STOCHASTIC EMERGENCY MEDICAL SERVICE VEHICLE LOCATION PROBLEM: EQUITY, PERFORMANCE EVALUATION AND MATHEMATICAL MODELS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUHARREM ALTAN AKDOĞAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
INDUSTRIAL ENGINEERING

JANUARY 2023

Approval of the thesis:

STOCHASTIC EMERGENCY MEDICAL SERVICE VEHICLE LOCATION PROBLEM: EQUITY, PERFORMANCE EVALUATION AND MATHEMATICAL MODELS

submitted by **MUHARREM ALTAN AKDOĞAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** __________

Prof. Dr. Esra Karasakal
Head of Department, **Industrial Engineering** __________

Prof. Dr. Z. Pelin Bayındır
Supervisor, **Industrial Engineering, METU** __________

Prof. Dr. Cem İyigün
Co-supervisor, **Industrial Engineering, METU** __________

Examining Committee Members:

Prof. Dr. Yasemin Serin
Industrial Engineering, METU __________

Prof. Dr. Z. Pelin Bayındır
Industrial Engineering, METU __________

Prof. Dr. Onur Kaya
Industrial Engineering, Eskişehir Technical University __________

Assoc. Prof. Dr. A. Selin Kocaman
Industrial Engineering, Bilkent University __________

Assist. Prof. Dr. Sakine Batun
Industrial Engineering, METU __________

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Muharrem Altan Akdoğan

Signature :

ABSTRACT

STOCHASTIC EMERGENCY MEDICAL SERVICE VEHICLE LOCATION PROBLEM: EQUITY, PERFORMANCE EVALUATION AND MATHEMATICAL MODELS

Akdoğan, Muharrem Altan
Ph.D., Department of Industrial Engineering
Supervisor: Prof. Dr. Z. Pelin Bayındır
Co-Supervisor: Prof. Dr. Cem İyigün

January 2023, 174 pages

In this thesis, emergency medical service (EMS) vehicle location problem with uncertainties in demand, travel times, and incident handling time is studied in three layers. The performance measures of EMS systems are evaluated with discrete event simulation models due to the uncertainties incorporated. Firstly, we focus on the equity in service quality resulting from vehicle location decisions in emergency medical services. We address the unbalanced service quality among regions with respect to various mathematical models including conventional ones. An extensive numerical study is conducted to show the effect of modeling approaches and network features on equity. Several observations are drawn and it is shown that the use of overall performance measures in the objective functions of mathematical models ignores other essential criteria, and there is room for improvement in terms of equity. Secondly, we propose decomposition methods based on queueing theory to assess the performance measures of the EMS system under a given location solution without needing to construct computationally cumbersome queueing or simulation models. The proposed decomposition methods require a set of nonlinear equations to be solved simultaneously, yet they are still favorable to exact queueing or simulation models in terms of the computational burden. A genetic algorithm is proposed to use the decomposition method in the optimization problems and find solutions for the EMS vehicle location problem based on queueing theory. It is shown that the proposed methods perform well in assessing the performance measures and evaluating close enough solutions the best solution for the optimization problems. Lastly, we propose mixed integer nonlinear problems (MINLP) with various objective functions incorporating closed-form formulations for the performance measures of the system based on decision variables. With MINLP models, there is no need for estimating problem parameters such as busy probability of vehicles in advance. The proposed models are easier to construct and solve with respect to approximation algorithms or decomposition methods in the literature where stochastic processes are incorporated. Hence, the proposed MINLP models enable decision-makers to incorporate uncertainties in the problem environment directly in the estimation of the parameter of the models based on queueing theory while still keeping the models relatively easy to solve since only the objective functions are nonlinear.

Keywords: emergency medical services, ambulance location, simulation, stochastic processes

ÖZ

STOKASTİK ACİL TIBBİ HİZMET ARACI KONUM PROBLEMİ: EŞİTLİK, PERFORMANS DEĞERLENDİRME VE MATEMATİKSEL MODELLER

Akdoğan, Muharrem Altan
Doktora, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Z. Pelin Bayındır
Ortak Tez Yöneticisi: Prof. Dr. Cem İyigün

Ocak 2023 , 174 sayfa

Bu tezde, talep, seyahat süreleri ve yerinde müdahale süresinde belirsizlikler içeren acil tıbbi hizmet araç lokasyonu problemi üç katmanda incelenmiştir. Acil tıbbi hizmet sistemlerinin performans ölçümleri, dahil edilen belirsizlikler nedeniyle ayrık olay simülasyon modelleri ile değerlendirilmiştir. İlk olarak, acil tıbbi hizmetlerde araç yeri kararından kaynaklanan hizmet kalitesindeki eşitlik üzerine odaklanılmıştır. Bölgeler arasındaki dengesiz hizmet kalitesi, geleneksel modeller de dahil olmak üzere çeşitli matematiksel modellerle ele alınmıştır. Modelleme yaklaşımlarının ve ağ özelliklerinin eşitlik üzerindeki etkisini göstermek için kapsamlı bir sayısal çalışma yapılmıştır. Çeşitli gözlemler çıkarılmış ve matematiksel modellerin amaç fonksiyonlarında genel performans ölçümlerinin kullanılmasının diğer temel kriterleri göz ardı ettiği ve eşitlik açısından iyileştirmenin mümkün gösterilmiştir. İkinci olarak, hesaplama açısından kuyruk veya simülasyon modelleri oluşturmaya gerek kalmadan belirli bir lokasyon çözümü altında acil tıbbi hizmet sisteminin performans ölçümlerini değerlendirmek için kuyruk teorisine dayalı ayrıştırma yöntemleri öneriyoruz. Önerilen ayrıştırma yöntemleri, bir dizi doğrusal olmayan denklemin aynı anda çözülmesini gerektirir, ancak bunlar performansalarına bakılarak yine de kuyruk veya simülasyon modellerine tercih edilebilir. Ayrıştırma yöntemini optimizasyon problemlerinde kullanmak ve kuyruk teorisine dayalı EMS araç konum problemine çözüm bulmak için bir genetik algoritma önerilmiştir. Önerilen yöntemlerin, performans ölçülerini kestirmede ve optimizasyon problemlerinin en iyi çözümüne yeterince yakın çözümleri bulmada iyi performans gösterdiği gösterilmiştir. Son olarak, karar değişkenlerine dayalı sistemin performans ölçümleri için kapalı form formülasyonları içeren çeşitli amaç fonksiyonlarına sahip karma tamsayılı doğrusal olmayan problemler (MINLP) öneriyoruz. MINLP modelleri ile araçların meşgul olma olasılığı gibi problem parametrelerinin önceden tahmin edilmesine gerek yoktur. Önerilen modellerin oluşturulması ve çözülmesi, stokastik süreçlerin dahil edildiği literatürdeki yaklaşım algoritmaları veya ayrıştırma yöntemlerine göre daha kolaydır. Bu nedenle, önerilen MINLP modelleri, karar vericilerin problem ortamındaki belirsizlikleri doğrudan kuyruk teorisine dayalı modellerin parametre tahmininde hesaba katmaya olanak tanırken, yalnızca amaç fonksiyonları doğrusal olmadığı için modelleri çözmek nispeten kolay olmaktadır.

Anahtar Kelimeler: acil tıbbi yardım servisi, ambulans lokasyonu, benzetim, stokastik süreçler

To my dearest, Deniz

ACKNOWLEDGMENTS

This thesis was made possible not only by my efforts but also by the assistance of many others along the way. I want to thank them for making this study possible.

First and foremost, I would like to thank my advisors, Dr. Pelin Bayındır for guiding me throughout my studies, always pointing me in the right direction and not sparing me from her intellect, and Dr. Cem İyigün for always believing in me, helping me write various pieces with his exceptional writing and his motivational speeches that tremendously helped me complete this study.

I want to thank the members of the examining committee, Dr. Yasemin Serin and Dr. Onur Kaya, for their comments and guidance in developing this thesis. I thank Dr. Selin Kocaman and Dr. Sakine Batun for their time reviewing this work.

I thank my family, my mom, Feride; my dad, Rıza; and my little sister, Aydan for being the happy and whole family we are. I cannot thank you enough for your very existence and your love.

And, the long list of friends and colleagues who helped me during difficult times and with whom we enjoyed the happiest of times together; I thank Ezgi Evrim Özkol for her kindness, hospitality, and ability in sugar coating the most disturbing so that I can cope with at times, Gökhan Sarı for making me feel confident and competent at everything I try to accomplish and introducing me to the world of startups and business almost as an inconvenience for my progress in pursuing this degree, my office neighbor Melis Özates Gürbüz for always making the time for me to discuss what needs to be discussed and to have a cup of afternoon coffee with snacks, my colleagues and friends Sena Önen Öz, Dilay Aktaş for always having a seat in their office to have a chat about work, life, and my studies and for their refreshing, uplifting and of course entertaining attitude on things, Cansu Alakuş, Can Barış Çetin, Burak Öz and Yeti Ziya Gürbüz for making the funniest jokes and being there at times needed the most, Tuna Berk Kaya, Yağmur Caner, Dilay Özkan and

Muhammed Ceyhan Şahin for having discussions on various topics up to the point of ridiculousness and being great colleagues, and Elçin Ergin for her overseas support and friendship over the course of the years.

Dearest of all, my other half, Deniz Coşkun, deserves the highest of acknowledgments. Getting this degree has been a roller coaster, both emotionally and physically at times when I needed to travel back and forth between cities. Deniz, you have been overly patient, have always supported me, and have stood by me until I get this done. As we close one chapter, our daughter, Ada, has joined our family, making us parents. Deniz, I cannot thank you enough. You are dear to my heart.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Emergency medical service (EMS) is one of the most critical daily services, among others, for public service providers. Emergency services, in general, need planning at every level, from operational to strategic.

