
- $x_i$: The house/incidents locations
- $c_j$: The centroid of the $j^{th}$ EMS cluster
- $\varepsilon$: termination criterion, $\varepsilon = 1e-5$
- $U^y$: The fuzzy partition matrix
- $u_{ij}$: The degree of membership of house $x_i$ in the EMS cluster $j$.

The algorithm is composed of steps shown in Figure 14.
3.4.2.4. Constraints and Assumptions

There are many constraints and assumptions that have been set while implementing the fuzzy k-means clustering algorithms in conjunction with GIS simulation. These constraints are:

1. The targeted response time to reach to any incident or house is 10 minutes considering an average speed of 80 km/hour. The response time of 10 minutes was
chosen since it closely represents an ideal EMS system as discussed in the literature reviewed.

2. The number of EMS units required to be located is 49 units. Thus, the algorithm has been coded to generate 49 clusters and their centroids. The logic behind this constraint is that the investigation seeks to find out whether effective and efficient planning of an EMS system can provide better results without increasing the number of currently available EMS units.

Assumptions include the following:

1. Each house consist of 14 family members, typical of the culture in Qatar and the gulf region
2. The speed is 80km/hour. Thus, the distance between EMS unit and incidents is around 13.333 Km.
3. For clustering houses, it is assumed that each 1% of houses are accumulated near to each other and thus they are form one cluster. Therefore, the total number of clusters in and outside Doha is 100.
4. For clustering incidents, it is assumed that there is a total of 50 clusters in and outside Doha.

3.4.2.5. Verification and Validation of Clustering Techniques

Verification simply means building the system right. Thus, the built system should meet the requirements and specifications that have been specified in the design stage. In this research, verification is assured by using the fuzzy clustering model with the help of GIS software. Hence, it is assured that the response time needs to be within 10 minutes and the total number of EMS units is not exceeding the 49 units. In addition to that, Arc/GIS is used to verify that populations are allocated at the right district and the area of districts
is representing the real districts given by the Ministry of Development Planning and Statistics. Moreover, it is used to verify that the proposed EMS locations are in free, not occupied area.

While verification means matching the built system with the specification, validation means meeting the user’s expectations and specifications. It means in other word, building the right system which represents the real situation and meets the operational requirements. The difference between validation and verification is presented in Figure 15. The built system in this research will be validated by three different ways, namely; using Arc/GIS, using another clustering algorithm called k-means, and using silhouette method. Each method is used to validate specific thing as described in the below graphs.

![Figure 15: Verification and Validation](image)

Starting with the first method which is implemented using Arc/GIS, it is used to investigate and validate whether the incidents/population are served within the specified response target or not. A random number of locations incidents were generated and the network analyses were used to examine whether the EMS facility can reach to them within the specified target (10 minutes) or not. Moreover, the same concept where used also to assure that all populations' houses are covered within the acceptable response time. Results of this method are shown in section 4.2.3. of chapter 4.
Moving to the next method, it is used to validate the code itself and have some sense in whether the generated EMS locations are acceptable or not. This done by using one of well-known clustering algorithms called k-means. The difference between this algorithm and fuzzy k-means clustering is simply that this algorithm allows each point (incident/population) to belong only to one cluster, whereas each point in the fuzzy k-means has a weight associated with a particular cluster. Validation done by Matlab by finding the best ten clusters for the distribution of houses in 4 different zones in Qatar (Madinat Khalifa, Dahl al-Hamam, Almarkhyia, Freej Kulaib). The validation code is shown in Appendix H. The location of houses at these zones was given by the Ministry of Municipality and Urban Planning in Qatar. After implementing both algorithms in Matlab, the centroid of the ten clusters using both algorithms were found as shown in

Figure 16.

A comparison between the nearest centroids for each algorithm executed as following:

(i) Find a random point. Here I took the center of one of the zones which is Madinat Khalifa-south.
(ii) Find the distance between the center of that point and each centroid for each algorithm as shown in Table 5.
(iii) Considering the k-means centroid as the base, the percent difference was found by using the following equation.

\[
D\% = \left| \frac{\text{Dis}(c) - \text{Dis}(k)}{\text{Dis}(k)} \right| \times 100\% \tag{eq.3}
\]

where,
D\% = percent of distance difference
Dis(c): the distance between the center of the zone and the centroid of cluster i generated by fuzzy k-means clustering
Dis(k): the distance between the center of the zone and the centroid of cluster i generated by k-means clustering
Figure 16: Validation of Fuzzy k-means Clustering

It is found that the average difference between the two algorithms is ±13.1% as shown in Table 5.

The silhouette method was used to validate the number of clusters, which is here EMS units, in and out Doha. Thus, here we will examine if the number of existing EMS units are enough for serving the population. In the currently existing EMS system, there are in total 49 EMS units distributed in the State of Qatar. 13 EMS units are out of Doha and 36 units are in Doha.

Table 5: Distance and Percent Difference for Each Clustering Algorithm

<table>
<thead>
<tr>
<th>Distance between the center of the zone and the centroid of the cluster</th>
<th>%Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-means</td>
<td>Fuzzy k-means</td>
</tr>
<tr>
<td>723.051</td>
<td>612.408</td>
</tr>
</tbody>
</table>
Silhouette method finds how well each point lies within its cluster by going through the following steps (scikit-learn developers, 2011):

1- Find the average “a” distance between the point x and other points within the cluster.
2- Find the average “b” distance between the point x and all points in the nearest cluster.
3- Find the silhouette coefficient for x “S(x)”, which is the difference between step 2 and step 1 divided by the maximum of step 2 and step 1. Thus, the equation of the silhouette coefficient is as below in equation 4:

$$S(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}}$$  
(eq. 4)

The value of the coefficient will be within ±1. If the coefficient is near to 1, it means that the data are properly clustered, whereas it is badly clustered if it is near to -1 (The
MathWorks, Inc., 2014). Any values near to zero it means that clusters are overlapping. The best number of clusters for the dataset will be determined by looking at the beak value (maximum coefficient) in the graph.

For in Doha datasets, we found as seen in Figure 17 that almost all clustering sizes have almost similar coefficient values. Moreover, all of them have a value close to zero, which means that the clusters are overlapping and points are close to each other. We can say that clustering size of 33 has the best coefficient value. In addition to that, the silhouette coefficient values for the clustering groups of out Doha data points are also found to be near to zero except for clustering size of 12 clusters (Figure 18). The best coefficient value is found to be at clustering size of 13. Overall, we can say that all population data are near together, therefore, and accordingly clusters are overlapping. Based on this method, it can be inferred that the total number of EMS units should be around 46 (i.e. 33+13). Since the number of currently existing EMS units in the EMS system of Qatar is 49, the number of EMS units in and out of Doha can be considered acceptable.

Figure 17: Validating Number of Clusters - In Doha
3.4.2.6 GIS Method

ArcGIS 10 was used for location allocation analysis of the EMS systems. ArcGIS location-allocation analysis uses six problem types, namely; minimize impedance (p-median), maximize coverage, minimize facilities, maximize attendance, maximize market share and target market share. For the application in this thesis only two methods were found to be more suitable. These two methods were the minimize impedance and the maximum coverage models.

Minimize Impedance

In this model, facilities are located such that the sum of all weighted costs between demand points and solution facilities is minimized. In the description of the study area, it has already been observed that the EMS system in Qatar differentiates between
operations in summer and in winter. In addition, the EMS also differentiates between services in Doha and services out of Doha. Hence, there is no equitability.

Maximize Coverage

In this model, facilities are located such that as many demand points as possible are allocated to solution facilities within the impedance cutoff. For EMS location this model is important because emergency services are required to arrive at all demand points within a specified response time. The general procedure for location allocation in the GIS environment can be summarized as follows:

1. Click on Network Analyst > New Location-Allocation. A location-Allocation analysis layer will be created which shows six network analysis classes—Facilities, Demand Points, Lines, Point Barriers, Line Barriers, and Polygon Barriers.
2. Add all data points which are considered as candidate facilities. Specify the filed that represents their names and their type, weight and capacity if there is any.
3. Add all data points which are considered as incidents. Specify the filed that represents their names and weights if there is any.
4. In the Layer Properties dialog box, modify the analysis parameters as following:
   a. In the analysis setting tab, there is what is called Impedance: it reflects the cost attribute. It could be time, length, or anything which could be considered as cost. Choose the appropriate attribute which you are interested in.
   b. Specify the travel form and add any restriction if there is any.
   c. In the advanced setting tab, specify the type of the problem whether it is minimize impedance, maximize coverage, or any other problem from the list.
   d. Specify the number of facilities that you seek to choose from the candidate facilities.
   e. If required, specify the impedance cutoff value, in which GIS will stop searching for a facility whenever that value is triggered.
   f. Specify the type of transformation in the “impedance transformation”.
5. After that, click solve in the network analyst toolbar.
3.4.3. Hybrid Clustering and GIS Simulation

Another objective of this thesis is to simulate and compare the effectiveness of a variety of model-driven strategies for developing EMS systems. For EMS systems, the response time is a very important factor that determines the quality of the EMS. A number of factors affect the performance of EMS system. At the planning stage, the strategy and/or procedures used may affect the performance of the resultant EMS systems. A Hybrid clustering and GIS simulation framework was implemented to investigate the effects of a selected strategy or procedures for planning an EMS system. Figure 19 shows the process flow chart for implementing the hybrid clustering and simulation framework.

The objective of the GIS simulation was to investigate the performance of a number of model-driven strategies and procedures for planning EMS locations in order to achieve an acceptable EMS response time. Arc/GIS was used in this research to imitate the real system and to understand and examine how different EMS systems will respond to a given set of demand points (i.e. incidents in this work). This was achieved through the Network analyst in GIS. Using the Network Analyst, one is able to solve realistic network conditions and sophisticated models to find one of the following (Esri, 2014):

- The quickest way to get from point A to point B
- Houses/incidents which are within five minutes of a fire station
- Market areas covered by a business
- Ambulances or patrol cars can respond quickest to an incident
- Find which store should be closed if a company has to downsize and at the same time maintain the most overall demand

The network analyst was used to carry out two types of analysis; first to find the service area and second to find the closest facility. Finding a service area means for each EMS unit we find the area that is surrounding it for response times of 8, 10, and 15 minutes. It includes also all streets within that range. Although this investigation was focused on a
target of 10 minutes response time, response times of 8 and 15 minutes were included for comparison sake. On the other hand, the closest facility function determines the nearest facility (EMS units) and measures the cost of travelling between incidents and EMS units. In this regard, cost means the time taken to reach to the incidents.

In the GIS simulation, a total of nine models were developed, simulated and their performance evaluated based on the service area analysis and closest facility analysis. The aim of the simulation was to guide EMS planners in terms of; (a) the different types of simulation models that can be done in the GIS environment, (b) the performance of the various models that can be created in a GIS environment, and (c) the effect of planning strategies and/or planning procedures on the performance of and EMS system.
Figure 19: Flowchart for Implementing the Hybrid Clustering GIS Simulation
CHAPTER 4

RESULTS AND DISCUSSIONS

In order to meet the goal of this research, the following objectives were addressed: (1) to assess and evaluate the performance of the EMS system in the State of Qatar, (2) to compare the effectiveness of the existing and a proposed EMS system through service area analysis and closest facility performance metrics, (3) to develop a multi-facility location allocation model for an EMS system through integrated use of fuzzy k-means, data mining techniques and GIS simulation, and (4) to compare the effectiveness of model-driven strategies for developing EMS systems. The findings of these investigations are discussed in the following sections.

4.1. Data Analysis

In order to assess and evaluate the performance of the EMS system in the State of Qatar, data was collected from ambulance service in state of Qatar. The data was from October, 2012 to September, 2013. These data includes the following:

- Number of Alpha, Delta, and Charlie units available on hourly bases.
- Location of units in summer and winter.
- Incidents received, response time, and time to arrive to the hospital.

The hourly available EMS units in the current EMS system are provided in appendix E. Location of EMS units in winter and summer are different in the current system. The details of the location of EMS units in summer and winter are found in section 3.3.2. in chapter 3. The incidents served and response time data is provided in appendix B.
4.1.1. Response Times

An examination of the provided data on response times revealed that there were many outlier data points, and therefore, there was a need to specify the acceptable range for the response time. A box plot of the response times is shown in Figure 20.

![Box Plot (Minutes)](image)

Figure 20: Box Plot for Response Time.

It was decided that any response time falling beyond 47 minutes was considered unacceptable, and such data were treated as outliers. The reasons for such anomalies were attributed to incorrect recording of data by the ambulance service personnel on the raw data sheets provided (data was collected and recorded manually). Based on statistical analysis, the average response time for the current EMS system in Qatar was found to be 12.49 minutes and the median was 12.03 minutes. It is worth mentioning that most of the data recorded in the provided data sheets were between the EMS target ranges of 10 to 20 minutes. It is also important to highlight that on average a used EMS unit would be available for use again after 48.8 minutes. This represents the ‘recovery time’ of that EMS unit. Based on the calculated response times, it can be inferred that there is a “good enough” facility location plan of the currently existing facilities and resources.
4.1.2. Pattern of Incidents

From Figure 21, it is found that the number of incidents is somehow stable over the year and it ranged between 240 to 470 per day. However, the number of incidents increases slightly in the summer period (June, July, August, and September). Further observations of the number of incidents occurring every day shows that there is no correlation between the number of incidents and the day. Thus, incidents do not depend on the day of the week.

Figure 21: Incidents Over a Year

On the other hand, it was observed that there is correlation between the number of incidents and the time of the day. As shown in Figure 22, the number of incidents increases radically from 5:00AM to reach a peak by 11:00AM. Afterwards, the number of incidents decreases slightly from 11:00AM to around 4:00PM after which it increases again to reach another peak at 8:00PM and the cycle is repeated. It is worthy to mention that the number of incidents is minimum on Friday and Saturday morning in comparison to other days. However, it is maximum in the evening and night times. Moreover, the pattern of incidents on Fridays is different from the other days since it keeps increasing gradually while the pattern for other days have a specific shape with two peaks (one in the morning time and the other in evening time).
4.1.3. EMS Availability

According to the above discussion, it is expected to have a relation between the number of incidents and the number of EMS units available on an hourly basis. As shown in Figure 23 the number of EMS units increases on Friday afternoons until evening time. However, this increment in the number of EMS units applies also for Thursdays and Saturdays even though the incidents reach to their peaks in different time slots (in morning and evening). Moreover, the number of EMS units in other days is almost the same regardless of the incidents pattern.

4.1.4. Incidents-Zone Analysis

Qatar is divided into ninety two zones. Figure 24 shows the incidents density in various zones in the state of Qatar. Historical data for one year showed that there are some zones that could be considered as “black zones” with respect to the number of incidents occurring in these zones.
As shown in Table 6, these zones are 57, 55, 53, 37 and they are comprising more than 5% of the total incidents. Moreover, it was observed that around thirty zones are responsible for triggering 80% of the incidents. These zones are presented in green color in Figure 25 and are listed in the below table (Table 6) in descending order based on number of incidents. It is worth mentioning that other zones trigger less than 1% of incidents each.

Figure 23: Available EMS Units Every Hour.
Figure 24: Zone Incidents Density
Table 6: Zones That Have More Than 1% Occurrence of Incidents.

<table>
<thead>
<tr>
<th>Zone #</th>
<th>Number of Incidents</th>
<th>%</th>
<th>Zone #</th>
<th>Number of Incidents</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>16278</td>
<td>13.18%</td>
<td>34</td>
<td>1904</td>
<td>1.54%</td>
</tr>
<tr>
<td>55</td>
<td>8353</td>
<td>6.76%</td>
<td>91</td>
<td>1763</td>
<td>1.43%</td>
</tr>
<tr>
<td>53</td>
<td>7805</td>
<td>6.32%</td>
<td>38</td>
<td>1737</td>
<td>1.41%</td>
</tr>
<tr>
<td>37</td>
<td>7338</td>
<td>5.94%</td>
<td>46</td>
<td>1720</td>
<td>1.39%</td>
</tr>
<tr>
<td>81</td>
<td>5326</td>
<td>4.31%</td>
<td>92</td>
<td>1702</td>
<td>1.38%</td>
</tr>
<tr>
<td>60</td>
<td>4408</td>
<td>3.57%</td>
<td>54</td>
<td>1667</td>
<td>1.35%</td>
</tr>
<tr>
<td>56</td>
<td>4180</td>
<td>3.39%</td>
<td>70</td>
<td>1511</td>
<td>1.22%</td>
</tr>
<tr>
<td>71</td>
<td>4110</td>
<td>3.33%</td>
<td>26</td>
<td>1505</td>
<td>1.22%</td>
</tr>
<tr>
<td>90</td>
<td>3676</td>
<td>2.98%</td>
<td>30</td>
<td>1491</td>
<td>1.21%</td>
</tr>
<tr>
<td>39</td>
<td>3602</td>
<td>2.92%</td>
<td>32</td>
<td>1412</td>
<td>1.14%</td>
</tr>
<tr>
<td>51</td>
<td>3411</td>
<td>2.76%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>2610</td>
<td>2.11%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>2582</td>
<td>2.09%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>2505</td>
<td>2.03%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>2427</td>
<td>1.97%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>2175</td>
<td>1.76%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>2000</td>
<td>1.62%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 25: Number of Incidents s. Zones
As mentioned earlier, EMS providers in Qatar classify response priority into three main types. Figure 26 shows that life threat encompasses more than half of the incidents (i.e. 58.8% of the cases) whereas 39.2% of the cases classified as urgent and only 2.1% considered as transport.

![Response Priority](image)

Figure 26: Number of Incidents For Each Response Priority - 2012/2013

It is also worth to mention that 27.9% of the cases are considered as trauma, 69% are medical and 3.1% are transport (see Figure 27).

![Case Classification](image)

Figure 27: Total Number of Cases for Each Class - 2012/2013
It is obvious from Figure 28 and Figure 29 that the usage of Alpha units throughout the year is extremely more than the other units. However, it is important to mention that Charlie units have been used in more than the quarter of life threat cases. This is because Charlie units, as mentioned earlier, are used for critical cases as they are stuffed with advanced equipment and staffs.

Figure 28: EMS Usage with Respect to Case Classification

Figure 29: EMS Usage with Respect to Response Priority
4.2. Analysis of Alternative EMS Systems

In this section, the performance of three EMS systems will be compared. These include: (i) the existing EMS system, (ii) an EMS system developed through a GIS method (updated existing system), and (iii) an EMS system developed by the fuzzy k-means algorithm. The comparison of the EMS systems was based on two main performance metrics i.e. (a) service area, and (b) closest facility.

4.2.1. Existing EMS System

In order to evaluate the existing EMS, the locations of the EMS units were integrated into the GIS environment. For the purpose of relative comparisons, the locations entered into the GIS environment did not differentiate between hubs and spokes. In addition, these locations were superimposed on the district layer of the map of Qatar. This procedure was necessary since the proposed EMS system design is based on districts rather than the hub catchment areas as is the case with the existing system. Since the winter system was envisaged to be better than the summer system, the winter EMS system was used for all comparative analysis. The locations of the EMS units are shown in Figure 30. These locations were then analyzed through GIS network analysis. Two types of analyses were carried out namely, the service area and the closest facility.

4.2.1.1. Service Area Analysis

The service area analysis was applied for ambulance response times of 8, 10, and 15 minutes. The service areas covered are shown in Figure 31.

Table 7 summarizes the results. It can be observed from both Figure 31 and Table 7 that as the ambulance response time increases, the service coverage is improved for population, districts, and the populated districts. This trend is an expected result. It can also be observed that 70.82% of districts are covered considering a 15 minutes service area. Values for the 8 and 10 minutes service area are relatively low.
Figure 30: EMS Locations of Existing System
Table 7: Summary of Service Area Analysis – Existing System

<table>
<thead>
<tr>
<th>Data Related to The Population</th>
<th>Considering 8 Minutes</th>
<th>Considering 10 Minutes</th>
<th>Considering 15 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Population Covered</td>
<td>99.352</td>
<td>99.39</td>
<td>99.895</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
<td>0.648</td>
<td>0.61</td>
<td>0.105</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Related to The Districts</th>
<th>Percentage of Districts Covered</th>
<th>Percentage of Uncovered Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.398</td>
<td>49.602</td>
<td></td>
</tr>
<tr>
<td>55.57029</td>
<td>44.43</td>
<td></td>
</tr>
<tr>
<td>70.822</td>
<td>29.178</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Related to The Populated Districts</th>
<th>Percentage of Populated Districts Covered</th>
<th>Percentage of Uncovered Populated Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>58.342</td>
<td>41.658</td>
<td></td>
</tr>
<tr>
<td>77.0128</td>
<td>22.9872</td>
<td></td>
</tr>
<tr>
<td>78.8884</td>
<td>21.1116</td>
<td></td>
</tr>
</tbody>
</table>

4.2.1.2. Closest Facility Analysis

The closest facility analysis is used to assign the best EMS for each population and/or incident points by finding the minimum time to the closest facility. Results of the closest facility analysis are summarized in Table 8. It can be observed that the results generated for the population data is better (values are lower) than those generated for the incidents data. This implies that a system developed based on population data may perform better than that based on incidents data. As is usually the case elsewhere, some service providers choose to base their system design on historical incidents rather than actual population. For both data types, it is found that more than 90% of data points are served within 10 minutes or less.
Service area of 8 minutes
Service area of 10 minutes
Service area of 15 minutes

Figure 31: Service Areas of the Existing System
Table 8: Closest Facility Results of the Existing System (minutes)

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>36.03695</td>
<td>67.414</td>
</tr>
<tr>
<td>Average</td>
<td>2.286398</td>
<td>3.471</td>
</tr>
<tr>
<td>Min</td>
<td>7.53E-05</td>
<td>9.6E-05</td>
</tr>
<tr>
<td>Median</td>
<td>1.588092</td>
<td>2.002</td>
</tr>
<tr>
<td>STD</td>
<td>2.631439</td>
<td>4.314</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>96.987</td>
<td>89.569</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>0.591</td>
<td>2.785</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>1.037</td>
<td>5.125</td>
</tr>
<tr>
<td>% more than 15</td>
<td>1.385</td>
<td>2.521</td>
</tr>
</tbody>
</table>

4.2.2. Proposed GIS Based EMS System (Updated Existing System)

For the GIS based methods, two models were used in the experiments. These models are the minimize impedance (which is essential a p-median model) and the maximize coverage. The location of the EMS units generated based on two GIS models (minimize impedance model and maximize coverage model) is shown in Figure 32.

Figure 32 shows that coincidentally, the maximize coverage and the minimize impedance methods in the GIS environment specifies exactly the same positions of the EMS. Therefore, further analysis was based on the maximize coverage model only since the other would give exactly the same performance measures for service area analysis and closest facility analysis.
Figure 32: Location of EMS units based on p-median and maximize coverage models
Using the maximize coverage model, a GIS based EMS system was developed by seeding existing locations of EMS units. This was regarded as an exercise to update the existing using the recent population data. The reason for this procedure was to find out if the GIS environment can be used iteratively i.e. in two steps to provide a better EMS system. Thus, all existing EMS units’ locations were seeded as candidates locations and the GIS asked to select 49 units (the exact number of existing EMS units) in such a way to meet the objective (i.e. maximize the coverage or minimize the response time). It was expected that all 49 units will be selected. However, even at a cut-off value of 20 to 60 minutes, it was observed that only 43 units of the existing were selected and the rest of the unselected EMS units were considered as redundancy i.e. units that do not add any value to the EMS system (see Figure 33). This could be an indication that the existing system is over designed. This suggestion also seems to be supported by the cluster validation method in which 46 EMS units were deemed sufficient for the EMS system in Qatar using the recent population data.