EMS vehicle location has been an important topic for decision-makers and a critical research area for researchers. Management of these systems is mostly regulated by governmental bodies which oblige operators of the systems to sustain a system satisfying specific criteria. One of the most prominent criteria among those is the response time of these systems. For instance, National Health Service England (NHS-England (2015)), one of the four national service institutions of the United Kingdom, enforces 90% of life-threatening incidents to be responded to within 15 minutes. Some studies focus on other important measures in the literature, such as the fraction of lost demand, coverage, or costs related to strategic or operational activities. Aringhieri et al. (2017), Bélanger et al. (2019) and Wang et al. (2021) present comprehensive reviews of EMS location planning literature.

Emergency medical service operations planning is an essential topic for governmental authorities. It requires several parties to work together, such as call centers, operating companies, hospitals. Emergency response is a critical part of emergency service and requires planning personnel, equipment, vehicles, and physical infrastructure.

There are various approaches in the planning of the emergency medical service (EMS) vehicle stations, from deterministic models to simulation studies in the literature. Vehicle location problems could be handled in a deterministic problem setting where demand and service are taken as static entities. This approach could disregard the overflow of emergency calls due to installed capacity, i.e., vehicles, or the uncertainty in demand calls, service times, or travel times. Regarding those concerns, a stream of studies in the literature focuses on modeling approaches incorporating the uncertainty in the environment.

1.1 Problem Definition and Environment

In this thesis, we study EMS vehicle location problem to locate predefined number of vehicles at candidate locations where the problem environment includes uncertainties in demand calls, travel time between vehicle locations and demand regions, and incident handling times at scene.

A geographic area where EMS vehicles are to be located is divided into demand regions, indexed by $j \in J$. Candidate vehicle locations are formed as a subset of demand regions, $I \subset J$. The number of vehicles, N, is predetermined by the decision-maker where all vehicles are identical.

A call for medical emergency service is called *demand*. We assume that the demand calls occur at the center of each demand region and follow a time-homogeneous Poisson Process. It is assumed that demand across regions is independent. The mean number of calls in a unit time in regions, denoted by λ_j for region j, are known and are allowed to be non-identical. When a call is received, the closest available vehicle is assigned to that call, and the service starts. If there is no available vehicle in the system at the time of the call, the demand is assumed to be lost. Demand calls are not queued and assumed to be handled by external systems, similar to a stream of studies in EMS vehicle location literature.

A service is composed of three modes: (i) traveling to the demand region, (ii) incident handling, and (iii) returning to the vehicle location. The incident handling time is assumed to be exponentially distributed with known mean ϕ_j for region j. Travel time is considered a significant component of the service. Travel time from a vehicle location $i \in I$ to a demand region j, T_{ij} is assumed to follow Exponential distribution with a known mean, ω_{ij} and it is symmetrical for the mode of traveling back. Although the variance in incident handling and travel times is high due to the assumption of Exponential distribution, it is justified since loss systems where demand calls are not queued is said to be relatively insensitive to the form of the service distribution by Jarvis (1975). Service time is random and as a result of being dependent on the demand and vehicle locations, it is the sum of three random variables: (i) travel time to the demand region, (ii) incident handling time, and (iii) travel time for returning to the vehicle location. Incident handling times are assumed specific to demand regions. It is assumed that it includes the travel time within the region if the region is served by a vehicle located in this region. Travel times are taken explicitly for the case of serving a region by a vehicle located in another region.

In this problem environment, performance measures of a candidate location solution, such as mean response time, could be evaluated by using queueing theory. A queuing model for the system resulting from a candidate location solution can be constructed as follows. We refer to this model as the exact queueing model in the rest of the thesis.

Demand calls arrive according to a time homogeneous Possion process with known means λ_j from region $j \in J$. Incident handling and travel times are exponentially distributed with known means ϕ_j for region j, and ω_{ij} from vehicle location i to demand region j, respectively. Since the service time for a demand call includes travel times which are distributed exponentially with known mean, the state definition for the exact queueing model should represent the location of each vehicle, the state of the vehicle (available or busy), the region it serves if the vehicle is busy and the mode of the service (en-route to the demand region, serving the demand, en-route to the vehicle location).

Under these assumptions, let $\{B_t, t \geq 0\}$ be a continuous time Markov chain with state space L.

Then, this queueing model can be represented by an N-dimensional state as follows:

$$B_t = (b_1, b_2, \ldots, b_N), \qquad t \geq 0$$

where b_k represents the status of the k^{th} vehicle at time t and denotes a 3-tuple as:

$$b_k = (i_k, s, m)\,,\ s \in J \cup \{0\}\,,\ m = 0, 1, 2, 3$$

where i_k represents the location that the k^{th} vehicle is located, s stands for "idle" or the region being served, and m represents the mode of the service (en-route to the demand region, serving the demand, en-route to the vehicle location).

If the vehicle is not busy, s and m are set to 0. Then, $b_k = (i_k, 0, 0)$ represents the state that the vehicle at location i_k is free. If the vehicle is busy serving demand region $j \in J$, s is set to j. When the vehicle is busy, m is set to $1, 2$ or 3 if the vehicle is en-route to demand region j, handling the incident or en-route to the vehicle location, respectively. Then, $b_k = (i_k, j, 1)$ represents the state where the vehicle at location i_k is busy with responding to the demand from region $j \in J$ and traveling to the demand region.

Let l_k be the set of possible statuses of b_k of the state B_t in the exact queueing model. Then, l_k is as follows with $(3\,|J| + 1)$ possible statuses:

$$\begin{aligned} l_k = \Big\{ & (i_k, 0, 0), (i_k, 1, 1), (i_k, 1, 2), (i_k, 1, 3), (i_k, 2, 1), (i_k, 2, 2), (i_k, 2, 3), \\ & \ldots,\ (i_k, |J|, 1), (i_k, |J|, 2), (i_k, |J|, 3) \Big\}, \qquad k = 1, ..., N \end{aligned}$$

where i_k is the location that the k^{th} vehicle is located.

Then, the state space L for the chain $\{B_t, t \geq 0\}$ is as follows:

$$L = \{(b_1, b_2, \ldots, b_N) | b_1 \in l_1, b_2 \in l_2, ..., b_N \in l_N\}.$$

Hence, the state definition results in a state space with $(3\,|J| + 1)^N$ states, and the size of the state space increases exponentially with the increasing number of vehicles. Due to this complexity and computational burden, the exact queueing model is not used to evaluate performance measures in this thesis. In order to account for the uncertainties, a discrete event simulation model is constructed to evaluate the performance measures of the system resulting from a candidate solution for the EMS system and proposed models are compared to the simulation of the exact system in the rest of the study.

In the next section, we briefly explain the research conducted on EMS vehicle location problem within the scope of this thesis.

1.2 Proposed Methods and Models

We study the EMS vehicle location problem in three layers. First, we focus on equity aspects resulting from the location decision of vehicles in emergency medical systems. Then, we propose decomposition methods based on queueing theory to assess the performance measures of the EMS system under a given location solution. Lastly, we propose mixed integer nonlinear problems (MINLP) with various objective functions incorporating closed form formulations for the performance measures of the system based on decision variables.

EMS systems provide service to the public under certain criteria to be met by service providers. Those criteria are mostly on response time to incidents where a certain threshold to be satisfied in at least a predetermined fraction of the incidents similar to the threshold by NHS-England (2015). However, the differences among regions in quality of service could be overlooked due to use of overall measures as the one in "90% of life-threatening incidents" . Hence, the equity in EMS systems is an important topic that needs to be discussed in EMS vehicle/facility location problems explicitly. However, the literature on equity in EMS systems are scarce as it is later discussed in Chapter 2 in more detail. Therefore, we study the EMS vehicle location problem from an equity perspective in Chapter 3. We use several location models with varying objective functions and/or constraints to show the effect of models on equity. An extensive computational study is conducted by using a discrete event simulation model and a ranking selection algorithm to find optimal solutions to the problem based on simulation study. The problem instances for the computational study are constructed by changing the network features such as distribution of demand regions over the plain, number of vehicles to be located and incident handling rate. In Chapter 3, we discuss the equity in emergency medical services by addressing the unbalanced service quality among regions due to limited resources. An extensive computational study shows the effect of modeling approach or network features on equity in emergency care.

The problem environment incorporates stochastic processes which enables us to use queueing theory to assess the performance of an EMS system. However, the size of the exact queueing model that needs to be constructed increases exponentially in the number of vehicles and regions. Therefore, we rely on simulation study for the experimental study in Chapter 3. Although we use a ranking and selection algorithm to select best solutions among feasible solutions for the problem instances, the assessment of the performance of the feasible solutions still requires significant computation power and time. Therefore, we study on decomposition methods to assess the performance of an EMS system based on queueing theory without requiring to construct the exact queueing model in Chapter 4. The exact queueing model is decomposed into interdependent queueing models in an effort to approximate the performance measures of the EMS system under the exact queueing model. The interdependence between queueing models in the decomposition methods results in a set of nonlinear equations that needs to be solved simultaneously in order to assess the steady state probabilities. In Chapter 4, a solution approach is proposed for the nonlinear equations and, the performance measures of the exact queueing model are estimated based on the steady state probabilities of the interdependent queueing models. An extensive computational study is conducted to show the performance of the decomposition method in estimating the performance measures of the EMS system, including real-life problem instances. In addition, an evolutionary meta-heuristic algorithm is proposed to find solutions to the optimization problems using the decomposition methods. The meta-heuristic algorithm uses the decomposition method to assess the objective function value of feasible solutions and finds a best solution for the optimization problem based on evolution. Hence, we propose a method to assess the performance measure of the EMS system and a genetic algorithm to find solutions for the optimization problems based on queueing theory for the EMS vehicle location problem in Chapter 4.

As it is explained above, the estimation of the performance measures of an EMS system relies on the queueing models in Chapter 4. EMS system is decomposed into interdependent queueing models, and performance measures of the EMS system are estimated based on the steady-state distributions of those interdependent models. The decomposition methods proposed requires algorithmic solution approaches in approximating the steady-state distribution which is the solution to a nonlinear set of equations. Therefore, the decomposition methods are computationally burdensome in comparison to closed-form mathematical models for which commercial solvers could be easily exploited. Therefore, approximate closed-form formulations for the performance measures of the EMS systems are studied in Chapter 5. A similar approach to Chapter 4 is taken and EMS system are represented as separate independent queueing models. The major difference is the assumption of independence among queueing models in Chapter 5. This assumption results in a linear set of equations which needs to solved to assess the steady-state distribution. Hence, the steady state distributions could now be expressed as a function decision variables in the constraints of a mixed integer program. Based on this structure, several objective functions are constructed based on the decision variables in nonlinear form. Thus, several mixed integer nonlinear programming models are proposed for EMS vehicle location problem in Chapter 5. In the computational study, those models are evaluated with commercial solvers and the results are compared with the optimal solutions found by complete enumeration. The performance of models are compared with the simulation-based ranking selection algorithm, proposed decomposition methods and a widely-known probabilistic model for EMS location problem, maximum expected coverage location problem by Daskin (1983).

1.3 Contributions and Novelties

The contribution of this thesis can be summarized as follows.

In Chapter 3, we lay the ground for equity discussions in the EMS vehicle location problem by addressing unbalanced service quality among demand regions resulting from the limited resources. We work on various modeling approaches to close the gap between existing research and fairness in emergency care from a Rawlsian perspective. We bring conventional models into the equity discussion as well as new ones in a stochastic problem environment. An extensive numerical study is conducted focusing on network features that could affect the performance of different approaches. By revealing the trade-off between overall performance and equity with the help of a structured experimental study, we show that models based on overall performance measures ignore other essential criteria, and conclude that there is a room for improvement in terms of equity.

In Chapter 4, decomposition methods are proposed to approximate the performance measures of exact system without needing to construct computationally cumbersome queueing models. For the proposed models, it is shown that there exists at least one solution to the resulting set of nonlinear equations and the performance measures could be estimated based on that solution. Differently from the studies in the literature, the vehicles at the same location are prioritized so that the effect of an additional vehicle at that location on the existing vehicle's performance could be observed. An easy-to-apply meta-heuristic algorithm is proposed to use the decomposition method in the optimization problems and find solutions for the EMS vehicle location problem based on queueing theory.

In Chapter 5, closed form formulations for performance measures are derived based on separate queueing models for the vehicles. Those formulations are used in the mixed integer nonlinear problems (MINLP) for which commercial solvers could be used to find optimal solutions. Differently from studies in the literature, there is no need for estimating the problem parameters like busy probability of vehicles in advance.The proposed models are easier to construct and solve with respect to approximation algorithms or decomposition methods in the literature where stochastic processes are included in the problem definition. Hence, the proposed MINLP models enable decision makers incorporate uncertainties in the problem environment directly in the estimation of the parameter of the models based on queueing theory while still keeping the models relatively easy to solve since only the objective functions are nonlinear.

1.4 The Outline of the Thesis

The outline of the thesis is as follows. In Chapter 2, a literature review on emergency vehicle location problems is presented including widely known deterministic models, probabilistic models and models using queueing theory to assess performance of such systems. A separate section is dedicated to the studies focusing on equity in EMS location problems. In Chapter 3, equity in EMS vehicle location problem is discussed and a comprehensive computational study is presented. In Chapter 4, we present decomposition methods to evaluate performance measures of EMS system. In Section

5, we present MINLP models for EMS vehicle location problem. In Section 6, the thesis is concluded with a summary of findings and contributions.

CHAPTER 2

LITERATURE REVIEW

In this chapter, a literature review is presented focusing on emergency facility/vehicle location problems.

Facility location problems could be separated into three categories as *p-median*, *p-center* and *covering* problems. The building blocks of *p-center* and *p-median* problems are first proposed by Hakimi (1964) to find optimum locations of switching centers, which later on extended in various studies for the EMS facility/vehicle location problem. In *p-center* problem, the maximum distance to a facility is minimized whereas a weighted measure is minimized in *p-median* problem. The first *covering* problem proposed for locating emergency facilities is the location set covering problem (LSCP) by Toregas et al. (1971).

Both deterministic and probabilistic models are used to locate service vehicles such as ambulances and corresponding facilities. We first present very well-known deterministic models used in emergency facility location problem and continue with probabilistic models. Lastly, the studies in the emergency facility/vehicle location problem focusing on equity are presented.

2.1 Deterministic models

As mentioned earlier, Toregas et al. (1971) is the first to formulate LSCP for emergency facilities. The coverage defined in terms of a maximal service distance where nodes are assumed to be covered if there are at least one facility within the maximal service distances. The objective of LSCP is to minimize the number of facilities given all nodes are covered. Let x_j be the binary decision variable which is equal to 1 if a facility is located at node j and 0; otherwise. d_{ji} is the distance from node j to node i and N_i is the set of nodes within s distance of node i, $N_i = \{j = 1, \ldots, n | d_{ji} \leq s\}$. Then, LSCP is constructed as follows:

$$\text{Minimize} \quad \sum_{j=1}^{n} x_j \tag{2.1}$$

$$\text{subject to:} \quad \sum_{j \in N_i} x_j \geq 1, \qquad \forall i = 1, \ldots, n, \tag{2.2}$$

$$x_j \in \{0, 1\}, \qquad \forall j = 1, \ldots, n. \tag{2.3}$$

Church and ReVelle (1974) propose maximal location covering model (MCLP) that is to maximize the number of regions covered where coverage of a region is assessed based on the existence of a facility within a travel distance under a threshold. Let I be the set of demand nodes, J be the set of facility sites, N_i be set of facility sites that can cover demand node i under threshold travel distance s, a_i be the population to be served at demand node i and P be the number of facilities to be located. Let x_j be the binary decision variable which is equal to 1 if a facility is located at node j and 0 and y_i be the binary variable which is equal to 1 if demand node i is covered by at least one facility. According to the parameters and decision variables, MCLP is written as follows:

$$\text{Maximize} \quad \sum_{i \in I} a_i y_i \tag{2.4}$$

$$\text{subject to:} \quad \sum_{j \in N_i} x_j \geq y_i, \qquad \forall i \in I, \tag{2.5}$$

$$\sum_{j \in J} x_j = P, \tag{2.6}$$

$$x_j \in \{0, 1\}, \qquad \forall j \in J, \tag{2.7}$$

$$y_i \in \{0, 1\}, \qquad \forall i \in I. \tag{2.8}$$

Daskin and Stern (1981) is the first to introduce multiple coverage of the demand regions into modeling with hierarchical objective set covering model (HOSC). In HOSC, the number of vehicles located is minimized while maximizing the number of times that demand regions are covered. Later, Hogan and ReVelle (1986) propose backup coverage problem, BACOP 1 which maximizes the population covered twice and BACOP 2 which maximizes the weighted average of population covered once and covered twice given the number of vehicles. While previous multiple coverage models use a single threshold for coverage, Gendreau et al. (1997) propose Double Standard Model (DSM) and use double coverage and varying coverage thresholds.

Berman and Krass (2002) extend MCLP and propose generalized maximal location covering model by allowing partial coverage based on distances. Drezner et al. (2004) propose gradual covering problem based on a linear coverage function. Karasakal and Karasakal (2004) extend MCLP and propose MCLP-P in the presence of partial coverage where the coverage function is allowed to be continous or discrete; linear or nonlinear. Differently from previous models, Berman et al. (2009) assume that every vehicle can cover a region at certain levels based on the distance and a region is covered only if the aggregate coverage provided by vehicles exceeds a certain threshold.

2.2 Probabilistic models

Probabilistic studies mainly focus on providing coverage with a probability which is based on a threshold response time or availability of the vehicles. Aly and White (1978) handle uncertainty that is related to problem inputs in coverage criterion where the response time for an emergency facility is a random variable and a demand is considered as covered if the response time is less than and equal to a threshold. Other studies focusing on the varying availability of the vehicles. Maximum expected covering location problem (MEXCLP) by Daskin (1983) incorporates busy probabilities of vehicles into the mathematical model. Let M be the number of facilities to be located, N be the number of nodes in the network, h_k be the demand generated at node k, d_{ik} be the distance between site i and node k, a_{ki} be a parameter which is equal to 1 if a vehicle at site i covers node k meaning $d_{ik} \leq D$ and 0 otherwise. p states the probability that a facility is not working. Let x_i be a decision variable which is equal to 1 if a vehicle is located at site i, and y_{jk} is a decision variable that is equal to 1 if node k is covered by at least j vehicles, 0 otherwise.