Figure 33: Network Analyst Message

4.2.2.1. Service Area Analysis
The service area analysis, showed that more than 50% of the total and populated districts in State of Qatar is covered considering a target of 10 minutes, while 70% is covered if the target is 15 minutes (Figure 34 and Table 9)


<table>
<thead>
<tr>
<th>Data Related to The Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Population Covered</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Related to The Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Districts Covered</td>
</tr>
<tr>
<td>Percentage of Uncovered Districts</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Related to The Populated Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Populated Districts Covered</td>
</tr>
<tr>
<td>Percentage of Uncovered Populated Districts</td>
</tr>
</tbody>
</table>

4.2.2.2. Closest Facility Analysis

Results of the closest facility analysis are shown in Table 10. From Table 10, it is clear that average response time does not exceed 3.2 minutes for both population and incidents data. Moreover, more than 90% of data points are served within 8 minutes.

Table 10: Closest Facility Results of the Updated Existing System (minutes)
4.2.3. New System Based on Fuzzy k-means

A new EMS system for Qatar was generated through integrated use of fuzzy k-means and GIS. The fuzzy clustering algorithm was used to propose new locations of EMS units using the same existing settings (i.e. 49 EMS units) based on incidents data for 2012/2013.

<table>
<thead>
<tr>
<th>Population</th>
<th>Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>40.304</td>
</tr>
<tr>
<td>Average</td>
<td>2.373</td>
</tr>
<tr>
<td>Min</td>
<td>0.00008</td>
</tr>
<tr>
<td>Median</td>
<td>1.638</td>
</tr>
<tr>
<td>STD</td>
<td>2.869</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>96.697</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>0.583</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>1.084</td>
</tr>
<tr>
<td>% more than 15</td>
<td>1.635</td>
</tr>
</tbody>
</table>

Figure 35 shows a Matlab graph of the various clusters and center of clusters as generated by the fuzzy k-means algorithm. In this map, the dark spots are the centers of the clusters.
i.e. EMS units and the colored polygons are the various clusters suggested by the fuzzy k-means algorithm.

The proposed location of EMS units generated by the fuzzy k-means algorithm was integrated into the GIS environment. Figure 36 shows the results of this integration in a GIS map showing the locations of the EMS units.
Figure 35: Matlab Graph of the Various Clusters and Center of Clusters for the EMS System Proposed by the Fuzzy k-means Algorithm
Figure 36: EMS Locations Based on Integrated Use of Fuzzy k-means Clustering and GIS
4.2.3.1. Service Area Analysis

Table 11 and Figure 37 summarize the service area analysis. The service area analysis shows that the percentage of districts and populated district covered is less than the uncovered when the target is 8 or less. However, the coverage increased as the target threshold increases as well (Table 11). Covered areas are shown in Figure 37.

Table 11: Summary of Service Area Analysis – fuzzy k-means Clustering System

<table>
<thead>
<tr>
<th>Data Related to The Population</th>
<th>Considering 8 Minutes</th>
<th>Considering 10 Minutes</th>
<th>Considering 15 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Population Covered</td>
<td>99.373</td>
<td>99.381</td>
<td>99.972</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
<td>0.627</td>
<td>0.619</td>
<td>0.028</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Related to The Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Districts Covered</td>
</tr>
<tr>
<td>Percentage of Uncovered Districts</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Related to The Populated Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Populated Districts Covered</td>
</tr>
<tr>
<td>Percentage of Uncovered Populated Districts</td>
</tr>
</tbody>
</table>

4.2.3.2. Closest Facility Analysis

Table 12 summarizes the results of the closest facility analysis. It can be observed that the results for both population and incidents are near to each other. However, the population data is slightly better than the incidents data.
Table 12: Closest Facility Results of the Fuzzy k-means Clustering System (minutes)

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>40.454</td>
<td>51.733</td>
</tr>
<tr>
<td>Average</td>
<td>3.455</td>
<td>3.682</td>
</tr>
<tr>
<td>Min</td>
<td>0.000156</td>
<td>3.05E-05</td>
</tr>
<tr>
<td>Median</td>
<td>2.268</td>
<td>2.110</td>
</tr>
<tr>
<td>STD</td>
<td>3.462</td>
<td>4.284</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>87.922</td>
<td>86.305</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>7.512</td>
<td>5.384</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>3.106</td>
<td>5.136</td>
</tr>
<tr>
<td>% more than 15</td>
<td>1.461</td>
<td>3.175</td>
</tr>
</tbody>
</table>

4.3. Performance Comparison of the Alternative EMS Systems

In order to compare the effectiveness of the existing and a proposed EMS systems results of service area analysis and closest facility performance metrics were used. The purpose of this section is to compare three alternative EMS system generated by three different methods. These three alternative EMS systems include: (i) Alternative I, which is the currently existing system that was designed based on the two-step floating catchment area method, (ii) Alternative II, which is a proposed system (updated existing system) based on the GIS method (maximal coverage), and (iii) Alternative III, which was designed based on the integrated use of fuzzy k-means and a GIS method. Figure 38, Figure 39, and Figure 40 show that in general the existing system performance (Alternative I) is more or less similar to and the EMS system designed based on the integrated use of the fuzzy k-means clustering algorithm and that from GIS (Alternative III). However, it can be observed that Alternative III gives slightly better results with respect to population coverage, while the existing system gives better results with respect to districts and populated districts coverage.
Service area of 8 minutes
Service area of 10 minutes
Service area of 15 minutes

Figure 37: Service Areas of The Fuzzy k-means Clustering System
Figure 38: Percentage of Covered Population for the Three Alternative EMS Systems

Figure 39: Percentage of Covered Districts - Existing, Updated, and Fuzzy k-means Systems

134
Based on the population data, it is found that the existing system generate better results if we are considering the population data points. However, the new system proposed by the GIS method gives almost the same results as the existing system although the EMS units is only 43 i.e. less by 6 units. This may suggest that the redundancy in the existing system makes the system more costly but without any gain in the performance of the system. Moreover, the updated existing system gives better results if we take into consideration the incidents data (see Table 13 and Table 14). The fuzzy k-means model implemented here does not give better results than either the existing or the updated system. In order to explain why this may be so, further investigations were conducted with the fuzzy k-means method. The inspiration for this further investigation rose from the fact that in implementing the fuzzy k-means methods a number of strategies and procedures can be used and may give different results.
Table 13: Closest Facility Results for the Existing, Updated, and Fuzzy k-means Clustering Systems- Population

<table>
<thead>
<tr>
<th></th>
<th>Existing System</th>
<th>Fuzzy k-means Clustering System</th>
<th>Updated Existing System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>36.03695</td>
<td>40.454</td>
<td>40.304</td>
</tr>
<tr>
<td>Average</td>
<td>2.286398</td>
<td>3.455</td>
<td>2.373</td>
</tr>
<tr>
<td>Min</td>
<td>7.53E-05</td>
<td>0.000156</td>
<td>0.00008</td>
</tr>
<tr>
<td>Median</td>
<td>1.588092</td>
<td>2.268</td>
<td>1.638</td>
</tr>
<tr>
<td>STD</td>
<td>2.631439</td>
<td>3.462</td>
<td>2.869</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>96.987</td>
<td>87.922</td>
<td>96.697</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>0.591</td>
<td>7.512</td>
<td>0.583</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>1.037</td>
<td>3.106</td>
<td>1.084</td>
</tr>
<tr>
<td>%more than 15</td>
<td>1.385</td>
<td>1.461</td>
<td>1.635</td>
</tr>
</tbody>
</table>

Table 14: Closest Facility Results for the Existing, Updated, and Fuzzy k-means Clustering Systems- Incidents

<table>
<thead>
<tr>
<th></th>
<th>Existing System</th>
<th>Fuzzy k-means Clustering System</th>
<th>Updated Existing System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>67.414</td>
<td>51.733</td>
<td>54.385</td>
</tr>
<tr>
<td>Average</td>
<td>3.471</td>
<td>3.682</td>
<td>3.183</td>
</tr>
<tr>
<td>Min</td>
<td>9.60E-05</td>
<td>3.05E-05</td>
<td>0.00006</td>
</tr>
<tr>
<td>Median</td>
<td>2.002</td>
<td>2.11</td>
<td>1.629</td>
</tr>
<tr>
<td>STD</td>
<td>4.314</td>
<td>4.284</td>
<td>4.358</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>89.569</td>
<td>86.305</td>
<td>90.01</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>2.785</td>
<td>5.384</td>
<td>2.299</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>5.125</td>
<td>5.136</td>
<td>4.637</td>
</tr>
<tr>
<td>%more than 15</td>
<td>2.521</td>
<td>3.175</td>
<td>3.054</td>
</tr>
</tbody>
</table>
4.4. Hybrid Data Clustering and GIS Simulations

4.4.1. Model-Based Strategies for Developing EMS Systems

The results discussed in section 4.3 have shown that in general different models for EMS planning can result in EMS systems that have different performance metrics. One of the objectives of this thesis is to simulate and compare the effectiveness of a variety of model-driven strategies for developing and planning EMS systems. Answers to such investigations can be used by EMS planners as recommendations for strategizing EMS planning with various objectives or for different seasons of the year. In order to provide information that can guide EMS planners in selecting a suitable model-driven strategy or procedure for designing and developing EMS systems using data mining techniques and GIS methods, a number of strategies for EMS planning were simulated in the GIS environment using the same data from the case study. The results of such simulations are reported in this section.

A total of eight models were configured for this simulation based on the flowchart method represented in Figure 19 in chapter 3. The Matlab code and results of fuzzy k-means clustering is shown in Appendices F and G respectively. In addition to the eight models, a comprehensive model is proposed by seeding the existing EMS locations and those generated by fuzzy k-means clustering. The significance of this comprehensive model lies in the logic that a lot of data from different data sources is usually available to EMS planners. Therefore using such data may result in a more informed EMS system which is expected to provide superior performance.

The GIS part of the comprehensive model uses the maximize coverage option in GIS. The eight proposed systems were initially divided into two types: (1) systems that consider the concepts of hubs and spokes as is the case with the EMS system in Qatar, and (2) systems that do not differentiate between hubs and spokes. In the design of EMS units using the integrated approach proposed in this thesis, this difference is basically procedural in the design process. Procedural alternatives are usually available to EMS
planners, for which the findings of this thesis will provide guidelines on which procedures provide better performance in the strategic planning of the EMS system. Other available strategies include options such as direct or indirect clustering. The direct clustering finds the best EMS locations directly without clustering first the houses into groups according to their locations. On the other hand, indirect clustering first groups the houses into 100 clusters based on their nearness, assuming that every 1% of the populations are accommodated near to each other. After this clustering, the model is then required to find the best 49 EMS locations. In addition to that, the direct and indirect clustering is divided further into two streams, one stream based on population and the other stream based on the incidents. Different divisions of proposed models are represented in Figure 41 and summarized in One of the main concepts used in this research is data mining. After being supplied by raw data (population data or incidents data), the fuzzy k-means algorithm clusters the data as shown in Figure 35. In order to derive meaning from these data, data mining algorithms were used to calculate performance metrics from the clustered data. For this analysis only models (a) to (h) were evaluated since in these models a number of strategies were coded in Matlab software in which the fuzzy k-means algorithm was implemented. The results of such analysis are reported in the following sections.

Table 15.

Figure 41: Proposed Systems.
One of the main concepts used in this research is data mining. After being supplied by raw data (population data or incidents data), the fuzzy k-means algorithm clusters the data as shown in Figure 35. In order to derive meaning from these data, data mining algorithms were used to calculate performance metrics from the clustered data. For this analysis only models (a) to (h) were evaluated since in these models a number of strategies were coded in Matlab software in which the fuzzy k-means algorithm was implemented. The results of such analysis are reported in the following sections.

Table 15: Description of Model Strategies Simulated in the GIS Environment

<table>
<thead>
<tr>
<th>Model code</th>
<th>Description</th>
<th>Model Abbreviations</th>
</tr>
</thead>
</table>
| (a)        | Direct: find EMS directly without clustering houses first  
- With Hubs and spokes: considering that there are two different types of EMS locations.  
- Taking into account the whole population in Qatar. | Direct-H&S-Qatar |
| (b)        | Direct: find EMS directly without clustering houses first  
- With Hubs and spokes: considering that there are two different types of EMS locations.  
- Taking into account incidents only. | Direct-H&S-Incidents |
| (c)        | Direct: find EMS directly without clustering houses first  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account incidents only. | NoH&S-Incidents |
| (d)        | Direct: find EMS directly without clustering houses first  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account the whole population in Qatar. | Direct-NoH&S-Qatar |
| (e)        | Indirect: cluster houses first and then find EMS locations  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account the whole population in Qatar. | Indirect-NoH&S-Qatar |
| (f)        | Indirect: cluster houses first and then find EMS locations  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account incidents only. | Indirect-NoH&S-Incidents |
| (g)        | Indirect: cluster houses first and then find EMS locations | Indirect-H&S- |
Running the Matlab code is differs slightly from one model to another in terms of the inputs and/or the execution pattern as shown in Figure 19. However, the final output for all is the same which is mainly the location of the proposed EMS units. In addition to that, the number of people served by each EMS unit and the distance between the nearest EMS units and each data point are calculated from the clustered data. Using a computer with 2.67GHz processor and 4GB RAM, the time required to execute different strategies is represented in Table 16. All of the other results are shown Appendix G.

Table 16: Time Execution of Matlab Code

<table>
<thead>
<tr>
<th>Model</th>
<th>Execution Time (Seconds)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>147.895</td>
<td>1</td>
</tr>
<tr>
<td>b</td>
<td>323.7</td>
<td>6</td>
</tr>
<tr>
<td>c</td>
<td>304.806</td>
<td>5</td>
</tr>
<tr>
<td>d</td>
<td>156.422</td>
<td>2</td>
</tr>
<tr>
<td>e</td>
<td>276.566</td>
<td>4</td>
</tr>
<tr>
<td>f</td>
<td>327.705</td>
<td>7</td>
</tr>
<tr>
<td>g</td>
<td>276.108</td>
<td>3</td>
</tr>
<tr>
<td>h</td>
<td>335.51</td>
<td>8</td>
</tr>
</tbody>
</table>

The rank of the models with respect to execution time is also shown in Table 16. Table 16 shows that different strategies require different execution time. The shortest execution time is only about 2.5 minutes while the largest execution time is about 5 minutes. These differences in implementing an available algorithm for fuzzy clustering and information
extracting are of no practical significance. Therefore, it can be concluded that algorithm run time when using the fuzzy k-means algorithm for the development of an EMS units under the conditions described in this thesis is very small and hence has little impact on the design of an EMS system. This result can also mean that EMS planners can evaluate and re-evaluate the performance of an existing EMS in no time and hence update plans for an existing system can be easily generated. Therefore, most of the effort in terms of time and cost will be in the practical implementation of the proposed solution for example re-location EMS facilities.

In Figure 41, the comprehensive model (model (i) in Table 15) was created by seeding all the possible EMS locations available from different models created through different methods and procedures. This scenario is necessary for comparison since it simulates a situation in which the designers and developers of the EMS system have valuable information regarding how the EMS system should work. It is important to mention that most of EMS locations specified in the comprehensive model (41 in this case) were taken from the location seeded by the fuzzy clustering algorithm and only 8 of them were taken from the existing system. The location of proposed EMS units for the nine models is shown in Figure 42 (a-i).

For simplicity and to mimic the real situation, it is assumed that each house consist of 14 family members (Table 17).

Table 17: Number of Family Members in Qatar

<table>
<thead>
<tr>
<th>Zone #</th>
<th>Number of Buildings</th>
<th>Number of Population</th>
<th>Family Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>1217</td>
<td>14725</td>
<td>12.09942</td>
</tr>
<tr>
<td>33</td>
<td>453</td>
<td>5197</td>
<td>11.47241</td>
</tr>
<tr>
<td>34</td>
<td>2167</td>
<td>35125</td>
<td>16.20904</td>
</tr>
<tr>
<td>35</td>
<td>516</td>
<td>7702</td>
<td>14.92636</td>
</tr>
<tr>
<td></td>
<td><strong>Average =</strong></td>
<td></td>
<td><strong>13.68≈14</strong></td>
</tr>
</tbody>
</table>

134
This is calculated using the data of Qatar Area Referencing system for four zones in Doha given by the Ministry of Municipality and Urban Planning and by considering the below equation.

Number of Family Members = Average (population at each zone/buildings at each zone).
In this study, the districts were divided into In-Doha and Out-Doha as shown in Figure 43.
4.4.2. Comparison of Model Based Strategies for Developing EMS Systems

4.4.2.1. Service Area Analysis
After implementing service area analysis, the area covered for each system is represented in Figure 44 and detailed results provided in appendix C. The analysis will take mainly three ways; namely, the analysis of the districts, the analysis of the population, and the analysis of the covered populated area.

For districts, it is logically clear that as we reduce the response time (the travel time), the percentage of uncovered districts will be increased. Thus, if the response time set to be 8 minutes, larger number of districts will not be covered but among the whole ten models, the (indirect-No H&S-incidents) model gave the best results while the (indirect-H&S-Qatar) model gives the worst (Figure 45). For response times of 10 and 15 minutes, the best performance was observed for indirect-No H&S-incidents model while the worst performance was for the indirect-NoH&S-Qatar model. In overall, it can be inferred that the indirect-No H&S-incidents model provides the best performance for all time ranges in terms of district coverage.

In the previous paragraph, models were analyzed and compared with respect to the districts covered. However, the results of such an analysis do not help us in determining properly how many people are covered. Therefore, it is necessary to analyze and compare among these models in terms of how many populations are not covered. As represented in Figure 46, it is clear that there are three models that can be considered as extremes and these will be discarded from the analysis. These extremes are:

- Indirect, No hubs and Spokes, Incident Model
- Indirect, Hubs and Spokes, Qatar Model
- Indirect, Hubs and Spokes, Incidents Model.
Figure 44: Service Area of Proposed and Existing Models.
Figure 45: Percentage of Uncovered Districts

Figure 46: Number of Uncovered Population.

After filtering out the three models (see Figure 47), it is found that the indirect-NoH&S-Incident model and the comprehensive models have the minimum values (population uncovered) in all time ranges while the worst model is shared between the (Direct-H&S-Incidents) model and the (Direct-NoH&S-Incidents) model. It can be observed that the comprehensive model has more improved outputs in terms of population coverage as the response time target is reduced to 10 and 8 minutes, while indirect-NoH&S-Incident model has better results if the target is 15 minutes (see Table 18). It is also found that the uncovered population in the existing system is more than 79% of the uncovered
population in the comprehensive and indirect-NoH&S-Incident model in all time ranges. Thus, it can be inferred that the improvement of both the comprehensive and the indirect-NoH&S-Incident model is more than 79% than the existing model.

![Percentage of uncovered population](image)

**Figure 47: Number of Uncovered Population-Filtered.**

Table 18: Uncovered Population- (Indirect-NoH&S-Incident) and Comprehensive Models

<table>
<thead>
<tr>
<th></th>
<th>Indirect-NoH&amp;S-Incidents</th>
<th>The comprehensive model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 min</td>
<td>10 min</td>
</tr>
<tr>
<td>Uncovered population</td>
<td>4310</td>
<td>1307</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
<td>0.277</td>
<td>0.084</td>
</tr>
</tbody>
</table>

In order to decide which model has better coverage, an analysis that focuses on determining how much populated area covered by each model is carried out. As shown in Figure 48, it is found that the best model in terms on covered populated area is the comprehensive model with a percentage exceeding 83% in all time ranges and more specifically when the response times become less.
Thus, from the above analysis, it could be concluded that the comprehensive model has more improved results especially for small response time targets.

4.4.2.2. Closest Facility Analysis

In this analysis, EMS service will be analyzed based on both population and incidents distribution. GIS is tasked to find the nearest EMS unit to each population/incident point using the closest facility tool under the network analysis-ArcGIS. Thus, GIS is used to simulate and find the smallest response time required to serve the incidents. Two tables were generated to summarize the results of simulating the populations and incidents data points. They are represented in Table 20 and Table 20 for population and incidents respectively. Ideally a small response time is often sought after. Therefore, among all aforementioned statistics i.e. maximum, minimum and average, the minimum times are highlighted in green and it is found that all of these minimum values can be obtained from the eight suggested models and none related to the existing EMS system. For instance, the minimum average value is found from model (g): Indirect-H&S-Qatar if we are taking into consideration the population data points, and it is model (a): Direct-H&S-Qatar if we are considering the incidents data points.
Table 19: Closest Facility Results - Population

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>Existing</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>46.44373</td>
<td>39.98912</td>
<td>40.454</td>
<td>46.572</td>
<td>47.40103</td>
<td>37.90408</td>
<td>47.65405</td>
<td>30.34855</td>
<td>36.03695</td>
<td>45.94061</td>
</tr>
<tr>
<td>Average</td>
<td>2.035561</td>
<td>3.331505</td>
<td>3.4553</td>
<td>1.927</td>
<td>2.631258</td>
<td>3.346153</td>
<td>2.698915</td>
<td>3.670968</td>
<td>2.286398</td>
<td>1.962472</td>
</tr>
<tr>
<td>Min</td>
<td>0.000301</td>
<td>0.000534</td>
<td>0.000156</td>
<td>5.89E-05</td>
<td>0.000181</td>
<td>0.000191</td>
<td>2.37E-05</td>
<td>0.000154</td>
<td>7.53E-05</td>
<td>0.000534</td>
</tr>
<tr>
<td>Median</td>
<td>1.336902</td>
<td>1.953867</td>
<td>2.267503</td>
<td>1.351</td>
<td>1.432455</td>
<td>2.178519</td>
<td>1.59097</td>
<td>2.293847</td>
<td>1.588092</td>
<td>1.504223</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>97.433</td>
<td>88.762</td>
<td>87.922</td>
<td>93.336</td>
<td>89.980</td>
<td>93.716</td>
<td>89.245</td>
<td>96.987</td>
<td>98.710</td>
<td></td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>1.650</td>
<td>6.318</td>
<td>7.512</td>
<td>1.400</td>
<td>1.387</td>
<td>6.116</td>
<td>1.272</td>
<td>5.185</td>
<td>0.591</td>
<td></td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>0.532</td>
<td>3.684</td>
<td>3.106</td>
<td>0.303</td>
<td>1.796</td>
<td>2.907</td>
<td>1.839</td>
<td>3.016</td>
<td>1.037</td>
<td></td>
</tr>
<tr>
<td>%more than 15</td>
<td>0.385</td>
<td>1.235</td>
<td>1.461</td>
<td>0.374</td>
<td>3.482</td>
<td>0.996</td>
<td>3.173</td>
<td>2.555</td>
<td>1.385</td>
<td></td>
</tr>
</tbody>
</table>

Table 20: Closest Facility Results - Incidents

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>Existing</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>81.066</td>
<td>49.60076</td>
<td>51.73285</td>
<td>81.336</td>
<td>82.335</td>
<td>50.532</td>
<td>61.736</td>
<td>54.158</td>
<td>67.414</td>
<td>80.934</td>
</tr>
<tr>
<td>Min</td>
<td>1.924E-05</td>
<td>0.000249</td>
<td>3.05E-05</td>
<td>0.000498</td>
<td>0.000390</td>
<td>0.000134</td>
<td>6.989E-05</td>
<td>8.55E-05</td>
<td>9.6E-05</td>
<td>0.000248</td>
</tr>
<tr>
<td>Median</td>
<td>2.132</td>
<td>1.82827</td>
<td>2.110581</td>
<td>2.193</td>
<td>2.219</td>
<td>2.294</td>
<td>1.648</td>
<td>2.274</td>
<td>2.002</td>
<td>2.252</td>
</tr>
<tr>
<td>STD</td>
<td>5.053</td>
<td>4.374296</td>
<td>4.284488</td>
<td>5.0546</td>
<td>5.422</td>
<td>3.804</td>
<td>5.441</td>
<td>5.234</td>
<td>4.314</td>
<td>4.807</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>85.971</td>
<td>85.690</td>
<td>86.305</td>
<td>86.449</td>
<td>84.119</td>
<td>89.812</td>
<td>85.304</td>
<td>87.131</td>
<td>89.569</td>
<td>87.564</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>3.857</td>
<td>5.510</td>
<td>5.384</td>
<td>4.134</td>
<td>3.065</td>
<td>4.062</td>
<td>2.728</td>
<td>3.760</td>
<td>2.785</td>
<td>4.153</td>
</tr>
</tbody>
</table>
The other part of the analysis is related to the number of points served in specific range of time and in this research we are focusing on response time of 10 minutes or even less to meet other international standards and also to be able to serve most of emergency cases such as cardiac arrests which needs response time of less than 10 minutes. We can also determine from this analysis which model meets the equity metric and serve fairly most of the population/incidents without differentiating between population/incidents in or out of Doha. It can be observed that the comprehensive model has better results than other models in terms of population data points and model (f): Indirect-NoH&S-Incidents is better if we are taking into consideration the incidents.

In addition to that and since equity is one of the important metric the response time of rural areas in the State of Qatar was studied separately to detect and have general idea about the performance of EMS units in such areas. According to the current ambulance service provider in Qatar, the targets of the current system in urban areas focuses on serving incidents within 10 to 15 minutes whereas the target for rural areas is within 15 to 20 minutes. Thus, there is a difference between urban and rural areas. However, as seen in Table 21, there is improvement in rural areas response time. For example, the average and median of the response time is around 4 minutes in model (d) considering the population data points. Moreover, most of the points (around 94%) will be served within 10 minutes. This percentage is higher in comparison to the existing system which does not aim to exceed 80%. However, the difference in the rural and urban areas coverage in the existing system is probably due to other factors such as operational costs and population density which favors the densely populated areas in Doha City.

For the data generated by analyzing the incidents (Table 22), it is found that the median and average values are higher than the results generated with respect to the population. However, still the proposed systems have better results over the existing one. Thus, model (b) outperforms others by having the minimum average and median and maximum 8 to 10 minutes coverage percentage. This gives an indication that the suggested models address the equity metric better than the existing one.

4.4.3. Statistical Analysis
Simple comparison between these models using for example the average is not enough to generate conclusions due to random noise of the data. Thus, statistical analysis was used to derive more information from the results. The purpose of this section is to report comparisons among the distribution of response time generated using each model.

Using the response time data generated by the closest facility analysis, it is found that none of the data from the ten models follows a normal distribution (Appendix D) and data is independent. Therefore, the Kruskal-Wallis test, one of non-parametric tests, was used to find if there could be a statistical significance among the data sets for response times.

Figure 49 shows the results obtained from Minitab. The \( p\)-value for all response time data sets was found to be less than 0.05. Therefore, it can be inferred that the differences in response time among all models is statistically significant. Figure 49 shows \( p\)-values of 0.000 for the unadjusted and adjusted for ties methods indicating that the null hypothesis that there is no difference between the response time medians can be rejected.

| \( H = 61585.74 \) | DF = 9 | \( p = 0.000 \) |
| \( H = 61585.74 \) | DF = 9 | \( p = 0.000 \) (adjusted for ties) |

Figure 49: Results of Kruskal-Wallis Test. The Spanish EMS was built upon some critical operational factors (Fogue, et al., 2013) which include: cost, resource utilization, and most importantly the assistance quality usually measured in terms of response time and rate, survival rate, and equity. These factors depend heavily on EMS capacity.
1.6.1. EMS Capacity

EMS capacity has many dimensions. It could mean the number of emergency vehicles and personnel available at each EMS facility. It also could mean the coverage area that each emergency facility/vehicle could cover, or the number of patients that each vehicle can transport at a time. These aspects play an important role in determining the best location of capacitated facilities in order to maximize the response and survival rates. On the other hand, the concern in incapacitated facilities is only to meet the desired response and survival rates without limitations in the number of facilities placed. Therefore, the survival and response rates are mostly proportional to the number and capacity of the facilities that need to be located. In most cases it is hard to achieve the desired response and survival rates in the capacitated EMS location problems. Likewise, the response and survival rates also depend on the number and location of health care facilities and hospitals.
1.6.2. Equity

One of the important objectives that affects the strategic planning decision and used to measure the system performance in EMS systems is equity (Beraldi, et al., 2004; Hosseini & Saeed, 2011). The meaning of equity in EMS systems is to provide fair treatment and service among all residence regardless of their locations. As mentioned by Rawlinson and Crews, there is usually a large discrepancy between rural and urban areas when providing EMS services (Chanta, et al., 2011). Such discrepancies contradict the equity objective in designing an EMS system. In certain situations, equity is critical to a proposed EMS design.

1.6.3. Response Rate and Time

Response rate and time are among the most important indicators for assessing the performance of EMS systems. Response rate simply means the percentage of on-time arrival of an ambulance in response to a community emergency need whereas response time covers the period from receiving an urgent call until an ambulance reaches the scene (Breen, et al., 1999). As stated by American Heart Association within four to six minutes after cardiac arrest, the brain death and permanent death occur and the chance to survive is reduced by 7% to 10% with every minute that passes without advanced life support assistance (Ludwig, 2004). Since very few people under cardiac arrest can survive after ten minutes, the EMS response time should be close to ten minutes.

Based on the National Fire Protection Association 1710 standard, 90% of EMS calls have to be covered within four minutes or less. Furthermore, the association recommended an arrival of 8 minutes 90% of the times, if the service provided is an ALS service. This does not preclude the four-minute initial response (Office of Strategic Health Autohrities, 2009; Wilde, 2009).

1.6.4. Survival Rate

The survival rate is related to the response time. It is increased as we decreasing the response time. It has been found that the survival rate will be increased by 12%, if the patients were
reached within five minutes instead of eight minutes (Office of Strategic Health Authorities, 2009). This means the life of about 86 persons will be saved/survived if the response time is decreased (Office of Strategic Health Authorities, 2009). While the response rate and survival rate heavily depend on EMS capacity a distinction between capacitated and incapacitated EMS facilities is crucial in addressing operational problems effectively.

As mentioned before, incapacitated facilities are facilities that have unlimited number of emergency vehicles and personnel so that it can transport as many patients as possible, and can cover unlimited area. Such facilities usually suffer from over-design problems as a result of trying to achieve performance targets. On the other hand, capacitated facilities have limited number of emergency vehicles and personnel which in turn will limit the coverage area and the number of transported patients. In this study, the focus is on multi-facility location allocation issues within EMS systems.

1.7. Multi-Facility Location Allocation

Facility location allocation problems require spatial analysis techniques to find the best location for supplying objects with respect to their demands by using a set of objectives and constraints. In reality, most of the facilities; especially those facilities consisting of single type (mostly static facilities), are constrained with limited resources which makes it incapable to meet the demand sufficiently. Examples of such systems include the health care systems that have a limited number of hospitals. In general, hospitals are usually located in urban areas and therefore are capable of covering only the need of residences who are living in the urban cities, whereas the people who are living in rural areas will not be covered properly. In such a case, there is a need to provide another type of facilities (temporary or transportation facilities) so that it can meet all demand points.

Moreover, using hospitals only, which represent static facilities, is not enough to cover all demand points in urban arenas. In some cases the time taken by a patient to reach a hospital is very long. Therefore, transportation facilities such as ambulance vehicles and temporary facilities such as temporary ambulance stations play an important role in maximizing the
coverage area and minimizing the transportation time which is very critical for the patients’ health and life (Gu & Wang, 2010). Figure 2 represents different types of healthcare facilities.

<table>
<thead>
<tr>
<th>Static Facility</th>
<th>Temporary Facility</th>
<th>Transportation Facility</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Static Facility Image]</td>
<td>![Temporary Facility Image]</td>
<td>![Transportation Facility Image]</td>
</tr>
</tbody>
</table>

Figure 2: EMS Multi-facilities

It is worthy to mention that the multi-facility location allocation problem is more flexible than the single-type facility location problem and therefore it is expected to have better solutions and better results. Taking the EMS example, there are many objectives which an EMS provider seeks to meet. For instance, EMS provider may seek to have enough capacity to cover all demand points in different areas of the country. This objective cannot be achieved with single-type facility. On the other hand, by having transportation facilities such as ambulances located in different areas in the country, the coverage area of EMS will increase. In other words, it will increase the capacity of EMS as well as reduce the response time and increase the survival rate since the ambulances will provide the required pre-hospital care which will reduce the disabilities and death.

1.8. Facility Location Allocation Models
For the past decades, facility location allocation models have been applied to solve EMS problems. Facility location allocation problems can be classified in accordance with the objective function, constraints, solution methods, and demand patterns. In what follows, different classes of facility location allocation models will be discussed based on various solution approaches.

Solution methods for facility location allocation models can be classified into classical and non-classical or contemporary approaches. The classical approach is simply the use of mathematical programming and enumeration methods to optimally solve particular problems. As mentioned by Vanegas, et al. (2008) the generated solutions from this kind of models are a tradeoff between different objectives so as to achieve the contiguity and compactness of the requirements. These kinds of models are mostly used as an optimality reference to validate the results generated from the non-classical approaches. Some famous examples of classical models are the linear and integer programming. It is worthy to mention that these kinds of models are not capable of solving complex, large size real problems.

On the other hand, the non-classical and contemporary approaches are used to solve complex, large size problems. They are usually used to search among extremely large spaces to generate a solution which consists of only one point or small set of points (Lindeskov, 2002; Vanegas, et al., 2008). However, there is no guarantee that the generated solution is optimal but it is usually considered to be good or near optimal. Non-classical approaches are iterative such that at each successive stage an enhanced pattern of the solution is achieved. The search continues until a solution with a desired degree is reached. Some examples of non-classical and contemporary approaches are genetic algorithm, simulated annealing, and data mining. Recently, Geographic Information Systems have been proposed as crucial additions and enhancements to contemporary approaches to the solution of location allocation problems. It is worthy to mention that nowadays, many facility location allocation models combine more than one tool in order to generate more advanced and more practical solutions.

Although there is a lot of research accomplished in the location allocation field, relatively few of these researches have addressed the multi-facility location allocation problems. This means that
the exact nature of the multi-facility location allocation problems is still not well understood even-more so if a different context is considered. This research provides a unique contribution to location-allocation modelling by proposing an innovative approach to modeling multi-facility location allocations in emergency medical service systems.

1.9. Problem Description

EMS design and operational issues differ from country to country. In the State of Qatar, for example, the population is continuously increasing. As a result, the demand for emergency medical services will continue to rise. This requires an effective and efficient emergency medical service system that is capable of providing the EMS service proficiently. Albeit, it is sometimes necessary to evaluate and re-evaluate the performance of an existing EMS in order to avoid operating a sub-optimal EMS system. Evaluation and re-evaluation of the EMS system may be necessary because of; (a) rapid population growth, (b) rapid expansions in infrastructures such as new residential areas and new cities, which must be covered by the existing EMS, and (c) changes in both residential, commercial or industrial locations and/or re-locations. As such, evaluations and re-evaluations of EMS systems are crucial inputs required for strategic decisions and strategic planning for an effective EMS system.

In the state of Qatar, the EMS system incorporates more than one type of facility, namely; (i) static facilities, which are ambulance hubs and, (ii) transportation facilities, which are mainly represented as ambulance vehicles. As such, the EMS system naturally lends itself as a multi-facility location allocation problem. This represents a challenging problem because two types of facilities are considered and locations of transportation facilities depend on locations of static facilities and demand points or vice-versa (GU, et al., 2009). As such, an appropriate mixture of facilities is required to serve a given population. One way to address this problem is to determine optimal locations for both static and transportation facilities in such a way that cost are minimized. However, static and transportation facility location allocation problems are relatively new in facility location allocation sciences (GU, et al., 2009). Despite the growing interest in the field of location allocation, very few researches have addressed the multi-facility location
problems i.e. addressing both static and transportation facilities. Therefore, the importance of extending location allocation models to multiple-facility location-allocation problems can never be overemphasized.

In some attempts at solving the multi-facility location allocation problems, geographical information systems (GIS) have recently evolved as contenders for addressing these problems. As such, GIS is a very promising tool which offers multiple functions that could in principle be used for solving multiple facility location allocation problems. Therefore, this research focuses on integrating GIS and location-allocation models in order to solve location-allocation problems for the EMS system in the state of Qatar.

1.10. Objectives

The goal of this research is to develop an innovative approach for modeling, simulating and evaluating multi-facility location allocation problems in emergency medical service systems. This goal is achieved through an integrated use of; a fuzzy k-means algorithm, data mining techniques and GIS simulation. These tools are used concurrently to analyze and evaluate the performance of a designed EMS system. Since a lot of data will be used in this analysis, data mining techniques will be employed to extract valuable information that can be used for strategic decision making or for planning for an effective and efficient EMS system. Data mining applications to facility location allocation problems is in itself a relatively new area of research whose strengths will be exploited in this research. The ambulance service system in the State of Qatar will be used as a case study.

In order to meet the goal of this research, the following objectives will be addressed:

5. To assess and evaluate the performance of the EMS system in the State of Qatar
6. To compare the effectiveness of the existing and a proposed EMS system through service area analysis and closest facility performance metrics
7. To develop a multi-facility location allocation model for an EMS system through integrated use of fuzzy k-means, data mining techniques and GIS methods
8. To simulate and compare the effectiveness of a variety of model-driven strategies for developing EMS systems

1.11. Scope

In this research, a mathematical model that addresses both static and transportation facilities will be constructed. The constructed model will be implemented through fuzzy k-means and the solution will be visualized, analyzed and simulated using Arch/GIS. Both tools are combined to take advantage of the strengths of location allocation modeling techniques and modern GIS techniques. The merits and demerits for using single and combined tools will be analyzed through the case study on the emergency medical service system in Qatar. For the case study, data was collected from; the operations of the emergency medical service system in the State of Qatar, Ministry of Development Planning and Statistics, and Ministry of Municipality and Urban Planning.

In Qatar, if there is a need for an ambulance, a person dials (999) to connect to the call center. In response to the 999 call, an ambulance dispatch system is initiated. The efficiency and effectiveness of the dispatch system mainly depends on the location and allocation of EMS units. This research focuses only on finding the best location and allocation of EMS units. Analysis of the dispatch system is outside the scope of the research. Moreover, this study does not put into consideration different types of ambulance vehicles. Thus, it does not differentiate between BLS and ALS vehicles and it does not include the dispatching policies and procedures of these vehicles. In other words, the study will focus on the optimal location of EMS units regardless of vehicle type and capability. The quoted experimental response times assume that issues of receiving 999 calls and dispatch of the EMS units consume minimum possible times without any delays.
The results of this research will contribute to a better understanding of the multi-facility location allocation problems and the merits and demerits of the various strategies, options, procedures and techniques for determining the optimal solutions. In addition, the case study will provide valuable information that can be used to improve emergency medical service systems in Qatar.

1.12. **Outline of This Study**

This research is organized in the following manner: In chapter 2 a review of the literature on EMS systems, location-allocation sciences, geographical information systems and their relation to each other is presented. Chapter 3 describes the methodology adopted and the various methods used to conduct this research. This includes a detailed discussion on the research process, tools used, and data gathering pre-processing and post processing of the collected data. Chapter 4 gives a report of the results and findings of this thesis including data analysis and a discussion of results. Finally, Chapter 5 gives a summary of the findings, recommendations and a brief description of the future work related to this study.
CHAPTER 2

LITERATURE REVIEW

The main purpose of EMS is to minimize the effect of emergency incidents on human lives and health. Besides these objectives, minimization of cost, and maximization of equity, back-up coverage and service level to the uncovered zones are other additional objectives of EMS systems (Araz, et al., 2007). As stated in Disease Control Priorities Project “Emergency health incidents could typically occur through a sudden insult to the body or mind, often through injury, infection, obstetric complications, or chemical imbalance. They may also occur as the result of persistent neglect of chronic conditions” (Jamison, et al., 2006). Therefore, to prevent disabilities and to control morbidity and mortality, which result from health incidents, EMS has to provide rapid and appropriate pre-hospital care, and efficiently transport the patient to the nearest healthcare facility. Furthermore, Kobusingye, et al. (2005) showed that with efficient planning, EMS systems in developing and low income countries can implement better outcomes without increase in costs. Therefore, systematic analysis is one major area of concerns in EMS systems (Setzler, 2007).

For the EMS location allocation problems, intensive studies have been done in order to find the best location of EMS facilities especially ambulance (GU, et al., 2009; Comber, et al., 2011; Knight, et al., 2012). The purpose of these studies was to reduce the response time and/or to improve the survival rate. Comber, et al. (2011) showed that by relocating the ambulances in Niigata city-Japan, one can get better results in terms of distance and response time. They found that the average response time is reduced from 5.35 minutes to 4.12 minutes which is a great improvement since several studies shows that the survival improved when the time is less than 4 to 5 minutes.

In analyzing a vehicle location problem for EMS, Araz, et al. (2007) used different methods in order to meet multiple objectives such as; maximizing the population covered by each vehicle, maximizing the backup coverage, and increasing the service level. It was found that fuzzy goal
programming (FGP) generates better results than other methods. Even though the percentage of first covering was reduced by about 3% than the original setup, it is substituted in the backup coverage and it is found that the backup coverage was increased by more than 100%.

In an effort to locate facilities and vehicles with respect to minimizing the response time, Lightner, et al. (2006) used a MOFLEET mixed integer programming model. It was observed that the coverage area was improved and the response time was reduced. They also found that adding one vehicle to an existing EMS system may increase the coverage by 1.5% which means it will respond to 230 calls more.

The research reported in this document will focus on the location allocation problems in EMS systems. As mentioned earlier on, EMS systems are considered to be complex and large. Therefore, solution approaches typically favor the use of non-classical approaches or a combination of both classical and non-classical approaches in the search for a desired solution. Section 2.2 will address in more details different models used by other researches. In the following section the most critical EMS performance indicators will be discussed.

2.4. **EMS Performance Indicators**

EMS decision makers, administrators are service providers are always seeking to enhance the system’s performance. EMS performance is mainly measured by equity, response time and survival rate. Equity nowadays becomes an important performance measure in many services including EMS systems. Consequently, many studies have addressed such an important factor in their models (Beraldi, et al., 2004; Sorensen & Church, 2010; Chanta, et al., 2011). There is no standard threshold for these indicators. However, a lot of safety/health associations and departments have carried out a number of studies that focus on finding the best thresholds that will ultimately increase health outcomes. It is worthy to mention that in many studies, it was observed that there is a correlation between the survival rate and response time especially for specific medical conditions such as cardiac arrest which has high mortality risk but it accounts only for about 1% of the total emergency cases (Ludwig, 2004; Wilde, 2009; Emergency Medical Services Authority, 2013). Pell, et al. (2001) stated that the potential survival rate could
increase from 6% to 8% if the target improved from 14 minutes to 8 minutes. Moreover, if 90% of calls have been responded within 5 minutes, the survivor rate would increase by 10% to 11%. Vukamir also reached similar conclusions in his study. He found that the survival of cardiac arrest patient is increased when the response time decreases from 6.81 minutes to 5.52 minutes for BLS and from 9.49 minutes to 7.29 minutes for ALS (Vukmir, 2006).

It was found that in cardiac emergency the survival rate is significantly associated with the response time only for critical cases (Jena & Adibabu, 2009). The chart in Figure 3 represents the survival rate versus the time for cardiac arrest patients as represented by Sund (2012).

Despite most of the aforementioned researches, Wilde found that reducing the response time will increase the survival rates not only for cardiac arrest but also for other types of emergency conditions (Wilde, 2009). Based on the cost benefit analysis, Wilde argued that the anticipated benefits of reducing the response time exceed the costs. He also stated that the mortality rate increases by 8% to 17%, for every minute increases in response time. In addition to that,
Blackwell concluded that the mortality is increased by 1.58% for patients whose response time was exceeding 5 minutes and 0.51% for those whose response time is less than 5 minutes (Blackwell & Kaufman, 2002). Recall that the American Heart Association stated that the brain and permanent death occur after 4 to 6 minutes in the case of a cardiac arrest. Therefore, the suitable response time should fall within this range in order to satisfy cardiac arrest patients.

2.5. Location Allocation Models

The purpose of location allocation models is to locate an optimal number of facilities in an area of interest such that it satisfies the customer demand and minimizes the corresponding cost (Azarmand & Jami, 2009). As such, location allocation models are considered as crucial tools for strategic decision making which will help in matching facilities to demand points (Simchi-Levi, et al., 1997).

Facility location modeling takes more interests when used to locate healthcare facilities since the effect of poor location systems will go beyond cost and customer satisfaction (Daskin & Dean, 2005). Thus, it will also increase the mortality and morbidity of customers.

2.5.1. Continuous and Discrete Location Models

Continuous location models occur in a plane which tends to look to a problem from a macro level perspective. The locations are determined by coordinates, e.g. x, y, (Eiselt & Sandblom, 2010). On the other hand, the discrete location models occur almost in the nodes of a transportation network. Thus, discrete location models tend to look to the problem in a micro level. In this respect, the locations are determined by a point (node). Therefore, continuous location models come under the linear or nonlinear optimization, while discrete location models considered as integer programming problems (Eiselt & Sandblom, 2010).

Since the introduction of location allocation models, many models have been proposed (Azarmand & Jami, 2009). Such models can be divided based on their objectives, namely: covering problem, P-median problem, and P-centered problem. These models are the main
models used in location problems in healthcare (Daskin & Dean, 2005). They will be discussed in more details in the following sections.

2.5.2. Covering Models

Covering models are popular and have been extensively used in research because they are applicable in many real-life problems, especially in emergency systems (Eiselt & Sandblom, 2010; Farahani, et al., 2012). The purpose behind this kind of models is to cover the demand points that are located within a predefined distance from a facility (Daskin & Dean, 2005; Fallah, et al., 2009; Eiselt & Sandblom, 2010; Farahani, et al., 2012). This predefined distance is called coverage distance or coverage radius. Consequently, a demand point is generally said to be covered by a facility if the distance between the demand point and the facility is less than or equal to the coverage distance (Daskin & Dean, 2005). As mentioned by Farahani, et al. (2012), the notion of this kind of models is related to a satisfactory method rather than a best possible one since the purpose behind them is to satisfy as much demand points as possible. For example, an increase in the time to respond to an area for fire protection will mean that the fire has greater chance to spread and it may decrease the chance to save life (Eiselt & Sandblom, 2010).

Covering models have many applications such as: designing of switching circuits, data retrieving, assembly line balancing, airline staff scheduling, locating defend networks (at war), distributing products, and warehouse locating (Fallah, et al., 2009). Application of covering models is also found in emergency facilities such as fire stations, ambulances, police cruisers, or any other facilities (Eiselt & Sandblom, 2010). There are mainly two objectives from covering models which are either to cover all demand points with minimum number of facilities or to cover as many demand points as possible with a predefined set of facilities (Eiselt & Sandblom, 2010). These kinds of models are widely used in set covering problems (SCP) and maximal covering location problems (MCLP), respectively.