Then, MEXCLP as follows:

$$\text{Maximize} \quad \sum_{k=1}^{N}\sum_{j=1}^{M}(1-p)p^{j-1}h_k y_{jk} \tag{2.9}$$

$$\text{subject to:} \quad \sum_{j=1}^{M} y_{jk} - \sum_{i=1}^{N} a_{ki}x_i \leq 0, \qquad k = 1, ..., N \tag{2.10}$$

$$\sum_{i \in I} x_i \leq M \tag{2.11}$$

$$x_i \leq M, integer, \qquad i = 1, ..., M \tag{2.12}$$

$$y_{jk} \in \{0, 1\}, \qquad j = 1, ..., M, k = 1, ..., N. \tag{2.13}$$

Some deterministic models are extended as well in order to introduce the uncertainties in the system into the models. ReVelle and Hogan (1988) extend LSCP by introducing a reliability constraint and propose probabilistic location set covering problem (PLSCP). The reliability constraint relies on the estimation of the probability of a demand being covered by one the vehicles that are closer to the demand region than a certain threshold. For the constraint, the number of vehicles that needs to be located within the threshold radius of every region is calculated based on the estimation of the local buys fractions of vehicles and the required reliability level.

PLSCP is as follows:

$$\text{Minimize} \quad \sum_{j=1}^{n} x_j \tag{2.14}$$

$$\text{subject to:} \quad \sum_{j \in N_i} x_j \geq b_i, \qquad \forall i = 1, \ldots, n, \tag{2.15}$$

$$x_j \geq 0, integer \qquad \forall j = 1, \ldots, n. \tag{2.16}$$

where b_i is the smallest integer satisfying $1 - (F_i/b_i)^{b_i} \geq \alpha$. F_i is the daily fractional workload that is shared among $\sum_{j \in N_i} x_j$ many vehicles which can cover node i. Then, $(F_i/b_i)^{b_i}$ is the probability of all b_i vehicles covering node i being busy and $1 - (F_i/b_i)^{b_i}$ is the reliability level for node i given b_i.

Later, ReVelle and Hogan (1989) propose the maximum availability location problem (MALP) based on the same reliability structure of PLSCP. In MALP, the objective is to maximize the population covered with reliability α. The version of the MALP that uses local busy fractions is as follows:

$$\text{Maximize} \quad \sum_{i \in I} f_i y_{ib_i} \tag{2.17}$$

$$\text{subject to:} \quad \sum_{k=1}^{b_i} y_{ik} \leq \sum_{j \in N_i} x_j, \qquad \forall i \in I, \tag{2.18}$$

$$y_{ik} \leq y_{ik-1}, \qquad k = 2, ..., b_i, i \in I, \tag{2.19}$$

$$\sum_{j \in J} x_j = p \tag{2.20}$$

$$x_j \geq 0, integer, \qquad i = 1, ..., M \tag{2.21}$$

$$y_{ik} \in \{0, 1\}, \qquad k = 1, ..., b_i, i \in I. \tag{2.22}$$

where f_i the population at node i, y_{ik} is equal to 1 if there exists k servers covering node i and 0 otherwise, and b_i is the smallest integer satisfying $1 - (F_i/b_i)^{b_i} \geq \alpha$.

Later, several models are constructed based on the models MEXCLP and MALP. Ball and Lin (1993) propose a reliability model by imposing an upper limit for the probability that a demand call is not met while incorporating the randomness of demand calls in the calculation of the probability of failure, differently from MALP. Sorensen and Church (2010) propose local reliability-based maximum expected covering location problem (LR-MEXCLP) where the maximum expected coverage objective of MEXCLP and local reliability estimation of MALP are integrated. El Itani et al. (2019) propose a bi-objective model combining MEXCLP and MALP to maximize expected coverage while minimizing the cost of using external resources (ambulances) to increase coverage.

Another deterministic model that is extended in a way to incorporate uncertainties is DSM by Gendreau et al. (1997). Liu et al. (2016) propose a probabilistic DSM where every demand region is covered at east once within the secondary coverage threshold while ensuring some portions of the demand regions are covered with a given service reliability level in the first and secondary coverage threshold.

Differently from previous studies, Erkut et al. (2008) propose a new measure for the objective function. They construct a survivability measure based on travel times and propose maximal survival location problem where the expected number of patients who survive is maximized. All the models mentioned utilize objective functions that are representative of system-wide (overall) performance such as minimum average weighted distance, maximum number of covered location and maximum expected coverage.

Another stream of research in probabilistic studies uses Queuing Theory to assess the performance measures of emergency service systems. Queuing Theory is introduced in location models by Larson (1974) with the Hypercube Queuing Model (HQM). The model analyzes vehicle location-allocation and districting in emergency response units operated as server-to-customer services. With HQM, various performance measures such as travel times and work-loads are obtained from the steady-state probabilities of the system. Jarvis (1985) proposes an approximation algorithm based on HQM to find the performance of systems with distinguishable servers and general service time distributions.

Several studies use HQM to measure the performance of the systems such as Sacks and Grief (1994), Brandeau and Larson (1986), Iannoni and Morabito (2007), Takeda et al. (2007). HQM is extended for different service rates for each server by Mendonca and Morabito (2001) to assess the mean response time of the system. In the studies of Iannoni and Morabito (2007) and Takeda et al. (2007), service time variations resulting from the variations in travel times to the demand location are considered insignificant with respect to the sum of variations in set-up time, on-scene service time and travel time back to the vehicle location. Halpern (1977) states that the estimations for service times in the study of Mendonca and Morabito (2001), where variations in the travel times are considered to be of second-order, give questionable results where travel time is a significant part of the service time.

Saydam and Aytuğ (2003) use HQM in a genetic algorithm to find location-specific server busy probabilities and employ these probabilities in the estimation of expected coverage. Iannoni et al. (2008) use a genetic algorithm to find the locations for EMS servers, allowing only one server at a single location while using service rates specific to servers. Iannoni et al. (2009) and Iannoni et al. (2011) use HQM in an optimization environment for location and districting decisions of EMS servers on highways with alternative objectives. Geroliminis et al. (2009) extend HQM and develop a Spatial Queuing Model (SQM) by defining non-identical service times for servers that take into account the demand call's characteristics (inter-district or intra-district response). In a different study, Geroliminis et al. (2011) work on a large-scale system to deploy emergency response mobile units. They first divide the area into districts and find the optimal locations in these districts with the help of a genetic algorithm. Akdogan et al. (2018) extend SQM further differentiating service rates specific to vehicle location-demand region pairs and allowing multiple vehicles in a single location.

Approximation algorithms are studied in the location analysis of emergency service vehicles as well. Boyacı and Geroliminis (2015) propose approximation algorithms for large-scale networks with spatially distributed demand. They propose a partitioning algorithm that is used to find optimal server locations. Atkinson et al. (2008) propose ad-hoc heuristics to assess the probability of loss using algorithms based on the HQM. Budge et al. (2009) propose an algorithm to find the dispatch frequencies of vehicles. Differently from the previous studies, they allow locating multiple vehicles at a single location. Neither of the studies in Atkinson et al. (2008) and Budge et al. (2009) considers an optimization problem. Toro-Díaz et al. (2015) focus on reducing disparities between the mean response time of different regions while employing algorithmic approximation method for vehicle dispatch fractions by Budge et al. (2009) in an optimization problem.

Another stream of studies use queueing theory to estimate busy probabilities of emergency vehicles and utilize these in previously well studied models such as MEXCLP, PLSCP and MALP. Marianov and ReVelle (1994) extends the assumption of server independence in the PLSCP and propose Q-PLSCP. They use queueing theory in order to find the right hand side value in (2.15). In another study, Marianov and ReVelle (1996) extends MALP and propose queueing maximal availability location problem (Q-MALP) by again relaxing the server independence assumption and calculating the right hand side in (2.15) based on queueing models. Galvão et al. (2005) extends the classical MEXCLP and MALP by dropping the assumptions of server independence and common workload for servers. The authors embed HQM by Larson (1974) into the problems and propose EMEXCLP and EMALP for which they use local search methods to find solutions.

2.3 Studies on Equity in EMS Vehicle Location Problem

Although various studies consider different aspects of the EMS location problem, literature focusing on equity and fairness is scarce as it is also pointed out in two literature review studies by Li et al. (2011) and Aringhieri et al. (2017). On equity, no principle is commonly accepted. The studies incorporating equity have different emphases on specific issues. Brandeau and Larson (1986) use inequity in ambulance availability among regions as one of the primary performance measures in evaluating alternative location decisions. Drezner et al. (2009) use a model to locate facilities, not specific to EMS, minimizing the Gini coefficient as the equity measure. Mclay and Mayorga (2010) use priority ratings for emergency calls and propose a model that maximizes the number of high priority demand calls satisfying performance requirements and report disparities in patient survival between urban and rural areas according to changing response time thresholds. Chanta et al. (2014) propose bi-objective models to reduce the disparity in covered demand between urban and rural areas. Chanta et al. (2011) develop a p-envy location problem, modeling customer dissatisfaction as a distance-based function and minimizing overall envy. Toro-Díaz et al. (2015) focus on reducing disparities among the mean response time of different regions while employing the algorithmic approximation method for vehicle dispatch fractions proposed by Budge et al. (2009). Along with Iannoni et al. (2008), Iannoni et al. (2009) and Iannoni et al. (2011), the study of Toro-Díaz et al. (2015) is one of the few studies that uses the Queuing Theory to obtain performance measures while considering equity. Khodaparasti et al. (2016) propose a bi-objective model to locate EMS facilities while maximizing the efficiency and equity of the system where minimizing the total number of uncovered demand zones is used as the equity criterion. Although regulations impose restrictive performance requirements for these systems, studies in the literature are scarce addressing equity in a stochastic environment. Our study aims to explore some conventional models in relation to equity systematically under different network features.