2.5.2.1. Set Covering Problem (SCP)
The set covering problem is one of the first models which was developed in early 1970s to locate emergency facilities (Eiselt & Sandblom, 2010). As mentioned earlier, the objective of set covering problem is to minimize the cost of facilities by locating the smallest number of required facilities to cover all demand points (Daskin & Dean, 2005; Eiselt & Sandblom, 2010; Farahani, et al., 2012). The demand could be covered by one or more facilities if its distance does not exceed the predefined coverage distance (Eiselt & Sandblom, 2010). There are several variant forms of set covering problems, including; Location SCP, capacitated SCP, quadratics SCP, multiple optimal SCP, covering tour problem, path covering problem, probabilistic SCP, Stochastic SCP, Fuzzy SCP, multiple coverage SCP, backup coverage SCP, and multi-criteria SCP (Farahani, et al., 2012).

2.5.2.2. Maximal covering location problem (MCLP)

In real world, using the set covering model is associated with a number of issues. For example, covering all demand points is usually not practical due to the budget and resource constraints (Fallah, et al., 2009; Farahani, et al., 2012). Moreover, SCP considers all demand points to be the same regardless of their demand volume (Daskin & Dean, 2005; Fallah, et al., 2009). Consequently, the MCLP was developed in order to take into consideration the issues discussed above (Farahani, et al., 2012). This model tries to cover as much demand points as possible by sufficiently locating a given number of facilities. The demand points have to be within the predefined distance. There are many types of variants of MCLP such as: MCLP implicit and explicit, planar maximal covering, capacitated MCLP, MCLP with a criticality index analysis metric, MCLP with mandatory closeness constraints, probabilistic MCLP, MALP I, MALP II, Partial coverage problem, and gradual coverage.

2.5.3. P-center Models

In certain circumstances, we may be restricted to work within a certain budget and at the same time we may be restricted to locate P numbers of facilities. One solution is to relax the total coverage requirement. Another strategy to solve this problem is by using the P-center concept which focuses on minimizing the maximum distance between the demand point and the facility.
P-center model is different from the covering problem in sense that in this model instead of minimizing the number of facilities within a given coverage distance, we minimize the coverage distance using a predefined set of facilities such that all demand points are covered (Biazaran & SeyediNezhad, 2009). Therefore, unlike the coverage problem, the coverage distance requirement in P-center problem is relaxed. Thus, the objective of this model is to locate P facilities in such a way that all demand points are covered while minimizing the maximum distance between a demand point and the nearest facility. Accordingly, the focus here is to make the worst case as good as possible (Biazaran & SeyediNezhad, 2009). This concept was inspired from Rawls’s theory of justice in which the quality of a solution depends on the least well-served entity (Eiselt & Sandblom, 2010). It is worthy to mention that each demand point could have a weight which could represent the time, cost, or any other weight’s figure per unit distance (Daskin, 1995). As mentioned by Eiselt & Sandblom (2010) one of the limitations of P-center model is its exclusive focus on the farthest demand point which could lead to undesirable outcomes.

The P-center model can be applied in hospital emergency services, fire stations, warehouses, parks, hotels and location of post boxes and bus stops (Biazaran & SeyediNezhad, 2009). It is worthy to mention that the facilities in discrete models could be established either in any place in the network or could be placed only on nodes. These problems are called absolute and vertex models respectively.

2.2.5. P-median Models

P-median models are used to find the location of P facilities in order to minimize the total coverage distance subject to a requirement that all demand points have to be covered (Daskin & Dean, 2005). Thus, the objective is to minimize the average distance between demand points and facilities, so that the sum of costs can be minimized through this target (Jamshidi, 2009). In this way, the fixed cost for a facility disappears from the objective function. Unlike coverage problem and P-center problem, which treat service as binary (a demand point is either covered or not covered), the P-median problem focuses on minimizing the average/total weighted distance
that a provider or a customer has to travel to get a service. It is suited mostly to the public sector such as: schools, hospitals, firefighting, ambulance, technical audit stations of cars, and etc. As mentioned previously, it focuses on the average distance rather than the cost of candidate locations (Chan, 2005).

2.6. Location Allocation Models Related to EMS

Three main location models used in healthcare are: set covering model, maximal covering model, and P-median model (Daskin and Dean, 2005; Shuib & Zaharudin, 2010). These models can be classified into deterministic or probabilistic models. Deterministic models ignore stochastic conditions regarding the availability of healthcare facility such as ambulances. On the other hand, probabilistic models reflect the fact that ambulances could be busy at a point of time and cannot always answer a call (Brotcorne, et al., 2003).

It is worthy to mention that covering models (including both the SCP and MCLP) are the most widespread location models in the emergency location-allocation problem (Jia, 2006; Farahani, et al., 2012). One of the earliest models used was location set covering model by Toregas in 1971, where the location of ambulances that covers all demand points with respect to a pre-specified distance was determined (Toregas, et al., 1971). Later on, the MCLP was proposed in order to maximize the demand covered with respect to one condition, which is minimizing the number of ambulance allocated. MCLP was used by Eaton, et al. (1985) in order to reorganize the EMS system in Austin, Texas. It was found that this new model saved millions of dollars in construction and operating costs.

Sadigh, et al. (2010) used contemporary edge covering model, an extension of covering model, where the partial covering of an edge through vertices is permitted. Since this problem is considered as an NP-Hard problem, tabu search was used to find a solution. The authors claimed that this type of model can be used in locating emergency facilities including EMS facilities. It was found that the proposed mathematical model could solve up to 40 vertexes with 456 edges optimally and find high quality solutions for situations with larger scale. For EMS systems, it is worthy to mention that if there are multiple types of vehicles to be dispatched to a scene, both
SCP and MCLP models cannot recognize them (Brotcorne, et al., 2003). Therefore, a number of models were proposed to address issues related to the types and availability of ambulance vehicles. For EMS systems, the coverage models are considered deterministic since they do not consider the busyness of an ambulance vehicle (Brotcorne, et al., 2003; Shuib & Zaharudin, 2010). Accordingly, new models which put into consideration the availability of ambulances have been developed by Daskin & Stern (1981) and Hogan & Revelle (1986). Fujiwara, et al. (1987) used the model developed by Daskin and Stern and concluded that reducing the number of ambulances from 21 to 15 will have similar expected covering and response time. Araz, et al. (2007) used multiple solution approaches such as lexicographic multi-objective linear programming and fuzzy goal programing (FGP) to solve backup coverage objective and service level in addition to the first covering objective in order to locate EMS vehicles. It is found that FGP model is an effective tool than others for generating a set of more realistic and flexible optimal solution.

Another model for consideration is the ambulance allocation capacity model (A ACM) which is a combine of location set covering model and probability measures. The main reason behind AACM was to provide equity to all residence (Knight, et al., 2012). Convex path, convex hull, and potential sites algorithms were used to solve this problem. The results show that by adding one ambulance, the coverage increased to 90% instead of 49.3%.

All of the above studies assumed that there is fixed capacity for the facility. Yin & Mu (2012) claimed that the capacity of a facility varies according to the number of vehicles available on it. For example, the capacity of an ambulance station can be changed based on the number of ambulance vehicles available in the station. Therefore, Yin & Mu (2012) proposed a new MCLP model; namely modular capacity maximal covering location problem, which takes the capacity into consideration. Besides the modular capacity, it considers the allocation of all demand points, and the proximity of the uncovered demand points. The model input files were constructed in visual basic for application (VBA) program of ArcObjects in ArcGIS. Where then they have been solved using a branch-and-cut technique in CPLEX to find the optimal solution.
Few other researches have used P-median models. For instance, GU, et al. (2009) proposed a new heuristic method called Static Transportation Facilities Location Searching Algorithms (STFLS) to solve both static and transportation facility location problem. The purpose of their research was to minimize the average weighted traveling distance. The solution using STFLS was compared with the solution generated from a GIS software. It was found that the solution using STFLS has better performance in average and maximum traveling distance but it is time consuming.

In EMS systems, very large amount of data and information are received on a daily bases. This data must be analyzed in order to assess the effectiveness and efficiency of a given EMS system. Moreover, EMS decisions, more specifically location-allocation decisions, involves many complex factors such as distribution of population and their characteristics, types of facilities and their capacities and resources, and road network and buildings’ distributions. This means that EMS systems are data intensive systems since different types of information and data are used to make informed decisions regarding the state of an EMS system. As such, it is important to extract valuable information for decision making. Data mining and GIS are two powerful contemporary tools that have been successfully used to handle data intensive systems. The capabilities of GIS and Data Mining techniques will be used in this research for handling large amounts of data required to assess the performance of an EMS system.

2.5. Knowledge Discovery from Data

As discussed earlier on, EMS systems of today are data intensive, i.e. data from various sources is available for use in EMS systems. Since data is from many sources, the data is complex and may not be useful for decision making and planning purposes. Therefore, it is important for decision makers to be provided with the necessary information extracted from available sources for the purpose of planning EMS systems. In this regard data mining techniques have been used to extract information from data (Witten & Frank, 2005). (Han & Pei, 2011). Figure 4, shows the necessary steps in knowledge discovery from data. Data mining can be used in four patterns, namely; characterization and discrimination, association and correlation, classification and regression, and clustering.
In EMS systems, there are no clear class labels in which instances as well as people could be grouped in. Moreover, the main focus for EMS systems is to locate each facility in such a way to serve properly a group of people who are living at a particular area. Thus, each EMS facility has to satisfy as much as possible the area that it is assigned to with the ignorance of other areas. As such it is deemed necessary in this research to investigate the impact and the extent to which clustering techniques can be used in the planning of an EMS system.

2.6.1. Clustering

Clustering can be defined as grouping set of data points into multiple groups, called clusters, where data points inside the cluster are very similar to each other and dissimilar to others that are belonging to other clusters. Unlike other data mining patterns, the grouping or classification criteria are unknown and need to be discovered by the user. The relation between points is determined using an attribute value which describes the point and it is usually represented as distance (Han & Pei, 2011). Clusters can be fully separated from each other, they can overlap or be continuously connected to each other (Abonyi & Feil, 2000). Applications using cluster
analysis includes; business intelligence, image pattern recognition, Web search, biology, and security (Han & Pei, 2011). It is worthy to mention that unlike other data mining methods, clustering is learning by observation not by example (Han & Pei, 2011). Thus, it does not need to specify class label. Therefore, it is known as unsupervised learning. Moreover, clustering can be used for outlier detection.

There are mainly four different clustering methods; partitioning methods, hierarchical methods, density-based method and grid-space methods (Han & Pei, 2011). Using the same dataset, different clustering methods leads to different clusters and results. This is because clustering methods are varied in the level of partitioning; is it hierarchical or not, level of clusters’ separation; whether they are mutually exclusive or not, the similarity measures used; is it based on distance or density or other measures, and the clustering space used; whether the entire data space used or it is sub-spaced. Table 1 summarizes the difference between the four clustering methods (Han & Pei, 2011).

Table 1: General Clustering Analysis Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partitioning</td>
<td>Find mutually exclusive clusters of spherical shape</td>
</tr>
<tr>
<td>Methods</td>
<td>Distance-based</td>
</tr>
<tr>
<td></td>
<td>Effective for small- to medium-size data sets</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>Clustering is represented in multiple levels</td>
</tr>
<tr>
<td>Methods</td>
<td>Cannot correct erroneous merges or splits</td>
</tr>
<tr>
<td></td>
<td>May incorporate other techniques like micro clustering</td>
</tr>
<tr>
<td>Density-based</td>
<td>Find arbitrary shaped clusters</td>
</tr>
<tr>
<td>Methods</td>
<td>Clusters are dense regions of objects in space that are separated by low-density regions</td>
</tr>
<tr>
<td></td>
<td>May filter out outliers.</td>
</tr>
<tr>
<td>Grid-based</td>
<td>Use multi-resolution grid data structure</td>
</tr>
<tr>
<td>Methods</td>
<td>Fast processing time</td>
</tr>
</tbody>
</table>
The distribution of EMS facilities heavily depends on the distance between the facility and incidents. Moreover, there is no specific classification for people in different zones to help in locating EMS facilities. Therefore, the best method suitable for the current research requirements is partitioning. The following section will discuss in details partitioning method; more specifically k-means clustering.

2.6.2. Partitioning

Partitioning is clustering, in which a set of objects (points) are partitioned into exclusive groups called clusters. The objective of partitioning is to maximize the intra-cluster similarities and minimize the inter-cluster similarities (Han & Pei, 2011). It is worth mentioning that achieving global optimum in partitioning clustering requires extremely high computational cost and exhaustive enumeration of all the possible partitions. Therefore, heuristic methods, such as k-means algorithms, are used to achieve an optimum by gradually improving clustering quality. The algorithms work well in small- to medium-size problems. In this research, only k-means clustering method will be explained in details in later sections.

k-means clustering results in assigning each object to only one cluster. This rigid cluster assignment rule does not work with some real life applications, where an object could be related to more than one cluster. Therefore, a more advanced k-means clustering method has been demonstrated to deal with such situations by assigning a probability of belonging (Abonyi & Feil, 2000; Han & Pei, 2011). This method is called “fuzzy clustering”.

Fuzzy clustering is also called soft clustering since it is flexible and allows an object to belong to more than one cluster at the same time (Abonyi & Feil, 2000; Esnaf & Kucukdeniz, 2009). Esnaf and Kucukdeniz (2009) used a fuzzy clustering-based method to assign customers to plants. It consisted of two sequential phases: fuzzy clustering of multi-facility location problem and then they used center of gravity method to find the optimal location for each facility at each cluster. It was a pioneer paper which combined both the clustering algorithms and center of gravity methods. It is found that this hybrid method helped in decreasing the transportation cost.
Kaundinya, et al. (2013) used a GIS-based k-medoid data mining algorithm to select suitable locations to install biomass power plants in rural regions in Tumkur, India. This approach helped in finding the optimal solution by minimizing the cost of power system installation, biomass transportation and power transmission. A hybrid spatial clustering method for the selection of customer service location has been used by Fan (Fan, 2009). The proposed method considered huge number of factors which cannot be handled by conventional mathematical models. Such factors were: the spatial restrictions, constraints, traffic, terrain, cost, and environmental factors.

2.7. Geographical Information Systems (GIS)

During the past thirty years, GIS has become a popular application which is used in many diverse areas such as businesses, universities, and governments (Environmental systems Research Institute. Inc., 1990). The reason of using such system is that GIS is capable of analyzing both spatial and non-spatial data (Cheng, et al., 2007; Zhang, et al., 2011). It is also a powerful and efficient instrument to (Harvey, 2008): i) make location decisions specially for dynamic and complicated conditions (Vafaeinezhad, et al., 2009), ii) find the best location, iii) find the best way to get to that location, and iv) optimize the use of the available resources (Albrecht, 2007).

GIS can be defined as “an organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information to support geographical decision making” (Environmental systems Research Institute. Inc., 1990; Murray, 2010). Geographic information and maps are representations that follow a number of principles and conventions that help deal with the complexity of the world and guide choices that improve communication (Harvey, 2008). The skeleton of a GIS is the use of a location referencing system where the data of a specific location could be represented and analyzed with respect to other locations (Church, 2002). This powerful tool is used particularly in many research facilities and offices in order to analyze and manage resources (Harvey, 2008). It is worthy to mention that location science have been advanced by the use of developed mathematical models that combine the spatial problem and by the usage of innovative optimization techniques (Murray, 2010). Also, general purpose
optimization software has been used to support advancement in model development and application. Moreover, heuristic solutions techniques have been emerged which leads to solve more complex systems in a reasonable processing time. However, all of the above methods are not able to handle spatial data (Zhang, et al., 2011), and moreover, nowadays more real world interactions are integrated into the mathematical optimization problems which increases its complexity (Murray, 2010). Therefore, there is a need to use more advanced tools such as GIS to accurately solve such complex problems.

As stated by Murray (2010) “much of the model advancement in location science can be directly or indirectly linked with the maturation of GIS”. In his paper, Murray (2010) discussed how GIS develops the analysis and modeling of location science. He detailed that GIS is used in location allocation sciences mainly in three functions: i) model input; where GIS has supported location sciences by extracting location coordinates and its attributes to be in location models, ii) visualization; where a better understanding of the model, its objective, geographic spaces, and even the model result is reached and thus determine whether the solution makes sense or not (Murray, 2005), and (iii) GIS helps problem solution; where a location problem could be solved through the usage of GIS alone or through the combination of other modeling approaches.

As mentioned previously, due to the increasing complexity of real-life systems and accordingly the location modeling, there is a need to use advanced tools and methods which are capable of solving such systems. Therefore, the usage of GIS has been growing in recent years since it has been proved that combining GIS with other methods is very effective in decision support systems for selecting locations (Zhang, et al., 2011). For instance, Vafaeinezhad, et al. (2009) modeled task allocation to persons in a totally dynamic and complicated environment where they used GIS to simulate data, and to generate and evaluate the results of the tasks for two groups of life-detectors and rubble-removers of earthquake rescue teams. Thus, GIS used here as both to generate inputs and to find problem solution.

GIS has also been used by Rodrigues, et al. (2012) as source of input data and to present model results. Similarly, Bender, et al. (2002) and Murawski & Church (2009) used GIS to export
attribute and coordinate information to be used in location modeling problems. It is worthy to mention that the allocation of task manpower in GIS was not implemented i.e. the dynamics of different elements such as tasks, environment and unpredicted events were not taken into account. Therefore, the research of Vafaeinezhad, et al. (2009) is worthwhile since it addresses such issues. In Bozakaya, et al (2010), a Tabu search heuristic algorithm was coupled with GIS software to solve a vehicle routing problem. It was also used to store, analyze and visualize the location routing solutions. On the other hand, Murray (2005) formulated a set covering problem (SCP) using GIS to reduce inconvenient spatial effects when it is represented using only the SCP model. Based on specific criteria and assumptions, Zhang, et al. (2011) used GIS in order to minimize the set of candidate locations for pulpwood-to-biofuel conversion facilities. Similarly, Vega, et al. (2011) used GIS to improve the traditional solution methods by discarding any location which is less than the specified thresholds or which does not meet specific criteria. Cheng, et al. (2007) took the advantage of GIS to find the best location for a super shopping mall in order to minimize distance, maximize demand coverage, and to maximize average monthly income coverage. It was found that GIS easily supported all of the aforementioned enquiries.

In a competitive sitting, Vega, et al (2012) used GIS to find the best location for a retailer, where mainly two tools were used. The first tool, called Market Share Calculation (MSC), calculates the projected market share of each new facility for each point within the feasible region. The second tool, called Weighted Market Share (WMS), was used to identify the optima for the objectives and then evaluate the trade-off between all of them to find a final solution to the problem.

GIS has also been used in location problems related to health and emergency sectors. For instance, GIS was used to assess the efficiency of fire station systems (Lui, et al., 2006). Moreover, Ant-colony algorithm was applied within using the GIS to locate new fire stations in order to decrease the response time to be within 5 minutes instead of 8 minutes. Sasaki, et al. (2010) implemented GIS to count the emergency cases in each census area under study and to measure distance between the ambulance units and those census areas. The purpose of Sasaki’s study was to find the optimal location of ambulances by expecting future EMS cases and to
create future EMS management strategies by expecting the demand. To improve the EMS services, GIS was implemented by Estochen, et al. (1998) to identify existing EMS service areas and compare these areas with traffic crash densities. Moreover, the response time were estimated and compared to actual response time. Based on the results generated, improvements were suggested.

2.8. Summary of Gaps in the Literature Review

The literature review has shown that there are many models, methods, tools and techniques for designing and developing EMS systems. It has also been shown that different approaches will result in different designs of EMS systems. While most researches prefer to focus on one method, this thesis will focus on the integrated use of two methods, i.e. data mining and GIS. In particular applications of data mining are relatively a new area of study in multi-facility location allocation problems. As such, this thesis will put an effort in investigating how some data mining techniques can be utilized in the multi-facility location allocation problem in EMS systems. More specifically, the effectiveness of the fuzzy k-means algorithm will be tested and evaluated in this thesis.

Although the idea of combining a number of methods is not entirely new, this approach is deemed worthy to pursue in this thesis in order to assess the merits and demerits of combining methods in seeking better solutions. Thus both location-allocation and GIS methods will be combined with the hope of finding a better solution within an acceptable execution time. For example, Comber, et al. (2011) used a modified grouping genetic algorithm in order to optimally select best ambulance locations to minimize the population-weighted average distance. The results showed an improvement by 1.23 minutes on average in the EMS response time.

Table 2 summarizes the various location allocation models used in the public literature. Literature review has also shown that with respect to the p-canter and the p-median, the set covering problems are more widely used in EMS systems. However, Carson & Batta (1990) used the P-median in order to find the dynamic ambulance location strategy to serve emergency cases in Campus of the State University of New York at Buffalo. The ambulances were relocated in
different scenarios in order to minimize the average response time to the service calls. As mentioned in Jia (2006) priority dispatching using P-median model was also used by Mandell in 1998 to optimally locate two types of ambulances ALS and BLS. GIS is a contemporary tool that is now widely used for location allocation planning. In this thesis GIS will be used in conjunction with fuzzy k-means to develop an EMS system model. GIS will also be used for analyzing the performance of various EMS models that will be created in the various investigations of this thesis. This thesis derives inspiration from the fact that if the p-median problem is configured without an impedance cut-off, the resulting location allocation model will provide more equitability that most models (ArcGIS version 10). Since the fuzzy k-means model can be considered as an advanced p-median model, its implementation and the results thereof will contribute significantly to the importance of the p-median model in the location allocation of EMS systems.