CHAPTER 3

AN ANALYSIS OF EQUITY IN STOCHASTIC EMS VEHICLE LOCATION PROBLEM UNDER VARIOUS LOCATION MODELS AND NETWORK FEATURES

This chapter focuses on planning emergency medical service vehicle locations as a strategic level problem from an equity perspective by examining important performance criteria. As it is mentioned in Chapter 2, the studies focusing on equity in this context is scarce although emergency medical service is one of the most crucial services within the public health system. EMS operations are regulated by governmental bodies as in the case of NHS England that (NHS-England (2015)) uses priority categories for the emergencies with strict target performance to be achieved by the operators such as 90% of life-threatening incidents to be responded to within 15 minutes. However, the differences of service quality among regions (which could be defined by neighborhoods or political divisions) under limited resources remains an important topic. In this chapter of the thesis, we study several conventional location models that differ in the performance measures used in the objective function or constraints for the emergency vehicle location problem in an effort to show the effect of modeling approaches and network features on equity. The aim of this chapter is not to propose models but rather assess the performance of some conventional models in terms of equity.

To discuss equity, one first needs to define how it is conceived. In the literature, various definitions for equity are employed. Bertsimas et al. (2011) identify three alternative theories for social equity: Aristotelian equity, classical utilitarianism, and Rawlsian equity. *Aristotelian equity* dictates the proportional allocation of resources according to some pre-existing claims or rights of each party. This creates a problem regarding the determination of these claims or rights in society. Another theory is *classical utilitarianism* which was widely influential in economics during the 19th century. Allocation of resources is realized in a way to maximize the sum of utilities of individual parties. The third theory is proposed by Rawls (1971) based on political philosophy. *Rawlsian equity* gives priority to the least well-off parties to guarantee the highest minimum utility level. The drawback of this approach is to impose inconvenience on almost all parties. In this chapter, we approach equity from a Rawlsian perspective and use measures to assess the disparities among demand regions to respect individuals' equal rights in access to emergency care resulting from modeling approaches.

We study the emergency vehicle location problem where the problem environment involves uncertainty in incident handling times, travel times, and the occurrence of emergency service demand. We use models to benefit the least well-off demand regions in addition to models with widely used system-wide (overall) performance measures as mean response time and expected coverage in Section 3.2. Region-wise measures that may show the difference in service quality between demand regions are utilized to compare the models. We specifically use the variance of region-wise mean response time, the variance of region-wise lost demand, and the Gini coefficient by Gini (1912). Regarding the uncertainty in problem parameters, we use simulation optimization based on a fully sequential ranking algorithm, *KN++*, which is proposed by Kim and Nelson (2006). The relationship between equity and network features such as distribution of regions in the area, number of ambulances to be located, and incident handling rate is investigated in the experimental study.

This chapter is organized as follows. The problem environment is further explained in Section 3.1, and models under investigation are presented in Section 3.2. In Section 3.3, the details of the experimental study are given, results are analyzed from an equity perspective. Lastly, concluding remarks are laid out in Section 3.4.

3.1 Modeling Approach

We construct a mathematical model with only one set of decision variables, $x_i, \forall i \in I$, which indicates the number of vehicles located in vehicle location i. Let $\vec{x}$ be the vector of decision variables, x_i's, representing a solution. The objective function is defined as a function of $\vec{x}$ in the models. Different performance measures are used in the objective function in the next section, also with the addition of different constraints, forming alternative models to explore equity in EMS location problem.

Since the problem environment includes probabilistic aspects and the objective function includes the average performance measures in the steady-state, there is no closed-form formulation of the performance measures as a function of $\vec{x}$. In order to evaluate the performance measures of the EMS system resulting from a candidate solution $\vec{x}$, one needs to find the steady-state distribution of the EMS system and evaluate the objective function value based on those. Therefore, the objective function of the model is defined as a function of a candidate solution without introducing further notations.

We use one of the commonly used objective functions in the literature, *mean response time*, as a base model to compare with the models introduced later on. Mean response time is one of the most critical measures that is taken into account for the evaluation of emergency systems and it is based on the time passed between a demand call and arrival of a vehicle at the corresponding demand location.

Base model $\boldsymbol{P_1}$, that uses the mean response time of the system as the objective function, is defined as follows:

$$(\boldsymbol{P_1}) \qquad \text{Minimize} \qquad R(\vec{x}) \tag{3.1}$$

$$\text{subject to:} \qquad \sum_{i \in I} x_i = N \tag{3.2}$$

$$x_i \geq 0, integer, \qquad \forall i \in I. \tag{3.3}$$

where $R(\vec{x})$ is the mean response time of the system under given solution $\vec{x}$ and (3.2) enforces N vehicles to be located.

In the next section, we present location models that differ in objective function or constraints from the base model. In an experimental study, these models are examined to understand how the equity imposed by the models changes.

3.2 Location Models

Eight different location models for EMS vehicle location problem is presented in this section. These models are constructed considering some of the important criteria related to the performance of emergency systems.

3.2.1 Deterministic coverage constraint for a balanced service

Using a deterministic coverage constraint based on mean travel times is a way of incorporating equity consideration into models, which narrows down the solution space in the meantime. The vehicle locations that can cover a specific demand region are constrained by a time threshold. A demand region is assumed to be covered if there is a vehicle located within the threshold time, considering the mean travel time between the region and vehicle location. Let τ be the threshold travel time in minutes and ω_{ij} be the mean travel time in minutes between vehicle location i and demand region j. For every region j, we define a set W_j consisting of vehicle locations covering demand region j as follows:

$$W_j = \{i \in I : \omega_{ij} \leq \tau\}, \forall j \in J.$$

Let u_j be the demand fraction of region j in total demand and equals $\frac{\lambda_j}{\sum_{k \in J} \lambda_k}$. The model $\boldsymbol{P_2}$ below enforces that at least α fraction of total demand is covered by at least one vehicle.

$$
\begin{aligned}
(\boldsymbol{P_2}) \qquad \text{Minimize} \quad & R(\vec{x}) && && (3.4)\\
\text{subject to:} \quad & \sum_{i \in W_j} x_i \geq y_j, && \forall j \in J && (3.5)\\
& \sum_{j \in J} u_j y_j \geq \alpha && && (3.6)\\
& \sum_{i \in I} x_i = N && && (3.7)\\
& x_i \geq 0, integer, && \forall i \in I && (3.8)\\
& y_j \in \{0, 1\}, && \forall j \in J. && (3.9)
\end{aligned}
$$

A new decision variable is introduced to formulate the coverage criterion in $\boldsymbol{P_2}$. y_j is a binary decision variable taking the value of one if at least one vehicle is located in one of the vehicle locations in set W_j and zero otherwise in (3.5) for every j. (3.5) determines the regions covered in a solution and (3.6) enforces the threshold coverage where a feasible solution should cover a set of regions whose demand fractions, u_j, add up to at least α fraction of the total demand.

3.2.2 Chance constraints in assessing coverage

Notice that coverage criterion in $\boldsymbol{P_2}$ is utilized to define feasible solutions by means of constraints in the model and the coverage definition is based on mean travel time between regions. This approach is deterministic and it disregards the uncertainty in travel times since a certain region is assumed to be covered if there is a vehicle located in a location having mean travel time to the region less than threshold time τ. All demand of this region is assumed to be served within the threshold time, however this is a clear overestimation of the demand met within this threshold due the uncertainty in travel times. In addition, some demand calls from "not-covered" regions could be covered within threshold time based on realizations as well. It is also noted that, in $\boldsymbol{P_2}$, there is no consideration about the amount of demand that could be served from each vehicle location, i.e. the server capacity. The assumption that a vehicle covers the whole demand of the regions which it is the nearest cannot be justified without any further analysis of the problem environment.

There are regulations in place related to the concerns above, for example the one imposed by one of the four national health service institutions of the United Kingdom, NHS-England (2015). This regulation requires 90% of the life-threatening medical incidents to be responded within 15 minutes. One should make sure that such deterministic coverage constraints serve the purpose of this requirement and it leads to a more balanced system in terms of service quality, rather than just narrowing down the solution space.

Due to the uncertainties involved as well as the regulations, chance constraints are proposed to handle those concerns. Since travel times are random variables defined by known probability distribution functions, it is possible to suggest probabilistic constraints.

In the definition of the coverage constraint of $\boldsymbol{P_2}$, it is required to cover α fraction of the total demand under threshold time τ. Since the travel time from vehicle location i to region j is a random variable, we can bound the probability of serving time for a demand call from region j with a vehicle from location i being under threshold τ by α. For a vehicle location i and region j, the constraint is given as:

$$P(T_{ij} \leq \tau) \geq \alpha, \tag{3.10}$$

where T_{ij} is a random variable representing travel time between location i and region j.

Recall that T_{ij} is assumed to be exponentially distributed with a mean of ω_{ij}. Let $F_{T_{ij}}$ be the cumulative distribution function of random variable T_{ij}. Then, (3.10) can be written as

$$F_{T_{ij}}(\tau) \geq \alpha.$$

Hence, we can further break down (3.10) into

$$\tau \geq -\omega_{ij} \ln(1-\alpha),$$

where $F_{T_{ij}}(\tau) = 1 - e^{\frac{\tau}{\omega_{ij}}}$.

therefore, we can define a set V_j as the set of vehicle locations that could serve region j under threshold time τ with probability greater than or equal to α as

$$V_j = \left\{ i \in I : \omega_{ij} \leq -\frac{\tau}{\ln(1-\alpha)} \right\}, \forall j \in I.$$

Then, the following model, named as $\boldsymbol{P_3}$, is constructed using this set.

$$
\begin{aligned}
(\boldsymbol{P_3}) \qquad \text{Minimize} \quad & R(\vec{x}) && && (3.11)\\
\text{subject to:} \quad & \sum_{i \in V_j} x_i \geq z_j, && \forall j \in J && (3.12)\\
& \sum_{j \in J} z_j \geq \beta && && (3.13)\\
& \sum_{i \in I} x_i = N && && (3.14)\\
& x_i \geq 0, integer, && \forall i \in I && (3.15)\\
& z_j \in \{0, 1\}, && \forall j \in J. && (3.16)
\end{aligned}
$$

A new binary decision variable, z_j, is introduced in $\boldsymbol{P_3}$. By (3.12), z_j takes the value of one if at least one vehicle is located in one of the locations in V_j meaning there exist at least one location with a vehicle that could serve demand region j under threshold time τ with probability greater than or equal to α. It is zero, otherwise. (3.13) requires that at least β fraction of the demand regions should have access to a vehicle under threshold time τ with probability greater than or equal to α. Therefore, $\boldsymbol{P_3}$ accounts for the uncertainty in travel times with the help of these constraints, and it enforces a more realistic coverage criterion, differently from $\boldsymbol{P_2}$.

3.2.3 Minimizing worst region-wise mean response time

The quality of service for individual regions is as important as system-wide performance for emergency systems. The variation in mean response time among the regions could be high, resulting in unfair service quality. In addition to coverage constraints, region-wise mean response time could be another important measure to include in modeling EMS vehicle location problems. We construct the model $\boldsymbol{P_4}$ based on this proposition by minimizing the maximum of mean region-wise response times. Let $R_j(\vec{x})$ be the mean response time for region j under solution $\vec{x}$, $\boldsymbol{P_4}$ is given as follows:

$$(\boldsymbol{P_4}) \qquad \text{Minimize} \qquad \max_{j \in J}(R_j(\vec{x})) \tag{3.17}$$

$$\text{subject to:} \qquad \sum_{i \in I} x_i = N \tag{3.18}$$

$$x_i \geq 0, integer, \qquad \forall i \in I. \tag{3.19}$$

The objective function in (3.17) is of min-max type and it puts the emphasis on the least well-off region in line with Rawlsian equity.

3.2.4 Maximizing coverage

For a given location solution, another important measure is actual or realized coverage, i.e. the actual fraction of demand calls covered within threshold time τ. The actual coverage could be incorporated into the EMS vehicle location models as an objective function, rather than just as a constraint.

The model, $\boldsymbol{P_5}$, is constructed to maximize the realized coverage. A demand call is treated to be covered if the realized travel time to the demand region is shorter than threshold time τ. Let $C(\vec{x})$ be the expected fraction of total demand that is covered under threshold time τ for solution $\vec{x}$.

$\boldsymbol{P_5}$ is written as follows:

$$(\boldsymbol{P_5}) \qquad \text{Maximize} \qquad C(\vec{x}) \tag{3.20}$$

$$\text{subject to:} \qquad \sum_{i \in I} x_i = N \tag{3.21}$$

$$x_i \geq 0, integer, \qquad \forall i \in I. \tag{3.22}$$

$C(\vec{x})$ is calculated as the ratio of total number of covered demand calls based on realized values to the total number of calls generated. Since a demand call is assigned a vehicle if there is any available at the time of the call and assumed to be lost otherwise, $\boldsymbol{P_5}$ takes both the uncertainty in travel times and vehicle availability into account.

3.2.5 Positive deviation from average region-wise mean response time

While considering the least well-off region like in $\boldsymbol{P_4}$ is one way of enforcing equity, focusing on variations in region-wise measures is another one. According to a Rawlsian equity perspective, variations in region-wise measures are not desired. For this reason, we introduce an objective function that minimizes the total positive deviation of individual region-wise mean response times from the overall average region-wise mean response time (i.e. the sample mean). By this way, it is aimed to minimize the positive deviation and variance with a single metric.

Following the above discussion, a new model, $\boldsymbol{P_6}$, is given as:

$$(\boldsymbol{P_6}) \qquad \text{Minimize} \qquad \sum_{j \in J} \left[R_j(\vec{x}) - \frac{\sum_{k \in J} R_k(\vec{x})}{|J|} \right]_+ \tag{3.23}$$

$$\text{subject to:} \qquad \sum_{i \in I} x_i = N \tag{3.24}$$

$$x_i \geq 0, integer, \qquad \forall i \in I. \tag{3.25}$$

where $[\,.\,]_+ = \max\{0, .\}$.

3.2.6 Positive deviation from an average threshold travel time

While minimizing the total positive deviation from the average region-wise mean response time is meaningful in decreasing the variance of region-wise mean response times, it does not impose any restriction on the overall mean response time of the system. To account for the mean response time of the system in the model, we use the threshold time τ and construct a new objective function that penalizes based on the mean region-wise response time values exceeding τ. In this model, the total positive deviation of region-wise mean response time from the threshold time, τ, is minimized.

$$(\boldsymbol{P_7}) \qquad \text{Minimize} \qquad \sum_{j \in J} [R_j(\vec{x}) - \tau]_+ \tag{3.26}$$

$$\text{subject to:} \qquad \sum_{i \in I} x_i = N \tag{3.27}$$

$$x_i \geq 0, integer, \qquad \forall i \in I. \tag{3.28}$$

where $[\,.\,]_+ = \max\{0, .\}$.

The objective function in (3.26) puts an indirect bound on mean response time, $R(\vec{x})$, by minimizing positive deviations from the threshold. Hence, this does not only decrease variations among regions but also considers system-wide average performance as well.

3.2.7 Positive deviation from average region-wise lost demand

Lost demand fraction is another measure that can be evaluated from an equity perspective. This measure corresponds to the part of the society that fails to receive service under an EMS system in consideration. Similar to $\boldsymbol{P_6}$, we propose an objective function which minimizes the total positive deviation of individual region-wise lost demand fractions from the overall average of fractions.

Let $H_j(\vec{x})$ be the fraction of lost demand in region j under solution $\vec{x}$. Model $\boldsymbol{P_8}$ can be expressed as:

$$(\boldsymbol{P_8}) \qquad \text{Minimize} \qquad \sum_{j \in J} \left[H_j(\vec{x}) - \frac{\sum_{k \in J} H_k(\vec{x})}{|J|} \right]_+ \tag{3.29}$$

$$\text{subject to:} \qquad \sum_{i \in I} x_i = N \tag{3.30}$$

$$x_i \geq 0, integer, \qquad \forall i \in I, \tag{3.31}$$

where $[\,.\,]_+ = \max\{0, .\}$.

3.2.8 Limiting positive deviation from threshold travel time

We introduce another model by imposing a limit on the total positive deviation of region-wise mean response time from threshold time τ. The new constraint requires the sum of positive deviation from τ should be less than or equal to $\sigma \in [0, 1/2]$ fraction of the total deviation, hence it favors solutions with R_j's around τ. Then, proposed model with the new constraint is

$$(\boldsymbol{P_9}) \quad \text{Minimize} \quad R(\vec{x}) \tag{3.32}$$

$$\text{subject to:} \quad \sum_{i \in I} x_i = N \tag{3.33}$$

$$\sum_{j \in J} [R_j(\vec{x}) - \tau]_+ \leq \sigma \sum_{j \in J} |R_j(\vec{x}) - \tau| \tag{3.34}$$

$$x_i \geq 0, integer, \qquad \forall i \in I, \tag{3.35}$$

where $[\,.\,]_+ = \max\{0, .\}$.

(3.34) makes the problem infeasible if there exist no solutions where the total positive deviation of region-wise mean response times from τ is smaller than or equal to σ fraction of the total absolute deviation. Lack of such feasible solutions indicates that the model would result in an optimal mean response time value higher than τ without considering (3.34) where $\sigma \in [0, 1/2]$. Through the new constraint, then, we focus on the least well-off regions and the mean response time of the system together by eliminating solutions that could lead to mean response time values higher than τ.

3.3 Experimental Study

To evaluate the models presented in Section 3.2, an extensive experimental study is conducted with various problem instances having different network configurations. Problem instances are generated in a specific way to observe the effect of network features such as geographical distribution of regions, number of vehicles, and incident handling rate. Four performance measures are used for assessing the models; the mean response time, the variance of region-wise mean response time, the variance of region-wise lost demand, and the Gini coefficient. The details of the computational framework is presented in the Section 3.3.1.

3.3.1 Computational framework for the experimental study

The experimental study consists of two main parts as simulation module and analysis module. In the simulation module, the best solution for a problem instance under each model is selected. In the analysis module, the performance measures given above for each best solution are evaluated and an analysis of modeling approaches is performed based on the corresponding measures. In Figure 3.1, a flowchart is given to present the experimental study steps from the problem instance generation to the analysis of the effect of modeling approaches and network features.