The literature review has shown that very little research has been done to study both static and transportation facilities. However, GU, et al. (2009) used both the optimal method and GIS to solve such problems. In their study, GU et al. (2009) found that the optimal method gave better solution in comparison to GIS. However, the GIS method was faster in reaching the solution when compared to the optimal method. Based on GU, et al. (2009) and since the area of static and transportation facility still new, this thesis attempts to address both static and transportation facilities in order to improve the service level of EMS in Qatar. In other words, this study aims at improving the survival rates by reducing the response time. This objective will be accomplished by combining GIS and data mining techniques, in search for a more effective solution to EMS system in Qatar.
### Table 2: Summary of Research Papers

<table>
<thead>
<tr>
<th></th>
<th>Name of the paper</th>
<th>Name of the author</th>
<th>Model</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Allocation of Emergency Ambulances to Fire Stations</td>
<td>James A. Fitzsimmons</td>
<td>p-median</td>
<td>1973</td>
</tr>
<tr>
<td>4</td>
<td>Application of an expected covering model to emergency medical service system design</td>
<td>Mark S. Daskin</td>
<td>Covering problem</td>
<td>1982</td>
</tr>
<tr>
<td>5</td>
<td>Determining Emergency Medical Service Vehicle Deployment in Austin, Texas</td>
<td>David J. Eaton, Mark S. Daskin, Dennis Simmons, Bill Bulloch and Glen Jansma</td>
<td>Covering problem</td>
<td>1985</td>
</tr>
<tr>
<td>6</td>
<td>A Hybrid Fleet Model For Emergency Medical Service System Design</td>
<td>Geoffrey Bianchi, and Richard L. Church</td>
<td>Covering problem</td>
<td>1988</td>
</tr>
<tr>
<td>7</td>
<td>Locating an ambulance on the Amherst campus of the State University of New York</td>
<td>Yolanda M Carson; Rajan Batta</td>
<td>P-median</td>
<td>1990</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implementing a mathematical model for locating EMS vehicles in Fayetteville, NC</td>
<td>Asad Tavakoli, Constance Lightner</td>
<td>Covering problem (FLEET)</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Designing robust emergency medical service via stochastic programming</td>
<td>P. Beraldi, M.E. Bruni, D. Conforti</td>
<td>P-median</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>Developing A Mathematical Model For Locating Facilities And Vehicles To Minimize Response Time</td>
<td>Constance Lightner, Asad Tavakoli, Yahya Fathi.</td>
<td>Covering problem (FLEET)</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Extensions to emergency vehicle location models</td>
<td>Othman Ibraheem Alsallouma, Graham K. Rand,</td>
<td>Covering problem</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>A fuzzy multi-objective covering-based vehicle location model for emergency services</td>
<td>Ceyhun Araz, Hasan Selim, Irem Ozkarahan</td>
<td>Covering problem (MCLP)</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>Heuristic Solutions for Locating Health Resources.</td>
<td>J. Pacheco; S. Casado; J.F Alegre</td>
<td>Covering problem</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>A multi-period set covering location model for dynamic redeployment of ambulances</td>
<td>Hari K. Rajagopalan, Cem Saydam, Jing Xiao</td>
<td>Covering problem</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>GIS-FLSolution: A Spatial Analysis Platform for Static and Transportation Facility Location Allocation Problem</td>
<td>Wei Gu, Xin Wang, and Liqiang Geng</td>
<td>mix of P-median and P-center</td>
<td>2009</td>
</tr>
<tr>
<td>17</td>
<td>A mixed integer linear program and tabu search approach for the complementary edge covering problem</td>
<td>Ali Naimi Sadigh, Marzieh Mozafari, Ali Husseinzadeh Kashan</td>
<td>Covering problem</td>
<td>2010</td>
</tr>
<tr>
<td>18</td>
<td>A modified grouping genetic algorithm to select ambulance site locations</td>
<td>Alexis J Comber, Satoshi Sasaki, Hiroshi Suzuki, Chris Brunsdon</td>
<td>P-median</td>
<td>2010</td>
</tr>
<tr>
<td>19</td>
<td>Integrating expected coverage and local reliability for emergency medical services location problems</td>
<td>Paul Sorensen, Richard Church</td>
<td>Covering problem</td>
<td>2010</td>
</tr>
<tr>
<td>20</td>
<td>A modified grouping genetic algorithm to select ambulance site locations</td>
<td>Alexis J. Combera, Satoshi Sasakib, Hiroshi Suzuki and Chris Brunsdona</td>
<td>p-median</td>
<td>2011</td>
</tr>
<tr>
<td>21</td>
<td>Improving emergency service in rural areas: a bi-objective covering location model for EMS systems</td>
<td>Sunarin Chanta, Maria E. Mayorga, Laura A. McLay</td>
<td>Covering problem</td>
<td>2011</td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>Authors</td>
<td>Method</td>
<td>Year</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>22</td>
<td>A bi-objective model for emergency services location-allocation problem with maximum distance constraint</td>
<td>Mansoureh Haj Mohammad Hosseinia, Mohammad Saeed Jabal Ameli</td>
<td>Covering, p-center, p-median</td>
<td>2011</td>
</tr>
<tr>
<td>23</td>
<td>A multi-period double coverage approach for locating the emergency medical service stations in Istanbul</td>
<td>Ayfer Basar; Bulemt Çatay; Tonguc Ünlüyurt</td>
<td>Covering problem</td>
<td>2011</td>
</tr>
<tr>
<td>24</td>
<td>Ambulance allocation for maximal survival with heterogeneous outcome measures</td>
<td>Vincent Knight; Paul Harper; Leanne Smith</td>
<td>Covering problem (Set covering )</td>
<td>2012</td>
</tr>
<tr>
<td>25</td>
<td>Modular capacitated maximal covering location problems for the optimal siting of emergency vehicles</td>
<td>Ping Yin; Lan Mu</td>
<td>Covering problem (MCLP)</td>
<td>2012</td>
</tr>
<tr>
<td>26</td>
<td>Joint location and dispatching decisions for Emergency Medical Services</td>
<td>Hector Toro-Diaz, Maria E. Mayorga, Sunarin Chanta, Laura A. McLay</td>
<td>Covering problem</td>
<td>2013</td>
</tr>
</tbody>
</table>
CHAPTER 3

RESEARCH METHODOLOGY

Functions of an emergency medical service (EMS) system include; (a) to preserve life, (b) to prevent further injury, and (c) to promote recovery. While these functions can be achieved through a variety of methods, it is important to ensure optimality in providing these functions. Although operational issues such as EMS units dispatch are critical to the success of an EMS system they can be enhanced and/or enabled if the EMS system is strategically planned for optimality. Such strategic planning is usually based on sound theoretical and technical principles that are formulated based on achieving desired targets in operations.

In the process of designing and developing an optimized location-allocation plan for an EMS system this research employs a mix of qualitative and quantitative methods. The qualitative method was used to investigate how EMS systems (e.g. the EMS in the State of Qatar) are designed and why EMS systems are operated in a particular manner. This method was envisaged to help in understating; (a) the dynamics of EMS systems, and (b) the various generic problems in EMS systems. Such an inquiry-based analysis can be used to derive relevant information required for developing innovative approaches and unique analytical methods that can be used to develop new, improved and futuristic EMS systems. The quantitative method was then used to seek empirical support for the proposed approach to modeling EMS systems. In this thesis, the EMS system in Qatar was used as a case study.

Data collection for the quantitative approach was carried out by; (a) studying and analyzing available data sets, and (b) conducting structured and semi-structured interviews and meetings with a number of stakeholders affiliated to the EMS system in the State of Qatar. These interviews were conducted through phone calls and/or planned meetings. In cases of an in-depth understanding of complex ideas, techniques and
procedures, face-to-face meetings with key stakeholders were preferred. Interviews were conducted with key personnel from the following organizations; (a) Ambulance Services in Qatar, (b) Qatar 2022 Supreme Committee, (c) Ministry of Development Planning and Statistics, and (d) GIS Center under the Ministry of Municipality and Urban Planning. Collected information and data were used in the empirical analysis of EMS systems. Figure 5 summarizes the key stakeholders consulted in this research. A number of tools were used to analyze and process data into useful information from which findings were determined and conclusions made.

![Figure 5: Main Stakeholders of the EMS System in State of Qatar](image)

3.5. **Overview of the Research Process**

Like any other engineering research problem, the approach taken in this thesis was based on the engineering design method. Available solution methods for similar and related work were extracted from the public literature, studied and their applicability investigated. New solutions and ideas were proposed and implemented and the effectiveness of these solutions was explored. Results were then analyzed and
accordingly conclusions were drawn. The overall research process is presented in the flowchart in Figure 6.

![Figure 6: Research Process](image)

3.6. Tools Used

Since the EMS system in Qatar was used as a case study, the core data required for this research is related to the population demographics, incidents, emergency medical services locations, as well as the road networks in the state of Qatar. Relevant mathematical models, tools, and software were identified. Two contemporary methods (i.e. Data mining and GIS location science) were used to optimally locate and relocate EMS
facilities. Data mining and GIS location sciences were used concurrently to implement optimal location/re-location of EMS units. Closest facility analysis, service area analysis, and normality tests were used to statistically analyze the results of this study. Recommendations were suggested based on the analysis of the obtained results.

A number of tools were used in this research. The main tools used were; ArcGIS software, Matlab software, Excel and Minitab software. Minitab and excel softwares were used for statistical analysis of the various data sets used in the investigation.

Matlab is a software that is capable of analyzing data, developing algorithms, plot functions and data, as well as create models and applications (The MathWorks, Inc., 2013). In this research, Matlab was used to develop and implement a fuzzy clustering algorithm. In this respect, Matlab was used to cluster the large amount of population/incidents demographics data collected from databases provided by stakeholders. Obtained data was then processed into useful information which was used in the design and development of a more effective EMS system.

Arc/GIS is a comprehensive system that allows users to collect, organize, manage, analyze, communicate, and distribute geographic information (Economic and Social Research Institute, 1995). Arc/GIS was used to; (a) visualize the existing EMS system, (b) visualize proposed EMS models developed through a variety of procedures and techniques, (c) conduct a network analysis of the existing and proposed EMS systems, and (d) simulate the performance of proposed EMS models.

3.7. Case Study
3.7.1. Study Area

The study area is the State of Qatar, a peninsula in the gulf region. The state of Qatar lies halfway along the west coast of the Arabian Gulf. Its total area is 11606.8 km² with a width of 85km and length of 185km (Ministry of Development Planning and Statistics, 2010). The population of State of Qatar is estimated to be 1,551,821 according to the
2010 census. More than 90% of them live in Doha and its suburbs. Local planners of emergency medical services (EMS) in Qatar are often faced with the need to assess and re-assess the performance of the current EMS system in order to avoid sub-optimal operations. This is because the performance of the EMS system is affected by unexpected population growth, rapid expansion of cities and industrial areas as well as random relocations of residential, commercial and industrial facilities. Figure 7 shows the spatial positions of the population and road network in the State of Qatar, while Figure 8 shows the distribution of EMS incidents and the road network.

Figure 7: Location and Distribution of Population and Road Network in Qatar
Figure 8: Distribution of emergency incidents in Qatar for the year 2013/2014
3.7.2. Case Study EMS

The case study EMS was designed in GIS and based on the gravity-based two-step floating catchment area method (Luo & Qi, 2009; McGrail & Humphreys, 2009; Ngui & Apparicio, 2011). This EMS operates on two modes, namely the winter model and the summer model. The EMS units are divided into hubs (ambulance stations) and spokes (ambulance vehicles). Figure 9 and Figure 10 show the positions of the EMS units and the hub catchment areas for the summer and winter models respectively for 2013. The development of the current location-allocation model for the EMS system in Qatar was based on historical data regarding the incident calls.

Figure 9: Summer Response Locations – 2013
In the winter EMS system, the ambulance units are spread more around the hubs than the summer. Thus, the system in winter is regarded (by ambulance service providers) to be more ideal than that in summer. The reason behind having two different modes is that most of the people travel to other countries for the summer holiday since most companies have either a shutdown or temporary closure during this period. Consequently, the number of incidents is expected to be less. There are seven hubs in summer while there are 6 in winter. These hubs cover the whole country as shown in Figure 9 and Figure 10. The colored polygons (hub catchment areas) in Figure 9 and Figure 10 represent the coverage of each hub while the circles represent the spokes locations.

Figure 10: Winter Response Locations – 2013
The EMS system in Qatar consist of ambulance hubs and ambulance spokes (i.e. vehicles), all of which will be referred to as EMS units. For the existing system the total number of hubs is seven and the total number of spokes is 49. It is important to mention that hubs are more advanced locations than spokes in sense that they have more facilities for staffs such as offices for supervisors, and spare equipment and ambulances. Three main ambulance vehicles are; Alpha, Delta, and Charlie as shown in Figure 11.

![Figure 11: EMS Facilities and Staffs in State of Qatar](image)

The Alpha unit has the highest number and is used almost in all emergency incidents. It is used to provide prehospital treatment and care for any emergency case and it is usually operated by two ambulance paramedics. The Delta unit is a non-clinical supervisory unit used to facilitate the process of EMS by managing resources and communicating with different agencies involved in various emergency cases. It consists of only one operational supervisor or an ambulance paramedic. The Charlie unit is used in critical
cases and when the Alpha units need advanced assistance. There are few Charlie units and these units consist of critical care paramedic and ambulance paramedic. Recently, the life flight unit has been used for incidents outside Doha city and for any critical case that need fast response. It consists of one critical care paramedic and one ambulance paramedic. Table 3 summarizes the resources available for each EMS type.

Table 3: Resources Summary of EMS System at State of Qatar

<table>
<thead>
<tr>
<th>Unit Type</th>
<th>Maximum Number Available</th>
<th>Man power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ambulance paramedic</td>
</tr>
<tr>
<td>Alpha</td>
<td>62</td>
<td>✔️</td>
</tr>
<tr>
<td>Delta</td>
<td>10</td>
<td>✔️</td>
</tr>
<tr>
<td>Charlie</td>
<td>9</td>
<td>✔️</td>
</tr>
<tr>
<td>Life flight</td>
<td>1</td>
<td>✔️</td>
</tr>
</tbody>
</table>

The current EMS prioritizes the response into three main categories; (a) “life threat” scenarios, which are characterized by light and sirens driving and has the highest priority; (b) “urgent” scenarios which are characterized by normal driving and has priority of 2, and; (c) “Transport” which means the response for scheduled patient or inter-facility transfer and has the lowest priority.

Besides the response classification, there is also the patient classification. They classify the patient into: (i) Trauma: if injury caused by external force, (ii) Medical: which refers to any kind of illnesses such as diabetic collapse, and (iii) Transport: which refers to scheduled patients or inter-facility transfer. Most emergency calls fall under life threatening and urgent cases.

One of the mostly used performance measures is the response time, i.e., the time spent from dispatching and EMS unit until it reaches the scene of the incident. The targets in the current EMS system in Qatar are as follows: (a) in urban areas the target is to serve
75% of the cases within 10 minutes, and 95% within 15 minutes, (b) in rural areas the
target is to serve 75% within 15 minutes and 95% within 20 minutes. It is obvious from
the aforementioned targets that there is a difference in response times between rural and
urban area. Thus, there is preference for the urban areas over the rural areas i.e. non-
equitability). However, human life is important regardless of location. Consequently, this
research investigation will consider equitability as an important measure of performance
of the EMS system in Qatar regardless of the location in Qatar.

3.7.3. Data Collection

This section provides a discussion about the data available for this research and how it
was used. Collected data included; population data, the number and type of incidents, as
well as details of the road-network in the State of Qatar. Details of collected data are
provided in appendices A and B.

Population data included all population figures in each district in Qatar. The available
data was divided into; (a) male and female, and (b) groups based on age such as; 0 to 4, 5
to 9, 10 to 14, 15 to 19, 20 to 59, and 60+.

Call incidents data was provided by the ambulance services. This data was used to
identify patterns about the emergency cases with respect to the time over the day and the
month of year. Available incidents data was for a period of one year starting from
October, 2012 till September, 2013. This data included many details such as the date and
time of the incident, the zone number, location name, case classification, response
priority, call time, dispatch time, response time, and time to reach to hospital, and even
the destination or in other words the name of the hospital. Road network data is important
in this research since it helps in calculating the distance and, accordingly, the time
required to reach to an incident. The available road network data did not take into account
the speed of the street, the traffic lights and any other specifications of the street that
might affect the performance of the EMS system.
3.8. Methods Used to Achieve the Objectives

Four principal methods were used to achieve the objectives of this thesis. These methods are: (i) statistical data analysis, (ii) GIS network simulation, (iii) integrated use of data mining techniques and GIS methods, and (iv) hybrid clustering and GIS simulation. Details of these methods are provided in the following sub-sections.

3.4.1 Data Analysis

One of the objectives of thesis is to assess and evaluate the performance of an EMS system. In order to assess and evaluate the performance of the EMS system in Qatar, collected data was statistically analyzed. Statistical analysis was carried out in Minitab software, while descriptive statistics were analyzed through Microsoft Excel. Data analysis aimed at classifying incidents, determining response priorities and response times for the existing EMS case study. Microsoft Excel was used to analyze patterns in the historical incidents EMS calls. Such patterns included; variation and scatter of incidence during the course of a year and incidents zone analysis.

3.8.1.1. GIS Data Mapping

GIS mapping was used to develop district population density maps and zone incidents density maps. The various GIS layers, objects and attributes used in this research are shown in Figure 12.

3.8.1.2. Statistical Analysis

Statistical analysis was used to study and explore the pattern of large amount of data generated and collected in this study. In this research, it is also used to compare among different models and distributions to study whether there is a statistical significance or not. Statistical analysis is used because simple measures such as average or median may
not be enough for decision making and hence may not be suitable for drawing meaningful conclusions (Zomaya, 2005).

Figure 12: GIS Layer Required/Generated for/from the Research

Usually t-tests or ANOVA are used to test the statistical significance of data points if the distribution of these points is normal. There are several statistical methods used depending on the type and number of distributions available. In order to use these methods, first we need to test the normality of the data points and then depending on the results, a suitable method is used to test the significance of data provided. Figure 13 shows a flow chart that was used to select suitable methods for statistical analysis.
3.8.2. GIS Network Simulation

Another objective of this thesis is to compare the effectiveness of the existing (case study EMS) and a proposed EMS system through service area analysis and closest facility performance metrics. In order to achieve this objective, the Network Analyst extension tools in ArcGIS were used in simulation of the operations of the EMS systems.
3.8.2.1. Service Area Analysis

A network service area is a region that includes all accessible streets. Therefore service area analysis helps to evaluate accessibility of EMS units. In ArcGIS, concentric service areas show how accessibility varies with impedance. Service areas can be used to identify how many people can be served within the neighborhood or region. The steps for service area analysis can be summarized as follows:

5. Click on Network Analyst > New Service Area. A service area analysis layer will be created which shows six networks analysis classes—Facilities, Lines, Polygons, Point Barriers, Line Barriers, and Polygon Barriers.
6. Add all facilities that you need to find their service area under the facilities class. Specify the filed that represents their names.
7. In the Layer Properties dialog box, modify the analysis parameters as following:
   a. In the analysis setting tab, there is what is called Impedance: it reflects the cost attribute. It could be time, length, or anything which could be considered as cost.
   b. Add resection if there is any.
   c. Modify “the default breaks” which is the value of the cost.
   d. In the polygon generation tab, choose the appropriate polygon setting. We commonly use “Generalized” as a polygon type and overlapping under the multiple facilities options.
8. After that, click solve in the network analyst toolbar.

3.8.2.2. Closest Facility Location

The closest facility solver measures the cost of traveling between incidents and EMS facilities and determines which are nearest to one other. The steps for the closest facility analysis can be summarized as follows:
6. Click on Network Analyst > New Closest Facility. A closest facility analysis layer will be created which shows six network analysis classes: Facilities, Incidents, Routes, Point Barriers, Line Barriers, and Polygon Barriers.

7. Add all data points which are considered as facilities. Specify the field that represents their names.

8. Add all data points which are considered as incidents. Specify the field that represents their names.

9. In the Layer Properties dialog box, modify the analysis parameters as following:
   e. In the analysis setting tab, there is what is called Impedance: it reflects the cost attribute. It could be time, length, or anything which could be considered as cost. Choose the appropriate attribute which you seek to minimize.
   f. Specify (if you need) the default cutoff value, in which GIS will stop searching for a facility whenever that value is exceeded.
   g. Specify the number of facilities that need to be found for each incident under the “Facilities to find” parameter.
   h. Specify the travel form and add any restriction if there is any.

10. After that, click solve in the network analyst toolbar.

3.8.3. Integrated use of Data Mining and GIS Methods

Another objective of this thesis is to develop a multi-facility location allocation model for an EMS system through integrated use of fuzzy k-means, data mining techniques and GIS methods. As discussed in the literature review chapter of this thesis, an EMS system can be modeled as a location-allocation problem. Models for modelling EMS systems are usually classified as classical methods, non-classical methods, and contemporary methods. This research employs both non-classical (fuzzy k-means clustering) and contemporary methods (GIS) in the modelling of EMS systems. In this integrated methodology, solution procedures for general location science problems are used in conjunction with GIS modeling and simulation. This solution procedure is envisaged to enable the determination of optimal facility locations as well as the allocation of demand
to those facilities. The mathematical basis of such procedures is built on the application of various location models.

As discussed in section 2.3, relatively very little research has been done to implement p-median models in EMS systems. Therefore, the impact of p-median models and their effectiveness and efficiency in modelling EMS systems is less known in comparison to other models. This research contributes to knowledge by implementing a modified p-median method in the form of a fuzzy clustering algorithm. In addition, this research further studies the usefulness of this model in EMS systems by integrating the fuzzy k-means algorithm (developed in Matlab software) with GIS simulation (developed in Arc/GIS software). For the purposes of this research, a comparison will be made between a GIS based location allocation model and the fuzzy k-means-based model. Various components of this integrated methodology are discussed in the following sub-sections.

3.4.2.7 Data Clustering

Large volumes of data sets were used in the quantitative analysis of the EMS system in Qatar. In addition, large volumes of data were also used in the design and development of the proposed EMS system in Qatar. In order to organize the collected data and extract relevant information for strategic planning and decision making data mining methods and techniques were employed. Data mining, a non-classical method, was employed to; (a) classify data, and (b) extract useful information from the generated data. Thus, data mining techniques and heuristic based algorithms were used to formulate the EMS system as a location allocation problem. Data clustering was implemented through the fuzzy clustering method.

3.4.2.8 Fuzzy Clustering Method

Based on the discussions in the literature review chapter, a fuzzy clustering model was used to represent an EMS system. The logic behind the choice of the fuzzy k-mean
approach is that since an incident can be served by more than one facility depending on; its nearness, the road network, EMS facility availability, and any other circumstances, the fuzzy k-means gives a better representation of the EMS problem. Thus, given a set of objects $X = \{x_1, x_2, ..., x_n\}$, a fuzzy set $C_j$ is a sub-set of $X$ that assign for each object at $X$ a membership degree that ranges from 0 to 1. Fuzzy clustering of $k$ fuzzy clusters can be represented by a partition matrix $U= [u_{ij}]$, where $u_{ij}$ is the membership degree of object $x_i$ in fuzzy cluster $C_j$, $1 \leq i \leq n$ and, $1 \leq j \leq k$. The partition matrix should satisfy the following requirements (Abonyi & Feil, 2000; Han & Pei, 2011):

(iv) $0 \leq u_{ij} \leq 1$

(v) For each object $x_i$, $\sum_{j=1}^{k} u_{ij} = 1$, this is to ensure that every object participates in the clustering equivalently.

(vi) For each cluster $C_j$, $0 < \sum_{i=1}^{n} u_{ij} < n$, this to ensure that for every cluster, there is at least one object for which the membership value is nonzero.

The logic in the concept of partitioning in fuzzy clustering can be explained as follows: consider an EMS example in which the letters A, B, C, and D in Table 4 represent EMS facilities and time column represents the time required by each EMS facility to reach a customer. The following formula (Z) will be used to represent the degree of nearness of EMS facility to the customer.

$$Z = \begin{cases} 
1, & \text{if time } \leq 10 \text{ min} \\
\frac{10}{t}, & \text{if time } > 10 \text{ min}
\end{cases}$$  \hspace{1cm} (eq.1)

Based on the relationship in equation 1 above, the degree of nearness for facilities A and B (Table 4) is 1, whereas the degree of nearness for facility C = 0.5 while that for facility D = 0.25. Therefore, facilities A and B are better than the rest to serve the patient.
Table 4: EMS Facility and Time Required to Reach a Patient

<table>
<thead>
<tr>
<th>EMS Facility</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
</tr>
<tr>
<td>D</td>
<td>40</td>
</tr>
</tbody>
</table>

Thus, based on specific criteria (which in this case is the distance), the fuzzy clustering algorithm was used to group residents and incidents in the state of Qatar into clusters. The middle of the cluster will represent the location of the EMS hubs or spokes as named by Qatar ambulance services.

### 3.4.2.9. Fuzzy k-Means Algorithms

The objective of the implemented fuzzy clustering algorithm is to minimize the following function:

\[
J_m = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^m ||x_i - c_j||^2 ,
\]

(eq.2)

where

\[1 \leq m < \infty\]

is a weighting exponent which determines the fuzziness of the resulting clusters.

\[U = [u_{ij}]\]
is a fuzzy partition matrix of $X$,

$$c_j = [c_1, c_2, ..., c_k], c_j \in \mathbb{R}^n$$

is a matrix of cluster centers, which have to be determined.