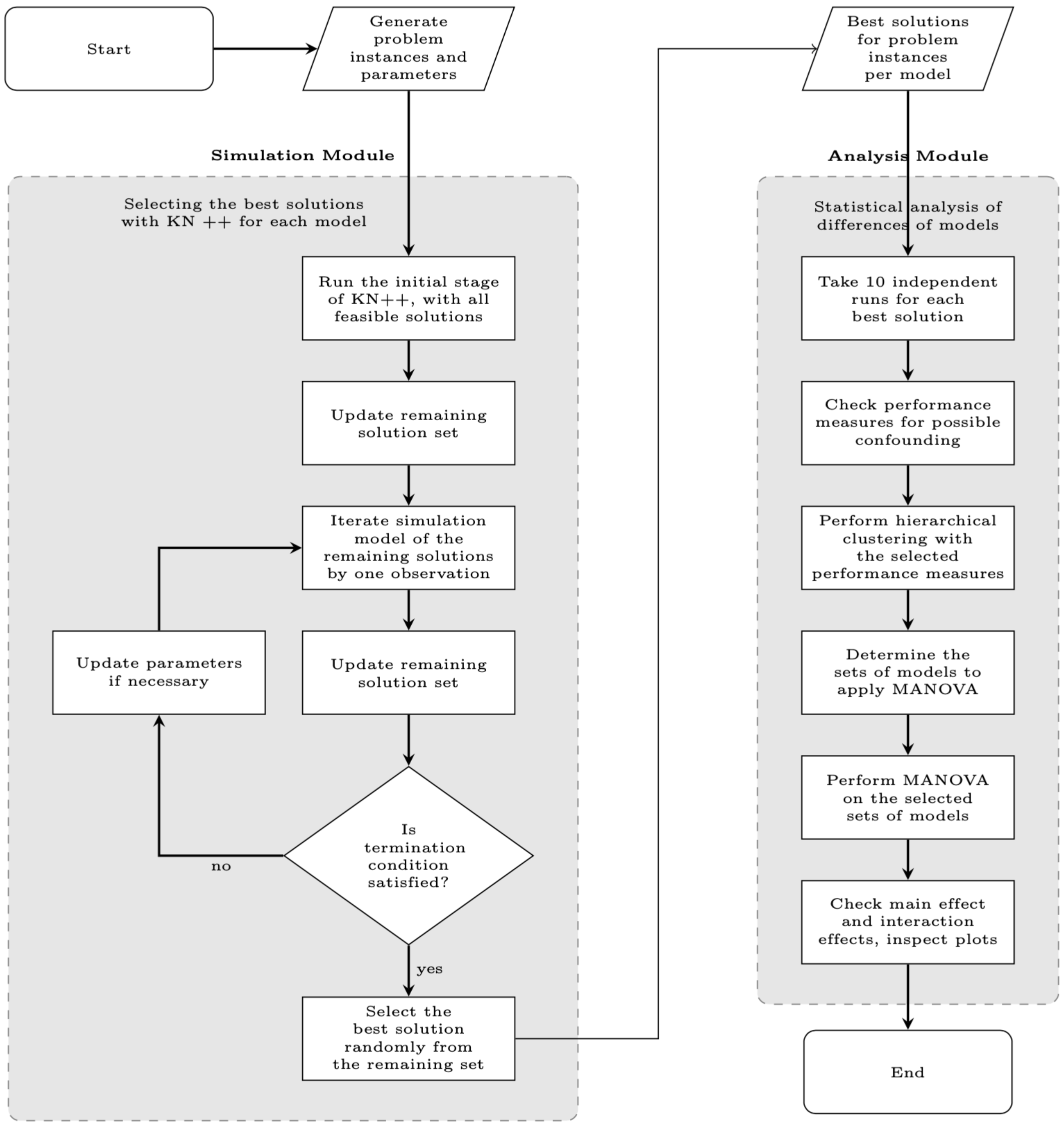


Figure 3.1: Computational framework

The computational burden of the experimental study mainly lies in the simulation module where *KN*++ algorithm is used to select the best solution for a problem instance under each model (See Section 3.3.3 for the details). In the execution of *KN*++ algorithm, all feasible solutions are evaluated with the simulation model for the initial stage in order to eliminate a set of inferior solutions. Since the simulation runs of the remaining solutions are iterated until a termination condition is satisfied, the computational log of the simulation run for each feasible solution such as demand arrival time, travel time realizations and incident handling time realization has to be stored. The computational burden of selecting the best solution under a model, hence, depends on the model of choice and the problem instance. The number of feasible solution that needs be assessed ranges from 66 to 38760 in the experiments. Due to the size complexity of the data stored in the computational environment, all data regarding the simulation run of a feasible solution is written to a separate text file. Then, the data are read from these text files to iterate the simulation runs of the remaining feasible solutions after the initial stage. The generation of the problem instances used in the experimental study and the selection of the best solutions are explained in Section 3.3.2 and 3.3.3, respectively.

3.3.2 Test bed

For the experiments, the problem instances are generated considering different network specifications. Three distribution patterns of demand regions over the plain are considered.

- *Uniform*: the regions are uniformly distributed over the area.
- *Center-accumulated*: there is an accumulation of regions in the center of the plain.
- *Outer-accumulated*, there is an accumulation of regions in an outer corner of the plain.

These patterns allow us to test the models under instances with different average pairwise distances between regions which would affect the resulting mean response time under the best solutions.

The numbers of regions and vehicles are chosen regarding the computational burden in the experiments. Every region is considered as a vehicle location candidate.

Therefore, the number of feasible solutions in the models increases exponentially with the number of vehicles and regions. All feasible solutions under a mathematical model are evaluated using a simulation model, meaning that, computational effort increases in the number of regions and vehicles. In accordance, the number of regions is set to 15 where all are defined as candidate vehicle locations, and 4, 5 or 6 vehicles are considered to be located.

Instances for the three distribution patterns with 15 demand regions are seen in Figure 3.2.

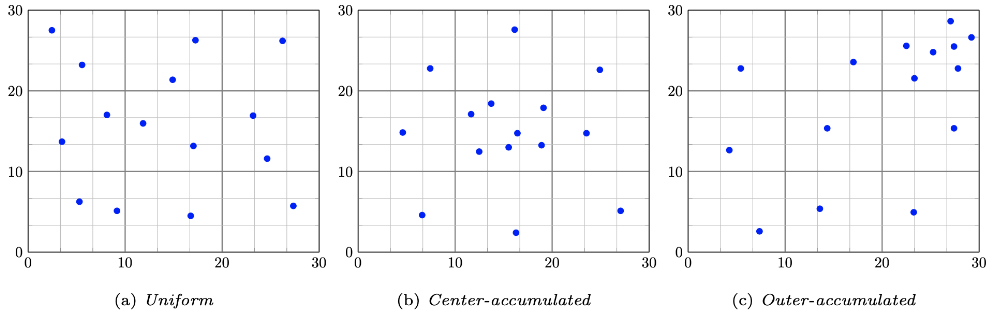


Figure 3.2: Distribution patterns of demand regions

The incident handling rate is set to 6 and 12 incidents per hour to observe the changes in the performance measures of the models when the proportions of the components in service time, namely travel time and incident handling time, change.

In total, 18 instances are generated based on full factorial design using three factors: *Distribution Pattern* with three levels, *Number of Vehicles* with three levels (4,5 and 6) and *Incident Handling Rate* with two levels (6 and 12).

Demand rate is assumed to be the same for every region as 0.5 unit per hour. Following values are set as the model parameters, minimum required coverage, α, is 0.90, the threshold time, τ, is 10 minutes, β in $\boldsymbol{P_3}$ is 0.5 and σ in $\boldsymbol{P_9}$ is 0.5.

3.3.3 Selecting the best solution for a model

Since there is no closed-form formulation for the models, a simulation model is constructed and coded in Matlab environment to simulate the emergency medical systems. The simulation model is constructed to simulate the system for the number of vehicles which corresponds to a given feasible solution in a model. Discrete event simulation is used, and performance measures of the emergency system are obtained accordingly.

The best solution for all models are found using *KN*++ algorithm by Kim and Nelson (2006) which relies on the estimates of the objective functions obtained by simulation. *KN*++ is a fully sequential selection algorithm that uses indifference zones to eliminate inferior solutions and terminates with a predefined number of alternatives remain.The algorithm tries to guarantee to select the best system with a predefined probability by comparing the long run average performance of the systems. It stops when the best system is at least a given amount better than the rest of the systems. In our context, for a given model, each feasible solution constitutes an alternative system that needs to be evaluated against others.

The procedure starts with a first-stage sample of all systems. The initial observation count is set by the decision-maker. Based on the first-stage sample, an estimator for the asymptotic variance of the difference between alternative systems is calculated. This estimator is used to find a continuation region, and the systems falling out of this region are eliminated. Then, simulation models are run for one more observation for each remaining system, and the objective function values are checked against the continuation region again for possible elimination. The algorithm terminates when a predefined number of systems remains.

Let O_{st} be the objective function value of feasible solution s after observation t, for $s = 1, 2, .., k$ and $t = 1, 2, ..., n_0$ where n_0 is the initial observation count, and $\bar{O}_s(n_0)$ is the mean of first n_0 observations from system s. An observation for a feasible solution corresponds to one demand call that is satisfied. (For $\boldsymbol{P_8}$, it corresponds to one demand call that arrives to the system since the objective function is calculated based on the lost demand fractions.)

KN++ uses batch means to calculate the asymptotic variance which is used to find a continuation region. Assume n observations $O_{s1}, O_{s2}, ..., O_{sn}$ are divided into b contiguous batches, each of length m. Let $\bar{O}_{s,a,m}$ be the a^{th} batch mean from system s. Then, the asymptotic variance, mV_b^2, that is used in the calculation of indifference zone can be found by the following estimator

$$mV_b^2 \equiv \frac{m}{b-1}\sum_{a=1}^{b}\left(\bar{O}_{s,a,m} - \bar{O}_s(n)\right)^2 \tag{3.36}$$

where $\bar{O}_s(n)$ is the mean of first n observations from system s.

The details of *KN++* algorithm can be seen in Algorithm 1.

Algorithm 1 *KN++*

Setup:

Select confidence level $1/k < 1-\gamma < 1$, indifference-zone parameter $\delta > 0$, first-stage sample size $n_0 \geq 2$, initial batch size $m_0 < n_0$.

Calculate η as the solution to the equation, $\sum_{l=1}^{c}(-1)^{l+1}\left(1-\frac{1}{2}I(l=c)\right)e^{-\frac{\eta}{c}(2c-l)l} = 1-(1-\gamma)^{1/(k-1)}$, where the constant c may be any non-negative integer.

Initialization:

Let $S = 1, 2, ..., k$ be the set of systems in contention, and let $h^2 = 2c\eta$.