The logic behind implementing Fuzzy k-means clustering is that it can be configured to locate EMS facilities by finding a meaningful grouping for a given data objects (i.e. population, residential houses and incidents in this research) based on the distance between these objects. Fuzzy k-means model is a variation from p-median model with squared Euclidean distances (Eislet, 2011). It is based on the minimization of the objective function shown in Eq. 2 iteratively by recognizing and updating $u_{ij}$ and $c_j$, such that:

$$u_{ij} = \frac{1}{\sum_{j=1}^{k} \left( \frac{\|x_i - c_j\|}{\|x_i - c_j\|} \right)^{m-1}}$$

$$c_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m}$$

The iteration will stop when the termination criterion $\varepsilon$ is met. Thus, the iterations will be terminated when:

$$\|U^{(y+1)} - U^{(y)}\| < \varepsilon$$

The notations and parameters used above can be defined as follows:

- $i$: set of houses/incidents, $i \in \{1, 2, 3, ..., n\}$
- $j$: set of EMS clusters, $j \in \{1, 2, 3, ..., k = 49\}$
- $y$: iteration steps, $y \in \{1, 2, 3, ...\}$
- $m$: A weighting exponent which determine the fuzziness of the resulting clusters, $1 \leq m < \infty$
$x_i$: The house/incidents locations

$c_j$: The centroid of the $j^{th}$ EMS cluster

$\epsilon$: termination criterion, $\epsilon = 1e^{-5}$

$U^y$: The fuzzy partition matrix

$u_{ij}$: The degree of membership of house $x_i$ in the EMS cluster $j$.

The algorithm is composed of steps shown in Figure 14.
3.4.2.10. **Constraints and Assumptions**

There are many constraints and assumptions that have been set while implementing the fuzzy k-means clustering algorithms in conjunction with GIS simulation. These constraints are:

3. The targeted response time to reach to any incident or house is 10 minutes considering an average speed of 80 km/hour. The response time of 10 minutes was chosen since it closely represents an ideal EMS system as discussed in the literature reviewed.

4. The number of EMS units required to be located is 49 units. Thus, the algorithm has been coded to generate 49 clusters and their centroids. The logic behind this constraint is that the investigation seeks to find out whether effective and efficient planning of an EMS system can provide better results without increasing the number of currently available EMS units.

Assumptions include the following:

5. Each house consist of 14 family members, typical of the culture in Qatar and the gulf region

6. The speed is 80km/hour. Thus, the distance between EMS unit and incidents is around 13.333 Km.

7. For clustering houses, it is assumed that each 1% of houses are accumulated near to each other and thus they are form one cluster. Therefore, the total number of clusters in and outside Doha is 100.

8. For clustering incidents, it is assumed that there is a total of 50 clusters in and outside Doha.
3.4.2.11. Verification and Validation of Clustering Techniques

Verification simply means building the system right. Thus, the built system should meet the requirements and specifications that have been specified in the design stage. In this research, verification is assured by using the fuzzy clustering model with the help of GIS software. Hence, it is assured that the response time needs to be within 10 minutes and the total number of EMS units is not exceeding the 49 units. In addition to that, Arc/GIS is used to verify that populations are allocated at the right district and the area of districts is representing the real districts given by the Ministry of Development Planning and Statistics. Moreover, it is used to verify that the proposed EMS locations are in free, not occupied area.

While verification means matching the built system with the specification, validation means meeting the user’s expectations and specifications. It means in other word, building the right system which represents the real situation and meets the operational requirements. The difference between validation and verification is presented in Figure 15. The built system in this research will be validated by three different ways, namely; using Arc/GIS, using another clustering algorithm called k-means, and using silhouette method. Each method is used to validate specific thing as described in the below graphs.

![Figure 15: Verification and Validation](image-url)
Starting with the first method which is implemented using Arc/GIS, it is used to investigate and validate whether the incidents/population are served within the specified response target or not. A random number of locations incidents were generated and the network analyses were used to examine whether the EMS facility can reach to them within the specified target (10 minutes) or not. Moreover, the same concept where used also to assure that all populations’ houses are covered within the acceptable response time. Results of this method are shown in section 4.2.3. of chapter 4.

Moving to the next method, it is used to validate the code itself and have some sense in whether the generated EMS locations are acceptable or not. This done by using one of well-known clustering algorithms called k-means. The difference between this algorithm and fuzzy k-means clustering is simply that this algorithm allows each point (incident/population) to belong only to one cluster, whereas each point in the fuzzy k-means has a weight associated with a particular cluster. Validation done by Matlab by finding the best ten clusters for the distribution of houses in 4 different zones in Qatar (Madinat Khalifa, Dahl al-Hamam, Almarkhyia, Freej Kulaib). The validation code is shown in Appendix H. The location of houses at these zones was given by the Ministry of Municipality and Urban Planning in Qatar. After implementing both algorithms in Matlab, the centroid of the ten clusters using both algorithms were found as shown in

Figure 16.

A comparison between the nearest centroids for each algorithm executed as following:

(iv) Find a random point. Here I took the center of one of the zones which is Madinat Khalifa-south.
(v) Find the distance between the center of that point and each centroid for each algorithm as shown in Table 5.
(vi) Considering the k-means centroid as the base, the percent difference was found by using the following equation.
\[
D\% = \left| \frac{\text{Dis}(c) - \text{Dis}(k)}{\text{Dis}(k)} \right| \times 100\% \quad \text{(eq.3)}
\]

where,

\(D\%\) = percent of distance difference

\(\text{Dis}(c)\): the distance between the center of the zone and the centroid of cluster \(i\) generated by fuzzy \(k\)-means clustering

\(\text{Dis}(k)\): the distance between the center of the zone and the centroid of cluster \(i\) generated by \(k\)-means clustering

---

**Figure 16: Validation of Fuzzy k-means Clustering**

It is found that the average difference between the two algorithms is ±13.1% as shown in Table 5.

The silhouette method was used to validate the number of clusters, which is here EMS units, in and out Doha. Thus, here we will examine if the number of existing EMS units
are enough for serving the population. In the currently existing EMS system, there are in total 49 EMS units distributed in the State of Qatar. 13 EMS units are out of Doha and 36 units are in Doha.

Table 5: Distance and Percent Difference for Each Clustering Algorithm

<table>
<thead>
<tr>
<th>Distance between the center of the zone and the centroid of the cluster</th>
<th>%Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-means</td>
<td>Fuzzy k-means</td>
</tr>
<tr>
<td>723.051</td>
<td>612.408</td>
</tr>
<tr>
<td>1698.807</td>
<td>1513.244</td>
</tr>
<tr>
<td>1905.520</td>
<td>1923.690</td>
</tr>
<tr>
<td>1339.377</td>
<td>1341.322</td>
</tr>
<tr>
<td>1384.489</td>
<td>1677.162</td>
</tr>
<tr>
<td>603.472</td>
<td>809.669</td>
</tr>
<tr>
<td>1844.732</td>
<td>1746.289</td>
</tr>
<tr>
<td>331.356</td>
<td>399.705</td>
</tr>
<tr>
<td>877.158</td>
<td>688.710</td>
</tr>
<tr>
<td>515.857</td>
<td>511.012</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
</tr>
</tbody>
</table>

Silhouette method finds how well each point lies within its cluster by going through the following steps (scikit-learn developers, 2011):

4- Find the average “a” distance between the point x and other points within the cluster.

5- Find the average “b” distance between the point x and all points in the nearest cluster.
Find the silhouette coefficient for $x$ “$S(x)$”, which is the difference between step 2 and step 1 divided by the maximum of step 2 and step 1. Thus, the equation of the silhouette coefficient is as below in equation 4:

$$
S(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}}
$$

(eq. 4)

The value of the coefficient will be within $\pm 1$. If the coefficient is near to 1, it means that the data are properly clustered, whereas it is badly clustered if it is near to -1 (The MathWorks, Inc., 2014). Any values near to zero it means that clusters are overlapping. The best number of clusters for the dataset will be determined by looking at the beak value (maximum coefficient) in the graph.

For in Doha datasets, we found as seen in Figure 17 that almost all clustering sizes have almost similar coefficient values. Moreover, all of them have a value close to zero, which means that the clusters are overlapping and points are close to each other. We can say that clustering size of 33 has the best coefficient value. In addition to that, the silhouette coefficient values for the clustering groups of out Doha data points are also found to be near to zero except for clustering size of 12 clusters (Figure 18). The best coefficient value is found to be at clustering size of 13. Overall, we can say that all population data are near together, therefore, and accordingly clusters are overlapping. Based on this method, it can be inferred that the total number of EMS units should be around 46 (i.e. $33 + 13$). Since the number of currently existing EMS units in the EMS system of Qatar is 49, the number of EMS units in and out of Doha can be considered acceptable.
Figure 17: Validating Number of Clusters - In Doha

Figure 18: Validating Number of Clusters - Out Doha
3.4.2.12. GIS Method

ArcGIS 10 was used for location allocation analysis of the EMS systems. ArcGIS location-allocation analysis uses six problem types, namely; minimize impedance (p-median), maximize coverage, minimize facilities, maximize attendance, maximize market share and target market share. For the application in this thesis only two methods were found to be more suitable. These two methods were the minimize impedance and the maximum coverage models.

Minimize Impedance
In this model, facilities are located such that the sum of all weighted costs between demand points and solution facilities is minimized. In the description of the study area, it has already been observed that the EMS system in Qatar differentiates between operations in summer and in winter. In addition, the EMS also differentiates between services in Doha and services out of Doha. Hence, there is no equitability.

Maximize Coverage
In this model, facilities are located such that as many demand points as possible are allocated to solution facilities within the impedance cutoff. For EMS location this model is important because emergency services are required to arrive at all demand points within a specified response time. The general procedure for location allocation in the GIS environment can be summarized as follows:

6. Click on Network Analyst > New Location-Allocation. A location-Allocation analysis layer will be created which shows six network analysis classes—Facilities, Demand Points, Lines, Point Barriers, Line Barriers, and Polygon Barriers.
7. Add all data points which are considered as candidate facilities. Specify the file that represents their names and their type, weight and capacity if there is any.
8. Add all data points which are considered as incidents. Specify the file that represents their names and weights if there is any.
9. In the Layer Properties dialog box, modify the analysis parameters as following:
   g. In the analysis setting tab, there is what is called Impedance: it reflects the cost attribute. It could be time, length, or anything which could be considered as cost. Choose the appropriate attribute which you are interested in.
   h. Specify the travel form and add any restriction if there is any.
   i. In the advanced setting tab, specify the type of the problem whether it is minimize impedance, maximize coverage, or any other problem from the list.
   j. Specify the number of facilities that you seek to choose from the candidate facilities.
   k. If required, specify the impedance cutoff value, in which GIS will stop searching for a facility whenever that value is triggered.
   l. Specify the type of transformation in the “impedance transformation”.

10. After that, click solve in the network analyst toolbar.

3.4.4. Hybrid Clustering and GIS Simulation

Another objective of this thesis is to simulate and compare the effectiveness of a variety of model-driven strategies for developing EMS systems. For EMS systems, the response time is a very important factor that determines the quality of the EMS. A number of factors affect the performance of EMS system. At the planning stage, the strategy and/or procedures used may affect the performance of the resultant EMS systems. A Hybrid clustering and GIS simulation framework was implemented to investigate the effects of a selected strategy or procedures for planning an EMS system. Figure 19 shows the process flow chart for implementing the hybrid clustering and simulation framework.

The objective of the GIS simulation was to investigate the performance of a number of model-driven strategies and procedures for planning EMS locations in order to achieve an acceptable EMS response time. Arc/GIS was used in this research to imitate the real system and to understand and examine how different EMS systems will respond to a given set of demand points (i.e. incidents in this work). This was achieved through the
Network analyst in GIS. Using the Network Analyst, one is able to solve realistic network conditions and sophisticated models to find one of the following (Esri, 2014):

- The quickest way to get from point A to point B
- Houses/incidents which are within five minutes of a fire station
- Market areas covered by a business
- Ambulances or patrol cars can respond quickest to an incident
- Find which store should be closed if a company has to downsize and at the same time maintain the most overall demand

The network analyst was used to carry out two types of analysis; first to find the service area and second to find the closest facility. Finding a service area means for each EMS unit we find the area that is surrounding it for response times of 8, 10, and 15 minutes. It includes also all streets within that range. Although this investigation was focused on a target of 10 minutes response time, response times of 8 and 15 minutes were included for comparison sake. On the other hand, the closest facility function determines the nearest facility (EMS units) and measures the cost of travelling between incidents and EMS units. In this regard, cost means the time taken to reach to the incidents.

In the GIS simulation, a total of nine models were developed, simulated and their performance evaluated based on the service area analysis and closest facility analysis. The aim of the simulation was to guide EMS planners in terms of; (a) the different types of simulation models that can be done in the GIS environment, (b) the performance of the various models that can be created in a GIS environment, and (c) the effect of planning strategies and/or planning procedures on the performance of and EMS system.
Figure 19: Flowchart for Implementing the Hybrid Clustering GIS Simulation
CHAPTER 4

RESULTS AND DISCUSSIONS

In order to meet the goal of this research, the following objectives were addressed: (1) to assess and evaluate the performance of the EMS system in the State of Qatar, (2) to compare the effectiveness of the existing and a proposed EMS system through service area analysis and closest facility performance metrics, (3) to develop a multi-facility location allocation model for an EMS system through integrated use of fuzzy k-means, data mining techniques and GIS simulation, and (4) to compare the effectiveness of model-driven strategies for developing EMS systems. The findings of these investigations are discussed in the following sections.

4.2. Data Analysis

In order to assess and evaluate the performance of the EMS system in the State of Qatar, data was collected from ambulance service in state of Qatar. The data was from October, 2012 to September, 2013. These data includes the following:

- Number of Alpha, Delta, and Charlie units available on hourly bases.
- Location of units in summer and winter.
- Incidents received, response time, and time to arrive to the hospital.

The hourly available EMS units in the current EMS system are provided in appendix E. Location of EMS units in winter and summer are different in the current system. The details of the location of EMS units in summer and winter are found in section 3.3.2. in chapter 3. The incidents served and response time data is provided in appendix B.
4.2.1. Response Times

An examination of the provided data on response times revealed that there were many outlier data points, and therefore, there was a need to specify the acceptable range for the response time. A box plot of the response times is shown in Figure 20.

![Box Plot (Minutes)](image)

Figure 20: Box Plot for Response Time.

It was decided that any response time falling beyond 47 minutes was considered unacceptable, and such data were treated as outliers. The reasons for such anomalies were attributed to incorrect recording of data by the ambulance service personnel on the raw data sheets provided (data was collected and recorded manually). Based on statistical analysis, the average response time for the current EMS system in Qatar was found to be 12.49 minutes and the median was 12.03 minutes. It is worth mentioning that most of the data recorded in the provided data sheets were between the EMS target ranges of 10 to 20 minutes. It is also important to highlight that on average a used EMS unit would be available for use again after 48.8 minutes. This represents the ‘recovery time’ of that EMS unit. Based on the calculated response times, it can be inferred that there is a “good enough” facility location plan of the currently existing facilities and resources.
4.2.2. Pattern of Incidents

From Figure 21, it is found that the number of incidents is somehow stable over the year and it ranged between 240 to 470 per day. However, the number of incidents increases slightly in the summer period (June, July, August, and September). Further observations of the number of incidents occurring every day shows that there is no correlation between the number of incidents and the day. Thus, incidents do not depend on the day of the week.

On the other hand, it was observed that there is correlation between the number of incidents and the time of the day. As shown in Figure 22, the number of incidents increases radically from 5:00AM to reach a peak by 11:00AM. Afterwards, the number of incidents decreases slightly from 11:00AM to around 4:00PM after which it increases again to reach another peak at 8:00PM and the cycle is repeated. It is worthy to mention that the number of incidents is minimum on Friday and Saturday morning in comparison to other days. However, it is maximum in the evening and night times. Moreover, the pattern of incidents on Fridays is different from the other days since it keeps increasing gradually while the pattern for other days have a specific shape with two peaks (one in the morning time and the other in evening time).
4.2.3. EMS Availability

According to the above discussion, it is expected to have a relation between the number of incidents and the number of EMS units available on an hourly basis. As shown in Figure 23, the number of EMS units increases on Friday afternoons until evening time. However, this increment in the number of EMS units applies also for Thursdays and Saturdays even though the incidents reach to their peaks in different time slots (in morning and evening). Moreover, the number of EMS units in other days is almost the same regardless of the incidents pattern.

4.2.4. Incidents-Zone Analysis

Qatar is divided into ninety-two zones. Figure 24 shows the incidents density in various zones in the state of Qatar. Historical data for one year showed that there are some zones that could be considered as “black zones” with respect to the number of incidents occurring in these zones.

Figure 22: Number of Incidents in Hourly Bases.
As shown in Table 6, these zones are 57, 55, 53, 37 and they are comprising more than 5% of the total incidents. Moreover, it was observed that around thirty zones are responsible for triggering 80% of the incidents. These zones are presented in green color in Figure 25 and are listed in the below table (Table 6) in descending order based on number of incidents. It is worth mentioning that other zones trigger less than 1% of incidents each.

Figure 23: Available EMS Units Every Hour.

As shown in Table 6, these zones are 57, 55, 53, 37 and they are comprising more than 5% of the total incidents. Moreover, it was observed that around thirty zones are responsible for triggering 80% of the incidents. These zones are presented in green color in Figure 25 and are listed in the below table (Table 6) in descending order based on number of incidents. It is worth mentioning that other zones trigger less than 1% of incidents each.
Figure 24: Zone Incidents Density
Table 6: Zones That Have More Than 1% Occurrence of Incidents.

<table>
<thead>
<tr>
<th>Zone #</th>
<th>Number of Incidents</th>
<th>%</th>
<th>Zone #</th>
<th>Number of Incidents</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>16278</td>
<td>13.18%</td>
<td>34</td>
<td>1904</td>
<td>1.54%</td>
</tr>
<tr>
<td>55</td>
<td>8353</td>
<td>6.76%</td>
<td>91</td>
<td>1763</td>
<td>1.43%</td>
</tr>
<tr>
<td>53</td>
<td>7805</td>
<td>6.32%</td>
<td>38</td>
<td>1737</td>
<td>1.41%</td>
</tr>
<tr>
<td>37</td>
<td>7338</td>
<td>5.94%</td>
<td>46</td>
<td>1720</td>
<td>1.39%</td>
</tr>
<tr>
<td>81</td>
<td>5326</td>
<td>4.31%</td>
<td>92</td>
<td>1702</td>
<td>1.38%</td>
</tr>
<tr>
<td>60</td>
<td>4408</td>
<td>3.57%</td>
<td>54</td>
<td>1667</td>
<td>1.35%</td>
</tr>
<tr>
<td>56</td>
<td>4180</td>
<td>3.39%</td>
<td>70</td>
<td>1511</td>
<td>1.22%</td>
</tr>
<tr>
<td>71</td>
<td>4110</td>
<td>3.33%</td>
<td>26</td>
<td>1505</td>
<td>1.22%</td>
</tr>
<tr>
<td>90</td>
<td>3676</td>
<td>2.98%</td>
<td>30</td>
<td>1491</td>
<td>1.21%</td>
</tr>
<tr>
<td>39</td>
<td>3602</td>
<td>2.92%</td>
<td>32</td>
<td>1412</td>
<td>1.14%</td>
</tr>
<tr>
<td>51</td>
<td>3411</td>
<td>2.76%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>2610</td>
<td>2.11%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>2582</td>
<td>2.09%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>2505</td>
<td>2.03%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>2427</td>
<td>1.97%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>2175</td>
<td>1.76%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>2000</td>
<td>1.62%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 25: Number of Incidents s. Zones
As mentioned earlier, EMS providers in Qatar classify response priority into three main types. Figure 26 shows that life threat encompasses more than half of the incidents (i.e. 58.8% of the cases) whereas 39.2% of the cases classified as urgent and only 2.1% considered as transport.

Figure 26: Number of Incidents For Each Response Priority - 2012/2013

It is also worth to mention that 27.9% of the cases are considered as trauma, 69% are medical and 3.1% are transport (see Figure 27).

Figure 27: Total Number of Cases for Each Class - 2012/2013
It is obvious from Figure 28 and Figure 29 that the usage of Alpha units throughout the year is extremely more than the other units. However, it is important to mention that Charlie units have been used in more than the quarter of life threat cases. This is because Charlie units, as mentioned earlier, are used for critical cases as they are stuffed with advanced equipment and staffs.

Figure 28: EMS Usage with Respect to Case Classification

Figure 29: EMS Usage with Respect to Response Priority
4.3. **Analysis of Alternative EMS Systems**

In this section, the performance of three EMS systems will be compared. These include: (i) the existing EMS system, (ii) an EMS system developed through a GIS method (updated existing system), and (iii) an EMS system developed by the fuzzy k-means algorithm. The comparison of the EMS systems was based on two main performance metrics i.e. (a) service area, and (b) closest facility.

4.3.1. **Existing EMS System**

In order to evaluate the existing EMS, the locations of the EMS units were integrated into the GIS environment. For the purpose of relative comparisons, the locations entered into the GIS environment did not differentiate between hubs and spokes. In addition, these locations were superimposed on the district layer of the map of Qatar. This procedure was necessary since the proposed EMS system design is based on districts rather than the hub catchment areas as is the case with the existing system. Since the winter system was envisaged to be better than the summer system, the winter EMS system was used for all comparative analysis. The locations of the EMS units are shown in Figure 30. These locations were then analyzed through GIS network analysis. Two types of analyses were carried out namely, the service area and the closest facility.

4.3.1.1. **Service Area Analysis**

The service area analysis was applied for ambulance response times of 8, 10, and 15 minutes. The service areas covered are shown in Figure 31.

Table 7 summarizes the results. It can be observed from both Figure 31 and Table 7 that as the ambulance response time increases, the service coverage is improved for population, districts, and the populated districts. This trend is an expected result. It can also be observed that 70.82% of districts are covered considering a 15 minutes service area. Values for the 8 and 10 minutes service area are relatively low.
Figure 30: EMS Locations of Existing System
Table 7: Summary of Service Area Analysis – Existing System

<table>
<thead>
<tr>
<th></th>
<th>Considering 8 Minutes</th>
<th>Considering 10 Minutes</th>
<th>Considering 15 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Related to The Population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Population Covered</td>
<td>99.352</td>
<td>99.39</td>
<td>99.895</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
<td>0.648</td>
<td>0.61</td>
<td>0.105</td>
</tr>
<tr>
<td><strong>Data Related to The Districts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Districts Covered</td>
<td>50.398</td>
<td>55.57029</td>
<td>70.822</td>
</tr>
<tr>
<td>Percentage of Uncovered Districts</td>
<td>49.602</td>
<td>44.43</td>
<td>29.178</td>
</tr>
<tr>
<td><strong>Data Related to The Populated Districts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Populated Districts Covered</td>
<td>58.342</td>
<td>77.0128</td>
<td>78.8884</td>
</tr>
<tr>
<td>Percentage of Uncovered Populated Districts</td>
<td>41.658</td>
<td>22.9872</td>
<td>21.1116</td>
</tr>
</tbody>
</table>

4.3.1.2. Closest Facility Analysis

The closest facility analysis is used to assign the best EMS for each population and/or incident points by finding the minimum time to the closest facility. Results of the closest facility analysis are summarized in Table 8. It can be observed that the results generated for the population data is better (values are lower) than those generated for the incidents data. This implies that a system developed based on population data may perform better than that based on incidents data. As is usually the case elsewhere, some service providers choose to base their system design on historical incidents rather than actual population. For both data types, it is found that more than 90% of data points are served within 10 minutes or less.
Service area of 8 minutes
Service area of 10 minutes
Service area of 15 minutes

Figure 31: Service Areas of the Existing System
Table 8: Closest Facility Results of the Existing System (minutes)

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>36.03695</td>
<td>67.414</td>
</tr>
<tr>
<td>Average</td>
<td>2.286398</td>
<td>3.471</td>
</tr>
<tr>
<td>Min</td>
<td>7.53E-05</td>
<td>9.6E-05</td>
</tr>
<tr>
<td>Median</td>
<td>1.588092</td>
<td>2.002</td>
</tr>
<tr>
<td>STD</td>
<td>2.631439</td>
<td>4.314</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>96.987</td>
<td>89.569</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>0.591</td>
<td>2.785</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>1.037</td>
<td>5.125</td>
</tr>
<tr>
<td>%more than 15</td>
<td>1.385</td>
<td>2.521</td>
</tr>
</tbody>
</table>

4.3.2. Proposed GIS Based EMS System (Updated Existing System)

For the GIS based methods, two models were used in the experiments. These models are the minimize impedance (which is essential a p-median model) and the maximize coverage. The location of the EMS units generated based on two GIS models (minimize impedance model and maximize coverage model) is shown in Figure 32.

Figure 32 shows that coincidentally, the maximize coverage and the minimize impedance methods in the GIS environment specifies exactly the same positions of the EMS. Therefore, further analysis was based on the maximize coverage model only since the other would give exactly the same performance measures for service area analysis and closest facility analysis.
Figure 32: Location of EMS units based on p-median and maximize coverage models
Using the maximize coverage model, a GIS based EMS system was developed by seeding existing locations of EMS units. This was regarded as an exercise to update the existing using the recent population data. The reason for this procedure was to find out if the GIS environment can be used iteratively i.e. in two steps to provide a better EMS system. Thus, all existing EMS units' locations were seeded as candidates locations and the GIS asked to select 49 units (the exact number of existing EMS units) in such a way to meet the objective (i.e. maximize the coverage or minimize the response time). It was expected that all 49 units will be selected. However, even at a cut-off value of 20 to 60 minutes, it was observed that only 43 units of the existing were selected and the rest of the unselected EMS units were considered as redundancy i.e. units that do not add any value to the EMS system (see Figure 33). This could be an indication that the existing system is over designed. This suggestion also seems to be supported by the cluster validation method in which 46 EMS units were deemed sufficient for the EMS system in Qatar using the recent population data.

Figure 33: Network Analyst Message

4.3.2.1. Service Area Analysis
The service area analysis, showed that more than 50% of the total and populated districts in State of Qatar is covered considering a target of 10 minutes, while 70% is covered if the target is 15 minutes (Figure 34 and Table 9)
### Data Related to The Population

<table>
<thead>
<tr>
<th></th>
<th>Considering 8 Minutes</th>
<th>Considering 10 Minutes</th>
<th>Considering 15 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Population Covered</td>
<td>97.709</td>
<td>97.747</td>
<td>99.895</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
<td>2.291</td>
<td>2.253</td>
<td>0.105</td>
</tr>
</tbody>
</table>

### Data Related to The Districts

<table>
<thead>
<tr>
<th></th>
<th>Considering 8 Minutes</th>
<th>Considering 10 Minutes</th>
<th>Considering 15 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Districts Covered</td>
<td>48.212</td>
<td>54.040</td>
<td>70.066</td>
</tr>
<tr>
<td>Percentage of Uncovered Districts</td>
<td>51.788</td>
<td>45.960</td>
<td>29.934</td>
</tr>
</tbody>
</table>

### Data Related to The Populated Districts

<table>
<thead>
<tr>
<th></th>
<th>Considering 8 Minutes</th>
<th>Considering 10 Minutes</th>
<th>Considering 15 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Populated Districts Covered</td>
<td>56.662</td>
<td>64.565</td>
<td>78.526</td>
</tr>
<tr>
<td>Percentage of Uncovered Populated Districts</td>
<td>43.338</td>
<td>35.435</td>
<td>21.474</td>
</tr>
</tbody>
</table>

#### 4.3.2.2. Closest Facility Analysis

Results of the closest facility analysis are shown in Table 10. From Table 10, it is clear that average response time does not exceed 3.2 minutes for both population and incidents data. Moreover, more than 90% of data points are served within 8 minutes.
4.3.3. New System Based on Fuzzy k-means

A new EMS system for Qatar was generated through integrated use of fuzzy k-means and GIS. The fuzzy clustering algorithm was used to propose new locations of EMS units using the same existing settings (i.e. 49 EMS units) based on incidents data for 2012/2013.

<table>
<thead>
<tr>
<th>Population</th>
<th>Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>40.304</td>
</tr>
<tr>
<td>Average</td>
<td>2.373</td>
</tr>
<tr>
<td>Min</td>
<td>0.00008</td>
</tr>
<tr>
<td>Median</td>
<td>1.638</td>
</tr>
<tr>
<td>STD</td>
<td>2.869</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>96.697</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>0.583</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>1.084</td>
</tr>
<tr>
<td>%more than 15</td>
<td>1.635</td>
</tr>
</tbody>
</table>

Figure 35 shows a Matlab graph of the various clusters and center of clusters as generated by the fuzzy k-means algorithm. In this map, the dark spots are the centers of the clusters
i.e. EMS units and the colored polygons are the various clusters suggested by the fuzzy k-means algorithm.

The proposed location of EMS units generated by the fuzzy k-means algorithm was integrated into the GIS environment. Figure 36 shows the results of this integration in a GIS map showing the locations of the EMS units.
Figure 35: Matlab Graph of the Various Clusters and Center of Clusters for the EMS System Proposed by the Fuzzy k-means Algorithm
Figure 36: EMS Locations Based on Integrated Use of Fuzzy k-means Clustering and GIS
4.3.3.1. Service Area Analysis

Table 11 and Figure 37 summarize the service area analysis. The service area analysis shows that the percentage of districts and populated district covered is less than the uncovered when the target is 8 or less. However, the coverage increased as the target threshold increases as well (Table 11). Covered areas are shown in Figure 37.

Table 11: Summary of Service Area Analysis – fuzzy k-means Clustering System

<table>
<thead>
<tr>
<th></th>
<th>Considering 8 Minutes</th>
<th>Considering 10 Minutes</th>
<th>Considering 15 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Related to The Population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Population Covered</td>
<td>99.373</td>
<td>99.381</td>
<td>99.972</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
<td>0.627</td>
<td>0.619</td>
<td>0.028</td>
</tr>
<tr>
<td>Data Related to The Districts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Districts Covered</td>
<td>46.286</td>
<td>55.17241</td>
<td>73.342</td>
</tr>
<tr>
<td>Percentage of Uncovered Districts</td>
<td>53.714</td>
<td>44.828</td>
<td>26.658</td>
</tr>
<tr>
<td>Data Related to The Populated Districts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Populated Districts Covered</td>
<td>43.3488</td>
<td>74.3685</td>
<td>75.4631</td>
</tr>
<tr>
<td>Percentage of Uncovered Populated Districts</td>
<td>56.6512</td>
<td>25.6315</td>
<td>24.5369</td>
</tr>
</tbody>
</table>

4.4.3.2. Closest Facility Analysis

Table 12 summarizes the results of the closest facility analysis. It can be observed that the results for both population and incidents are near to each other. However, the population data is slightly better than the incidents data.
Table 12: Closest Facility Results of the Fuzzy k-means Clustering System (minutes)

<table>
<thead>
<tr>
<th>Population</th>
<th>Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>40.454</td>
</tr>
<tr>
<td>Average</td>
<td>3.455</td>
</tr>
<tr>
<td>Min</td>
<td>0.000156</td>
</tr>
<tr>
<td>Median</td>
<td>2.268</td>
</tr>
<tr>
<td>STD</td>
<td>3.462</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>87.922</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>7.512</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>3.106</td>
</tr>
<tr>
<td>% more than 15</td>
<td>1.461</td>
</tr>
</tbody>
</table>

4.5. Performance Comparison of the Alternative EMS Systems

In order to compare the effectiveness of the existing and a proposed EMS systems results of service area analysis and closest facility performance metrics were used. The purpose of this section is to compare three alternative EMS system generated by three different methods. These three alternative EMS systems include: (i) Alternative I, which is the currently existing system that was designed based on the two-step floating catchment area method, (ii) Alternative II, which is a proposed system (updated existing system) based on the GIS method (maximal coverage), and (iii) Alternative III, which was designed based on the integrated use of fuzzy k-means and a GIS method. Figure 38, Figure 39, and Figure 40 show that in general the existing system performance (Alternative I) is more or less similar to and the EMS system designed based on the integrated use of the fuzzy k-means clustering algorithm and that from GIS (Alternative III). However, it can be observed that Alternative III gives slightly better results with respect to population coverage, while the existing system gives better results with respect to districts and populated districts coverage.
Service area of 8 minutes

Service area of 10 minutes

Service area of 15 minutes

Figure 37: Service Areas of The Fuzzy k-means Clustering System
Figure 38: Percentage of Covered Population for the Three Alternative EMS Systems

Figure 39: Percentage of Covered Districts - Existing, Updated, and Fuzzy k-means Systems
Based on the population data, it is found that the existing system generate better results if we are considering the population data points. However, the new system proposed by the GIS method gives almost the same results as the existing system although the EMS units is only 43 i.e. less by 6 units. This may suggest that the redundancy in the existing system makes the system more costly but without any gain in the performance of the system. Moreover, the updated existing system gives better results if we take into consideration the incidents data (see Table 13 and Table 14). The fuzzy k-means model implemented here does not give better results than either the existing or the updated system. In order to explain why this may be so, further investigations were conducted with the fuzzy k-means method. The inspiration for this further investigation rose from the fact that in implementing the fuzzy k-means methods a number of strategies and procedures can be used and may give different results.
Table 13: Closest Facility Results for the Existing, Updated, and Fuzzy k-means Clustering Systems - Population

<table>
<thead>
<tr>
<th></th>
<th>Existing System</th>
<th>Fuzzy k-means Clustering System</th>
<th>Updated Existing System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>36.03695</td>
<td>40.454</td>
<td>40.304</td>
</tr>
<tr>
<td>Average</td>
<td>2.286398</td>
<td>3.455</td>
<td>2.373</td>
</tr>
<tr>
<td>Min</td>
<td>7.53E-05</td>
<td>0.000156</td>
<td>0.00008</td>
</tr>
<tr>
<td>Median</td>
<td>1.588092</td>
<td>2.268</td>
<td>1.638</td>
</tr>
<tr>
<td>STD</td>
<td>2.631439</td>
<td>3.462</td>
<td>2.869</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>96.987</td>
<td>87.922</td>
<td>96.697</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>0.591</td>
<td>7.512</td>
<td>0.583</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>1.037</td>
<td>3.106</td>
<td>1.084</td>
</tr>
<tr>
<td>% more than 15</td>
<td>1.385</td>
<td>1.461</td>
<td>1.635</td>
</tr>
</tbody>
</table>

Table 14: Closest Facility Results for the Existing, Updated, and Fuzzy k-means Clustering Systems - Incidents

<table>
<thead>
<tr>
<th></th>
<th>Existing System</th>
<th>Fuzzy k-means Clustering System</th>
<th>Updated Existing System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>67.414</td>
<td>51.733</td>
<td>54.385</td>
</tr>
<tr>
<td>Average</td>
<td>3.471</td>
<td>3.682</td>
<td>3.183</td>
</tr>
<tr>
<td>Min</td>
<td>9.60E-05</td>
<td>3.05E-05</td>
<td>0.00006</td>
</tr>
<tr>
<td>Median</td>
<td>2.002</td>
<td>2.11</td>
<td>1.629</td>
</tr>
<tr>
<td>STD</td>
<td>4.314</td>
<td>4.284</td>
<td>4.358</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>89.569</td>
<td>86.305</td>
<td>90.01</td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>2.785</td>
<td>5.384</td>
<td>2.299</td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>5.125</td>
<td>5.136</td>
<td>4.637</td>
</tr>
<tr>
<td>% more than 15</td>
<td>2.521</td>
<td>3.175</td>
<td>3.054</td>
</tr>
</tbody>
</table>
4.6. Hybrid Data Clustering and GIS Simulations

4.4.2. Model-Based Strategies for Developing EMS Systems

The results discussed in section 4.3 have shown that in general different models for EMS planning can result in EMS systems that have different performance metrics. One of the objectives of this thesis is to simulate and compare the effectiveness of a variety of model-driven strategies for developing and planning EMS systems. Answers to such investigations can be used by EMS planners as recommendations for strategizing EMS planning with various objectives or for different seasons of the year. In order to provide information that can guide EMS planners in selecting a suitable model-driven strategy or procedure for designing and developing EMS systems using data mining techniques and GIS methods, a number of strategies for EMS planning were simulated in the GIS environment using the same data from the case study. The results of such simulations are reported in this section.

A total of eight models were configured for this simulation based on the flowchart method represented in Figure 19 in chapter 3. The Matlab code and results of fuzzy k-means clustering is shown in Appendices F and G respectively. In addition to the eight models, a comprehensive model is proposed by seeding the existing EMS locations and those generated by fuzzy k-means clustering. The significance of this comprehensive model lies in the logic that a lot of data from different data sources is usually available to EMS planners. Therefore using such data may result in a more informed EMS system which is expected to provide superior performance.

The GIS part of the comprehensive model uses the maximize coverage option in GIS. The eight proposed systems were initially divided into two types: (1) systems that consider the concepts of hubs and spokes as is the case with the EMS system in Qatar, and (2) systems that do not differentiate between hubs and spokes. In the design of EMS units using the integrated approach proposed in this thesis, this difference is basically procedural in the design process. Procedural alternatives are usually available to EMS
planners, for which the findings of this thesis will provide guidelines on which procedures provide better performance in the strategic planning of the EMS system. Other available strategies include options such as direct or indirect clustering. The direct clustering finds the best EMS locations directly without clustering first the houses into groups according to their locations. On the other hand, indirect clustering first groups the houses into 100 clusters based on their nearness, assuming that every 1% of the populations are accommodated near to each other. After this clustering, the model is then required to find the best 49 EMS locations. In addition to that, the direct and indirect clustering is divided further into two streams, one stream based on population and the other stream based on the incidents. Different divisions of proposed models are represented in Figure 41 and summarized in One of the main concepts used in this research is data mining. After being supplied by raw data (population data or incidents data), the fuzzy k-means algorithm clusters the data as shown in Figure 35. In order to derive meaning from these data, data mining algorithms were used to calculate performance metrics from the clustered data. For this analysis only models (a) to (h) were evaluated since in these models a number of strategies were coded in Matlab software in which the fuzzy k-means algorithm was implemented. The results of such analysis are reported in the following sections.

Table 15.

![Figure 41: Proposed Systems.](attachment:image.png)
One of the main concepts used in this research is data mining. After being supplied by raw data (population data or incidents data), the fuzzy k-means algorithm clusters the data as shown in Figure 35. In order to derive meaning from these data, data mining algorithms were used to calculate performance metrics from the clustered data. For this analysis only models (a) to (h) were evaluated since in these models a number of strategies were coded in Matlab software in which the fuzzy k-means algorithm was implemented. The results of such analysis are reported in the following sections.

Table 15: Description of Model Strategies Simulated in the GIS Environment

<table>
<thead>
<tr>
<th>Model code</th>
<th>Description</th>
<th>Model Abbreviations</th>
</tr>
</thead>
</table>
| (a)        | Direct: find EMS directly without clustering houses first  
- With Hubs and spokes: considering that there are two different types of EMS locations.  
- Taking into account the whole population in Qatar. | Direct-H&S-Qatar |
| (b)        | Direct: find EMS directly without clustering houses first  
- With Hubs and spokes: considering that there are two different types of EMS locations.  
- Taking into account incidents only. | Direct-H&S-Incidents |
| (c)        | Direct: find EMS directly without clustering houses first  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account incidents only. | Direct-NoH&S-Incidents |
| (d)        | Direct: find EMS directly without clustering houses first  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account the whole population in Qatar. | Direct-NoH&S-Qatar |
| (e)        | Indirect: cluster houses first and then find EMS locations  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account the whole population in Qatar. | Indirect-NoH&S-Qatar |
| (f)        | Indirect: cluster houses first and then find EMS locations  
- Without Hubs and spokes: dealing with EMS locations as EMS units regardless of it is located in hub or spoke location.  
- Taking into account incidents only. | Indirect-NoH&S-Incidents |
| (g)        | Indirect: cluster houses first and then find EMS locations | Indirect-H&S-
Running the Matlab code is differs slightly from one model to another in terms of the inputs and/or the execution pattern as shown in Figure 19. However, the final output for all is the same which is mainly the location of the proposed EMS units. In addition to that, the number of people served by each EMS unit and the distance between the nearest EMS units and each data point are calculated from the clustered data. Using a computer with 2.67GHz processor and 4GB RAM, the time required to execute different strategies is represented in Table 16. All of the other results are shown Appendix G.

<table>
<thead>
<tr>
<th>Model</th>
<th>Execution Time (Seconds)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>147.895</td>
<td>1</td>
</tr>
<tr>
<td>b</td>
<td>323.7</td>
<td>6</td>
</tr>
<tr>
<td>c</td>
<td>304.806</td>
<td>5</td>
</tr>
<tr>
<td>d</td>
<td>156.422</td>
<td>2</td>
</tr>
<tr>
<td>e</td>
<td>276.566</td>
<td>4</td>
</tr>
<tr>
<td>f</td>
<td>327.705</td>
<td>7</td>
</tr>
<tr>
<td>g</td>
<td>276.108</td>
<td>3</td>
</tr>
<tr>
<td>h</td>
<td>335.51</td>
<td>8</td>
</tr>
</tbody>
</table>

The rank of the models with respect to execution time is also shown in Table 16. Table 16 shows that different strategies require different execution time. The shortest execution time is only about 2.5 minutes while the largest execution time is about 5 minutes. These differences in implementing an available algorithm for fuzzy clustering and information...
extracting are of no practical significance. Therefore, it can be concluded that algorithm run time when using the fuzzy k-means algorithm for the development of an EMS units under the conditions described in this thesis is very small and hence has little impact on the design of an EMS system. This result can also mean that EMS planners can evaluate and re-evaluate the performance of an existing EMS in no time and hence update plans for an existing system can be easily generated. Therefore, most of the effort in terms of time and cost will be in the practical implementation of the proposed solution for example re-location EMS facilities.

In Figure 41, the comprehensive model (model (i) in Table 15) was created by seeding all the possible EMS locations available from different models created through different methods and procedures. This scenario is necessary for comparison since it simulates a situation in which the designers and developers of the EMS system have valuable information regarding how the EMS system should work. It is important to mention that most of EMS locations specified in the comprehensive model (41 in this case) were taken from the location seeded by the fuzzy clustering algorithm and only 8 of them were taken from the existing system. The location of proposed EMS units for the nine models is shown in Figure 42 (a-i).

For simplicity and to mimic the real situation, it is assumed that each house consist of 14 family members (Table 17).

Table 17: Number of Family Members in Qatar

<table>
<thead>
<tr>
<th>Zone #</th>
<th>Number of Buildings</th>
<th>Number of Population</th>
<th>Family Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>1217</td>
<td>14725</td>
<td>12.09942</td>
</tr>
<tr>
<td>33</td>
<td>453</td>
<td>5197</td>
<td>11.47241</td>
</tr>
<tr>
<td>34</td>
<td>2167</td>
<td>35125</td>
<td>16.20904</td>
</tr>
<tr>
<td>35</td>
<td>516</td>
<td>7702</td>
<td>14.92636</td>
</tr>
<tr>
<td><strong>Average =</strong></td>
<td></td>
<td></td>
<td><strong>13.68≈14</strong></td>
</tr>
</tbody>
</table>
This is calculated using the data of Qatar Area Referencing system for four zones in Doha given by the Ministry of Municipality and Urban Planning and by considering the below equation.

Number of Family Members = Average (population at each zone/buildings at each zone).
In this study, the districts were divided into In-Doha and Out-Doha as shown in Figure 43.
Figure 43: Districts in and Out Doha

4.4.4. Comparison of Model Based Strategies for Developing EMS Systems

4.4.4.1. Service Area Analysis
After implementing service area analysis, the area covered for each system is represented in Figure 44 and detailed results provided in appendix C. The analysis will take mainly three ways; namely, the analysis of the districts, the analysis of the population, and the analysis of the covered populated area.

For districts, it is logically clear that as we reduce the response time (the travel time), the percentage of uncovered districts will be increased. Thus, if the response time set to be 8 minutes, larger number of districts will not be covered but among the whole ten models, the (indirect-No H&S-incidents) model gave the best results while the (indirect-H&S-Qatar) model gives the worst (Figure 45). For response times of 10 and 15 minutes, the best performance was observed for indirect-No H&S-incidents model while the worst performance was for the indirect-NoH&S-Qatar model. In overall, it can be inferred that the indirect-No H&S-incidents model provides the best performance for all time ranges in terms of district coverage.

In the previous paragraph, models were analyzed and compared with respect to the districts covered. However, the results of such an analysis do not help us in determining properly how many people are covered. Therefore, it is necessary to analyze and compare among these models in terms of how many populations are not covered. As represented in Figure 46, it is clear that there are three models that can be considered as extremes and these will be discarded from the analysis. These extremes are:

- Indirect, No hubs and Spokes, Incident Model
- Indirect, Hubs and Spokes, Qatar Model
- Indirect, Hubs and Spokes, Incidents Model.
Figure 44: Service Area of Proposed and Existing Models.
After filtering out the three models (see Figure 47), it is found that the indirect-NoH&S-Incident model and the comprehensive models have the minimum values (population uncovered) in all time ranges while the worst model is shared between the (Direct-H&S-Incidents) model and the (Direct-NoH&S-Incidents) model. It can be observed that the comprehensive model has more improved outputs in terms of population coverage as the response time target is reduced to 10 and 8 minutes, while indirect-NoH&S-Incident model has better results if the target is 15 minutes (see Table 18). It is also found that the uncovered population in the existing system is more than 79% of the uncovered...
population in the comprehensive and indirect-NoH&S-Incident model in all time ranges. Thus, it can be inferred that the improvement of both the comprehensive and the indirect-NoH&S-Incident model is more than 79% than the existing model.

Table 18: Uncovered Population- (Indirect-NoH&S-Incident) and Comprehensive Models

<table>
<thead>
<tr>
<th></th>
<th>Indirect-NoH&amp;S-Incidents</th>
<th>The comprehensive model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 min</td>
<td>10 min</td>
</tr>
<tr>
<td>Uncovered population</td>
<td>4310</td>
<td>1307</td>
</tr>
<tr>
<td>Percentage of Uncovered Population</td>
<td>0.277</td>
<td>0.084</td>
</tr>
</tbody>
</table>

In order to decide which model has better coverage, an analysis that focuses on determining how much populated area covered by each model is carried out. As shown in Figure 48, it is found that the best model in terms on covered populated area is the comprehensive model with a percentage exceeding 83% in all time ranges and more specifically when the response times become less.
Thus, from the above analysis, it could be concluded that the comprehensive model has more improved results especially for small response time targets.

4.4.4.2. Closest Facility Analysis

In this analysis, EMS service will be analyzed based on both population and incidents distribution. GIS is tasked to find the nearest EMS unit to each population/incident point using the closest facility tool under the network analysis-ArcGIS. Thus, GIS is used to simulate and find the smallest response time required to serve the incidents. Two tables were generated to summarize the results of simulating the populations and incidents data points. They are represented in Table 20 and Table 20 for population and incidents respectively. Ideally a small response time is often sought after. Therefore, among all aforementioned statistics i.e. maximum, minimum and average, the minimum times are highlighted in green and it is found that all of these minimum values can be obtained from the eight suggested models and none related to the existing EMS system. For instance, the minimum average value is found from model (g): Indirect-H&S-Qatar if we are taking into consideration the population data points, and it is model (a): Direct-H&S-Qatar if we are considering the incidents data points.
### Table 19: Closest Facility Results - Population

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>Existing</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>46.44373</td>
<td>39.98912</td>
<td>40.454</td>
<td>46.572</td>
<td>47.40103</td>
<td>37.90408</td>
<td>47.65405</td>
<td>30.34855</td>
<td>36.03695</td>
<td>45.94061</td>
</tr>
<tr>
<td>Avg</td>
<td>2.035561</td>
<td>3.331505</td>
<td>3.4553</td>
<td>1.927</td>
<td>2.631258</td>
<td>3.346153</td>
<td>2.698915</td>
<td>3.670968</td>
<td>2.286398</td>
<td>1.962472</td>
</tr>
<tr>
<td>Min</td>
<td>0.000301</td>
<td>0.000534</td>
<td>0.000156</td>
<td>5.89E-05</td>
<td>0.000181</td>
<td>0.000191</td>
<td>2.37E-05</td>
<td>0.000154</td>
<td>7.53E-05</td>
<td>0.000534</td>
</tr>
<tr>
<td>Med</td>
<td>1.336902</td>
<td>1.953867</td>
<td>2.267503</td>
<td>1.351</td>
<td>1.432455</td>
<td>2.178519</td>
<td>1.59097</td>
<td>2.293847</td>
<td>1.588092</td>
<td>1.504223</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>97.433</td>
<td>88.762</td>
<td>87.922</td>
<td>93.336</td>
<td>89.980</td>
<td>93.716</td>
<td>89.245</td>
<td>96.987</td>
<td>98.710</td>
<td></td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>1.650</td>
<td>6.318</td>
<td>7.512</td>
<td>1.400</td>
<td>1.387</td>
<td>6.116</td>
<td>1.272</td>
<td>5.185</td>
<td>0.591</td>
<td></td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>0.532</td>
<td>3.684</td>
<td>3.106</td>
<td>0.303</td>
<td>1.796</td>
<td>2.907</td>
<td>1.839</td>
<td>3.016</td>
<td>1.037</td>
<td></td>
</tr>
<tr>
<td>%more than 15</td>
<td>0.385</td>
<td>1.235</td>
<td>1.461</td>
<td>0.374</td>
<td>3.482</td>
<td>0.996</td>
<td>3.173</td>
<td>2.555</td>
<td>0.297</td>
<td></td>
</tr>
</tbody>
</table>

### Table 20: Closest Facility Results - Incidents

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>Existing</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>81.066</td>
<td>49.60076</td>
<td>51.73285</td>
<td>81.336</td>
<td>82.335</td>
<td>50.532</td>
<td>61.736</td>
<td>54.158</td>
<td>67.414</td>
<td>80.934</td>
</tr>
<tr>
<td>Min</td>
<td>1.924E-05</td>
<td>0.000249</td>
<td>3.05E-05</td>
<td>0.000498</td>
<td>0.000134</td>
<td>6.989E-05</td>
<td>8.55E-05</td>
<td>9.6E-05</td>
<td>0.000248</td>
<td></td>
</tr>
<tr>
<td>Med</td>
<td>2.132</td>
<td>1.82827</td>
<td>2.110581</td>
<td>2.193</td>
<td>2.219</td>
<td>2.294</td>
<td>1.648</td>
<td>2.274</td>
<td>2.002</td>
<td>2.252</td>
</tr>
<tr>
<td>STD</td>
<td>5.053</td>
<td>4.374296</td>
<td>4.284488</td>
<td>5.0546</td>
<td>5.422</td>
<td>3.804</td>
<td>5.441</td>
<td>5.234</td>
<td>4.314</td>
<td>4.807</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>85.971</td>
<td>85.690</td>
<td>86.305</td>
<td>86.449</td>
<td>84.119</td>
<td>89.812</td>
<td>85.304</td>
<td>87.131</td>
<td>89.569</td>
<td></td>
</tr>
<tr>
<td>% 8 to 10 min</td>
<td>3.857</td>
<td>5.510</td>
<td>5.384</td>
<td>4.134</td>
<td>3.065</td>
<td>4.062</td>
<td>2.728</td>
<td>3.760</td>
<td>2.785</td>
<td></td>
</tr>
<tr>
<td>% 10-15 min</td>
<td>6.457</td>
<td>5.639</td>
<td>5.136</td>
<td>5.574</td>
<td>7.589</td>
<td>3.989</td>
<td>5.554</td>
<td>3.978</td>
<td>5.125</td>
<td></td>
</tr>
</tbody>
</table>
The other part of the analysis is related to the number of points served in specific range of time and in this research we are focusing on response time of 10 minutes or even less to meet other international standards and also to be able to serve most of emergency cases such as cardiac arrests which needs response time of less than 10 minutes. We can also determine from this analysis which model meets the equity metric and serve fairly most of the population/incidents without differentiating between population/incidents in or out of Doha. It can be observed that the comprehensive model has better results than other models in terms of population data points and model (f): Indirect-NoH&S-Incidents is better if we are taking into consideration the incidents.

In addition to that and since equity is one of the important metric the response time of rural areas in the State of Qatar was studied separately to detect and have general idea about the performance of EMS units in such areas. According to the current ambulance service provider in Qatar, the targets of the current system in urban areas focuses on serving incidents within 10 to 15 minutes whereas the target for rural areas is within 15 to 20 minutes. Thus, there is a difference between urban and rural areas. However, as seen in Table 21, there is improvement in rural areas response time. For example, the average and median of the response time is around 4 minutes in model (d) considering the population data points. Moreover, most of the points (around 94%) will be served within 10 minutes. This percentage is higher in comparison to the existing system which does not aim to exceed 80%. However, the difference in the rural and urban areas coverage in the existing system is probably due to other factors such as operational costs and population density which favors the densely populated areas in Doha City.

For the data generated by analyzing the incidents (Table 22), it is found that the median and average values are higher than the results generated with respect to the population. However, still the proposed systems have better results over the existing one. Thus, model (b) outperforms others by having the minimum average and median and maximum 8 to 10 minutes coverage percentage. This gives an indication that the suggested models address the equity metric better than the existing one.
4.4.5. Statistical Analysis

Simple comparison between these models using for example the average is not enough to generate conclusions due to random noise of the data. Thus, statistical analysis was used to derive more information from the results. The purpose of this section is to report comparisons among the distribution of response time generated using each model.

Using the response time data generated by the closest facility analysis, it is found that none of the data from the ten models follows a normal distribution (Appendix D) and data is independent. Therefore, the Kruskal-Wallis test, one of non-parametric tests, was used to find if there could be a statistical significance among the data sets for response times.

Figure 49 shows the results obtained from Minitab. The $p$-value for all response time data sets was found to be less than 0.05. Therefore, it can be inferred that the differences in response time among all models is statistically significant. Figure 49 shows p-values of 0.000 for the unadjusted and adjusted for ties methods indicating that the null hypothesis that there is no difference between the response time medians can be rejected.

![Figure 49: Results of Kruskal-Wallis Test.](image)

\[
\begin{align*}
H &= 61585.74 \quad DF = 9 \quad p = 0.000 \\
H &= 61585.74 \quad DF = 9 \quad p = 0.000 \quad (\text{adjusted for ties})
\end{align*}
\]
Table 21: Closest Facility Results in Rural Areas - Population

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>Existing</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>46.444</td>
<td>39.989</td>
<td>40.454</td>
<td>46.572</td>
<td>47.401</td>
<td>47.654</td>
<td>27.035</td>
<td>36.037</td>
<td>45.941</td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>0.0010</td>
<td>0.6388</td>
<td>0.3821</td>
<td>0.0018</td>
<td>0.0026</td>
<td>0.5295</td>
<td>0.0088</td>
<td>0.3821</td>
<td>0.0275</td>
<td>0.1019</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>92.610</td>
<td>43.975</td>
<td>54.256</td>
<td>93.219</td>
<td>52.724</td>
<td>52.226</td>
<td>53.649</td>
<td>73.223</td>
<td>86.523</td>
<td></td>
</tr>
</tbody>
</table>

Table 22: Closest Facility Results in Rural Areas - Incidents

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>Existing</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>60.525</td>
<td>39.900</td>
<td>40.365</td>
<td>60.654</td>
<td>61.483</td>
<td>37.814</td>
<td>61.736</td>
<td>36.827</td>
<td>44.744</td>
<td>60.022</td>
</tr>
<tr>
<td>Min</td>
<td>0.002996526</td>
<td>0.003459874</td>
<td>0.003567130</td>
<td>0.038950215</td>
<td>0.013340470</td>
<td>0.006745329</td>
<td>0.008020674</td>
<td>0.015433921</td>
<td>0.019303534</td>
<td>0.002996507</td>
</tr>
<tr>
<td>% 8 min or less</td>
<td>31.592</td>
<td>47.091</td>
<td>39.927</td>
<td>28.883</td>
<td>25.044</td>
<td>39.387</td>
<td>25.741</td>
<td>29.638</td>
<td>34.330</td>
<td>26.054</td>
</tr>
<tr>
<td>% more than 15</td>
<td>47.117</td>
<td>23.726</td>
<td>29.050</td>
<td>53.465</td>
<td>45.865</td>
<td>26.739</td>
<td>41.275</td>
<td>37.369</td>
<td>40.490</td>
<td>35.956</td>
</tr>
</tbody>
</table>
4.4.6. Comparison Based on Weighted Parameters

In this section, results from previous analysis are summarized and a simple comparison among all models is created just to help tentatively in determining which model could be considered as the best among all ten models by meeting the higher number of experiment parameters (experiment objectives). For each parameter there is the best model that gives the most appropriate values. For instance, if we need to find the best model in terms of area of population covered, the comprehensive model is the best. On the other hand, if we are looking at the minimum of minimums of the response time with respect to the population distribution, model (g): indirect-H&S-Qatar will be the best. Based on the analyses in the previous section, there are fifteen parameters (experiment objectives) which we can be used to compare among the models and find the one that meets the larger number of these parameters. These parameters are:

1. Number of covered districts; within 8 minutes, within 10 minutes, within 15 minutes
2. Number of covered population; within 8 minutes, within 10 minutes, within 15 minutes
3. Area of populated districts covered; within 8 minutes, within 10 minutes, within 15 minutes
4. Minimum value of maximum response time based on population
5. Minimum value of average response time based on population
6. Minimum value of minimum response time based on population
7. Minimum value of median of the response time based on population
8. Minimum value of standard deviation of the response time based on population
9. Equity and percentage of incidents served within 8 minutes or less based on population
10. Minimum value of maximum response time based on incidents
11. Minimum value of average response time based on incidents
12. Minimum value of minimum response time based on incidents
13. Minimum value of median of the response time based on incidents
14. Minimum value of standard deviation of the response time based on incidents
Equity and percentage of incidents served within 8 minutes or less based on incidents
Assuming the same weight for all parameters, Table 23 shows the ranks for all models
based on the results by choosing the minimum values for objectives 4 to 8 and 10 to 14,
and maximum values for others.

Table 23: Models Ranking

<table>
<thead>
<tr>
<th>Model obj</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
<th>Existing</th>
<th>Comprehensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1.a</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>1.b</td>
<td>10</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>1.c</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>2.a</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>2.b</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>2.c</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>3.a</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>3.b</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>3.b</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>10</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Sum</td>
<td>134</td>
<td>107</td>
<td>114</td>
<td>130</td>
<td>58</td>
<td>149</td>
<td>83</td>
<td>77</td>
<td>144</td>
<td>159</td>
</tr>
</tbody>
</table>

It is important to note that the best model takes score of 10, while the worst take 1. It can
be observed that the highest total is scored by the comprehensive model. Thus, it can be
said that it has the best overall performance that satisfies most parameters and trade-offs
between them. The ranks of the models are also shown in Table 23 along the rank row.
CHAPTER 5

CONCLUSION

Emergency medical services (EMS) must be accessible and distributed in a manner that allows reliable and timely response to emergency calls from demand areas. However, many factors, including ambulance location sites and population growth, can adversely influence EMS coverage and response time. As such, service providers are constantly seeking methods for evaluating, improving and optimizing existing EMS systems. This research has discussed the integrated use of a fuzzy k-means clustering algorithm and GIS network simulation in determining optimal location of EMS facilities and evaluating equitability and accessibility of EMS systems.

The main goal of this research is to develop an innovative approach for modeling, simulating and evaluating multi-facility location allocation problems in emergency medical service systems. This goal was achieved through an integrated use of; a fuzzy k-means algorithm, data mining techniques and GIS simulation.

In the proposed framework of analysis, the fuzzy k-means algorithm was successfully used to generate new EMS locations for existing and new demand points. Demand points were allowed to belong to several clusters with different degrees of membership. This allowed new demands to be split between candidate EMS centers in a bid to optimize EMS coverage, accessibility and equitability. GIS network analysis was applied to fine tune EMS locations obtained from the fuzzy k-means algorithm. In this way, eight different models were developed for GIS simulation and analysis. In addition, GIS based locations were created by seeding existing EMS locations, and those locations generated by fuzzy k-means, as candidate sites. Proposed methods were applied to the EMS system in the State of Qatar. Numerical results showed that the proposed approaches were able to generate EMS locations in real time and with better population coverage as well as better service area coverage.
10.1. Major Findings

In order to achieve the main goal of this thesis, the EMS system in Qatar (case study) was cast as a multi-facility location problem. The solution procedure was successfully provided through integrated use of fuzzy k-means and GIS simulation. A hybrid clustering and GIS simulation framework for analyzing various model-driven strategies for EMS planning and development was successfully developed and implemented to; (a) compare the effectiveness and efficiency of the existing and proposed EMS system, and (b) compare the effectiveness and efficiency of model-driven strategies/techniques that can be applied to solve the EMS location allocation problem.

10.1.1. Assessment and Evaluation of the EMS System in Qatar

The assessment of the EMS system in Qatar revealed a number of issues in the analytical process. It was discovered that the existing system design was acceptable since relatively good EMS response times can be achieved. However, two notable concerns were discovered, namely;

(i) The existing system design differentiates between summer and winter seasons. This is contrary to the concept of equitability. In fact, it was observed from the incidents data that the number of emergence incidents increase slightly in summer. However, according to the ambulance personnel, their winter model is said to be better than the summer model. The fact that currently there are no major concerns in the EMS system in Qatar can be partially explained by the fact that analysis in this thesis has discovered that the EMS system is actually over designed. This could also explain that even in summer, when a derated EMS operates, the EMS system in Qatar has not witnessed any major setbacks.

(ii) The existing system design differentiates between rural (out of Doha City) and urban (inside Doha City). Again, this is contrary to the concept of equitability.
The fact that there have been no major consequences related to this differentiation could also be explained by the same reasons discussed above.

10.1.2. Comparison of Existing EMS System with Alternative Systems

The overall performance of the existing EMS system in the State of Qatar can be rated as good at its best (since analysis has shown that it can achieve response time of about 12.5 minutes) and satisfactory at its worst. A comparison of the existing system with a proposed GIS based system and a fuzzy k-means based system showed that their performance metrics are relatively comparable. However, it can be noted that surprisingly, the GIS based proposed system was designed with only 43 EMS units while the existing system currently operates with 49 EMS units. This shows that the existing system is either over designed or was designed to operate with some redundancy. In engineering redundancy is often incorporated into a design to cover for any uncertainties that may be related to the operations of the system. Although no explanation could be provided from the service providers, redundancy is an extra cost. However, a properly designed-in redundancy can prove to be cost effective in the long run.

10.1.3. Multi-Facility Location Allocation Model

The development of a multi-facility location allocation model was explored in this research. This development was based on a fuzzy k-means algorithm. The model was successfully implemented in Matlab and worked well with data mining techniques which were used to extract valuable information from the algorithm generated clusters. The model was used to solve the multi-facility EMS problem for the State of Qatar without increasing the costs (i.e. using the same number of EMS units as in the existing system). It was observed that the fuzzy k-means algorithm could solve the multi-facility location allocation problem in real time. However, it was noted that the performance of the fuzzy k-means based EMS system was relatively comparable to that of the existing system. It is explored that fuzzy k-means approach should give superior results, it was concluded that
the performance of this algorithm could be improved if a number of model-based strategies for implementing the fuzzy k-means method are used. This early observation prompted further investigations into the effects of various model-based strategies on the performance of the resulting EMS system.

10.1.4. Comparison of Model-Based Strategies for EMS Development

Proposed model-based strategies were compared used two main performance metrics in the GIS simulation environment. These metrics are; the service area analysis and the closest facility analysis. The performance of nine variations of the fuzzy k-means algorithm was analyzed and compared to the existing EMS model in the State of Qatar. Included in these models was a comprehensive model that was implemented to take advantage of; (i) large amounts of data available to EMS planners, and (ii) the various sources of data available to EMS planners. As such, this model was configured to fully utilize available data, knowledge and information in the process of designing and developing an EMS system.

The overall results showed that for service area analysis, the comprehensive model outperformed other models including the existing EMS system. The overall results also showed that the comprehensive model outperformed other models including the existing EMS system in particular when equitability is an issue of major concern.

10.2. Conclusions Drawn from Findings

Based on the findings of this thesis, the following conclusions can be drawn:

✓ The GIS environment can be used to model, simulate, assess, evaluate and compare the performance of a number of EMS systems
✓ The fuzzy k-means method can be used to model multi-facility location allocation problems
The fuzzy k-means algorithm can be used to solve a multi-facility location allocation problem for EMS systems in real time.

Integrated use of fuzzy k-means methods and GIS based methods can be used to model EMS systems that give improved overall performance.

Different model-based strategies for designing and developing EMS systems based on the fuzzy k-means method result in EMS systems with different performance measures.

No one model-based strategy for developing EMS systems can be implemented to outperform other models if a large number of performance parameters are considered.

Knowledge informed comprehensive models can be used to design and develop EMS systems with superior performance.

10.3. Research Contributions

The contributions of this research can be summarized as follows:

1. Elsewhere, it has been claimed that one way of improving an existing EMS system is to find an optimal re-location of EMS facilities. The proposed EMS systems discussed in this thesis resulted in the re-location of various EMS facilities for an improved performance. Therefore, the results of this research agree with other results. Hence, the findings of this research can be considered as a further confirmation of various such and similar investigations.

2. It has also been observed by other researchers that an existing EMS can be improved without increasing the costs. In this thesis, an existing EMS was evaluated and re-designed without increasing the total number of EMS units. Since the resulting EMS system showed improved performance, then this thesis has contributed to knowledge by confirming the general results mentioned above.

3. Other researchers have integrated a number of different methods in order to successfully solve the multi-facility location allocation problem. This thesis has
contributed to this knowledge by further confirming the usefulness of an integrated approach. In addition, the integration of fuzzy k-means with GIS modeling and simulation is an innovative approach that has been demonstrated to be useful in this thesis.

4. Very little research on location allocation has incorporated the versatility of data mining methods and techniques. Application of data mining to location allocation research is in itself a novel contribution to knowledge.

5. This thesis has argued that with effective and efficient planning of EMS systems, it is possible to achieve better EMS performance. This argument has been successfully demonstrated by exploring the effects of different, methods, models, tools and techniques that can be used to strategically determine an optimal location and allocation of EMS units.

10.4. Limitations and Recommendations

Based on this research and analysis done for EMS system in state of Qatar, it is advised to improve the current reporting system as it has some limitations in terms of capturing data accurately for future use. First, it was noticed that there is no consistency when reporting the location of the incidents. Location of incidents is reported by mentioning the zone number and what is called “location site”. It was also found that there are many spelling mistakes which make it difficult to confidently ascertain location of a site.

Location site reporting was not synchronized with other geographical system in other organization such as Ministry of Municipality and Urban Planning. Thus, any data collected by the ambulance service will not properly match the ones in other organizations and thus leading to a number of inaccuracies.

The location of incidents is given by zone number which is a very large scale that will not help in identifying the actual location of the incidents and may include some areas that are not populated or where there are no incidents. Therefore, it is advisable first to
standardize the way of reporting among the ambulance service office and it is even recommended to synchronize their system with other organizations such as Ministry of Municipality and Urban Planning. Moreover, it is advised to scale down the reporting to be specific to the districts or even to streets. It is even recommended to use a coordinate system to specify exactly the location of incidents. This will help in having accurate studies and analysis of EMS systems.

It was planned also to do this study based on the ages of the patients and its relation to the type of the incident and location as there was a thought that there is a relation between the age, the incident type and even the location. However, it is observed that the incidents data provided was not classified according to the age of the patients. It is recommended to include the age of the patient as one of the important categories when reporting the incidents. By studying this category, analysts may find a relation between incidents types and ages and incidents location and ages as well and this may lead to more accurate and improved location of EMS units.

It was observed that the existing system was formed based only on incidents location. It is recommended to include other parameter such as the incidents type and/or priority as some incidents are not critical such as transport ones. Moreover, ages also considered as important information. All of these parameters may lead to find better EMS locations. For example, it may be found that life threatening incidents are concentrated in specific locations and as known this kind of incidents need to be served shortly within few minutes (not exceeding 8 minutes). Thus, this will inform the analyst that at these black spots more EMS units or closer EMS units need to be located.

It was observed that there is no enough information about the canceled incidents. Thus, the EMS service in Qatar does not provide information like when it is canceled; after dispatching or before? And if it is after, at what time? Moreover, no information was provided on whether the EMS unit has received the location or not. All of this information is important as we need to know when the EMS unit was busy or not utilized
It is also recommend to include how many incidents each EMS units served daily or even hourly. This will help in finding which EMS unit is not utilized well. In addition to that, it is also advised to study the number of incidents and even the type of these incidents with respect to time and improve the system based on that as it is shown earlier there is a relation between the number of incident and time of the day.

10.5. Future Work

In this research incidents and population distributions were the main parameters used to find the new locations of EMS units. These parameters were used separately. Future studies could combine between both parameters in such a way as to present their weights as they are important factors for determining the best location of EMS units.

Future studies could include the type of the incidents and its distribution all around the country. This important parameter may help in finding the black spots where there are always crucial cases that need to be served quickly and may result in a better location of EMS units. Also, including the ages of the patients could have value since it could be found that specific range of ages face specific type of incidents and relate that to geographical areas.

Future work could also consider the stochastic issues related to EMS units. Thus, it could take into consideration that at specific time, the required EMS unit could be busy. As such, the study may contain some probability or includes the concepts of pack-up. This will help in improving the response time of the EMS services. Furthermore, studies could also include the time pattern while meeting the coverage requirements as it is found that at specific time of the day, the number of incidents increased and decreased in others. This requires adding more resources at the peak time which accordingly will result in improving the response time.
Last but not least, more focus need to be given to rural areas and the issue of equity. This parameter is important as there is a need to deal with incidents equally regardless of its place. This is because we are dealing with human life and this life is our priority regardless of where it is. Therefore, more studies and models need to be implemented on this regard.
References


Appendix A – Population Data
Appendix B – Incidents Data
Appendix C – Results of Service Area Analysis
Appendix D – Normality Test
Appendix E – EMS Hourly Schedule
Appendix F – Matlab Code
Appendix G – Matlab Results
Appendix H – Validation Code