Obtain n_0 observations O_{st}, $t = 1, 2, ..., n_0$, from each system $s = 1, 2, ..., k$.

Set observation counter $r = n_0$ and $m_r = m_0$.

Update:

If m_r or b_r has changed since the last update, then for all $s \neq u$, $s, u \in I$, compute estimator $m_r V_{su}^2(r)$, the sample asymptotic variance of the difference between systems s and u based on b_r batches of size m_r.

$N_{su} = \lfloor \frac{h^2 m_r V_{su}^2(r)}{\delta^2} \rfloor$ and $N_s(r) = \max_{u \neq s} N_{su}(r)$.

If $r \geq \max_s N_s(r) + 1$, then stop and select the best system in S with the smallest $\bar{O}_s(r)$ as the best. Otherwise, go to **Screening**.

Screening:

Set $S^{old} = S$.

$S = \left\{s : s \in S^{old} \text{ and } \bar{O}_s(r) \leq \bar{O}_u(r) + W_{su}(r)\, \forall u \in S^{old}, u \neq s\right\}$ where $W_{su}(r) = \max\left\{0, \frac{\delta}{2cr}\left(\frac{h^2 m_r V_{su}^2(r)}{\delta^2} - r\right)\right\}$.

If $|S| = 1$, then stop and select the system whose index is in S as the best. Otherwise, take one additional observation $O_{s,r+1}$ for each system $s \in S$, set $r = r+1$, and go to **Update**.

KN++ requires a batching sequence (m_r, b_r) to introduce more updates to ensure convergence. Kim and Nelson (2006) use the following batching sequence from the study of Goldsman et al. (2002) to update m_r and b_r parameters as the algorithm continues. Let $n_0 = b_0 m_0$ and $r = n_0 + 1$, $m_r = m_0$, $b_r = b_0$ and $f = 2$, $u_r = \sqrt{r}$. Every time another observation is realized in *KN*++ algorithm (for each new value of r), m_r and b_r are updated with respect to Algorithm 2.

Algorithm 2 Update of parameters m_r and b_r

If ($u_r < m_0$ & $r = fn_0$)

Set $m_r = m_{r-1}$

Set $b_r = 2b_{r-1}$

Set $f = 2f$

Else if ($u_r \geq m_0$ & $u(r)$ is integer)

Set $m_r = u_r$

Set $b_r = \lfloor r/m_r \rfloor$

Else

Set $m_r = m_{r-1}$

Set $b_r = b_{r-1}$

End if

Each instance is solved with models $\boldsymbol{P_1}$ - $\boldsymbol{P_9}$ using the simulation model. The best solution for all models are found using *KN*++ algorithm by Kim and Nelson (2006). For *KN*++ algorithm, first-stage sample size n_0 is set to 50,000 demand calls for $\boldsymbol{P_1}$, $\boldsymbol{P_2}$, $\boldsymbol{P_3}$, $\boldsymbol{P_5}$ and 100,000 calls for $\boldsymbol{P_4}$, $\boldsymbol{P_6}$ and $\boldsymbol{P_7}$ where steady-state behavior is harder to reach due to the use of region-wise measures in the objective function. Initial number of batches, b_0, is set to 10 for all models. The confidence level γ, indifference-zone parameter δ and parameter c are set to 0.05, 0.01 and 1, respectively. *KN*++ algorithm is terminated when the best performing alternative has a 0.1 % difference in the objective function value from the worst performing one. Then, the best solution is selected randomly from the remaining alternatives.

3.3.4 Evaluation of the performance measures

After finding the best solutions for each instance under each model (by *KN++* algorithm), all performance measures are estimated by running the simulation model. A total of 550,000 demand calls is simulated for every best solution in the simulation model, and the performance measures are reported by using batch means where the warm-up period is selected as 50,000 calls. Ten batches are constructed from the last 500,000 calls, and the batch mean of the performance measure is found. Analysis of the performance measures is conducted based on this experimental procedure.

Let $\vec{x}^*$ be the optimal solution with the best objective function value for a model, and $O(\vec{x}^*)$ be the corresponding batch mean value of the measure O for this solution from the simulation run taken after choosing the best with KN++. To compare the modeling approaches, four performance measures are used, namely the mean response time, $R(\vec{x}^*)$, the variance of region-wise mean response time, $VarR_j(\vec{x}^*)$, the variance of region-wise lost demand, $VarH_j(\vec{x}^*)$ and the Gini coefficient, $G(\vec{x}^*)$.

Gini coefficient is a well known measure in economics and used as an indicator of economic inequality. It is calculated based on the line of equality and Lorenz curve. The Lorenz curve shows the proportion of cumulative income generated by the cumulative share of the population, and it is a non-decreasing function in the cumulative population. On the other hand, the line of equality represents a society in which every individual has the same income. Gini coefficient is equal to twice the size of the area in between the Lorenz curve and the line of equality.

In our problem setting, we use the Gini coefficient to find the inequality in mean response time among regions. Cumulative mean response time and cumulative demand are considered to calculate the Gini coefficient for the best solution for a model where each region is a demand source.

In Figure 3.3, twice the size of the shaded area $\mathcal{A}$ gives the Gini coefficient for a representative instance with five demand regions. Each point in Figure 3.3 shows the percent of cumulative mean response time generated by the regions whose percent cumulative demand adds up to the coordinate in x-axis.

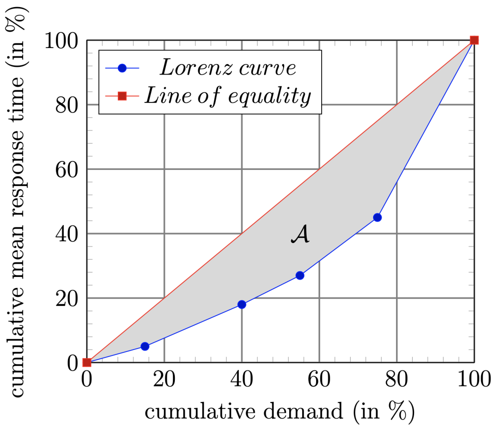


Figure 3.3: Graphical representation of Gini coefficient for a problem instance with 5 demand regions

3.3.5 Results and discussion

In this section, various measures are reported for the best solution for each problem instance under models $\boldsymbol{P}_1$ - $\boldsymbol{P}_7$.

$\boldsymbol{P}_8$ is excluded from the analysis since objective function values are very small for all feasible solutions of all instances, and *KN*++ reports a different best solution in every run. This inconsistency is attributed to the fact that demand rates for all regions are identical, resulting in lost demand fractions close to each other for all regions in each feasible solution since there is no queuing and prioritization of the calls. Due to the same reason, the variance of region-wise lost demand, $VarH_j(\vec{x})$, is also excluded from the performance measures used to assess the equity.

For Model $\boldsymbol{P}_9$, it is not possible to use *KN*++ algorithm since constraint (3.34) needs to be assessed after every demand call met. This procedure would lead some solutions to leave the alternative solution list and re-enter in another iteration, which is not possible in the structure of *KN*++ algorithm. Instead, the best solutions reported for $\boldsymbol{P}_1$ are checked whether they are feasible for $\boldsymbol{P}_9$ since both models have the same objective function. It is seen that the best solutions for $\boldsymbol{P}_1$ do not violate (3.34), and they are treated as the best solutions for $\boldsymbol{P}_9$ as well. Since $\boldsymbol{P}_1$ and $\boldsymbol{P}_9$ shares the same best solutions, $\boldsymbol{P}_9$ is excluded from the further analysis.

The experimental results are reported in various ways to analyze the effects of network specifications and models on different performance measures. In Figure 3.4, $R(\vec{x})$, $VarR_j(\vec{x})$ and $G(\vec{x})$ results are shown as bar charts with the first and third quartiles for each model. $(\vec{x})$ is dropped from the notation of the performance measures in the rest of the study to facilitate reading.

It is seen that $\boldsymbol{P_6}$ provides better equity among regions due to low $VarR_j$ and G values, whereas the mean response time, R, is higher than other models. In all 18 instances, $\boldsymbol{P_6}$ gives the highest R and the smallest G values. In thirteen instances, it provides the smallest $VarR_j$ values. These results show that $\boldsymbol{P_6}$ is the model that yields the smallest variance in service quality and betters off equity among regions the most. As expected, this results in a substantial increase in the overall mean response time value.

From Figure 3.4, we observe that $\boldsymbol{P_1}$, $\boldsymbol{P_3}$, $\boldsymbol{P_5}$ show similar results in the performance measures within each other. $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ are the other models which behave similarly. To test the similarity and significance of the differences between models, hierarchical clustering and statistical tests are used.

First, the performance measures used to quantify equity in the system, $VarR_j$ and G, are checked for correlation to prevent confounding in the tests. It is seen that $VarR_j$ and G are highly correlated with a correlation coefficient of 0.79. Since these two measures are highly correlated, $VarR_j$ is dropped in the application of hierarchical clustering. G is used as the equity measure since it is more sensitive to outliers than $VarR_j$ as stated by Yitzhaki and Schechtman (2013), therefore it would be better in assessing equity from a Rawlsian perspective.

Hierarchical agglomerative clustering is used to form clusters based on dissimilarity between the models. In agglomerative clustering, each observation starts in its own cluster, and pairs of clusters are merged based on a dissimilarity measure. In this study, Mahalanobis distance is used as the dissimilarity measure. It considers the correlation between multiple measures and removes the scale effect on them. Mahalanobis distance between models P_m and P_n is found as follows:

$$d(P_m, P_n) = \sqrt{(\vec{O}_{P_m} - \vec{O}_{P_n})S^{-1}(\vec{O}_{P_m} - \vec{O}_{P_n})^T} \tag{3.37}$$

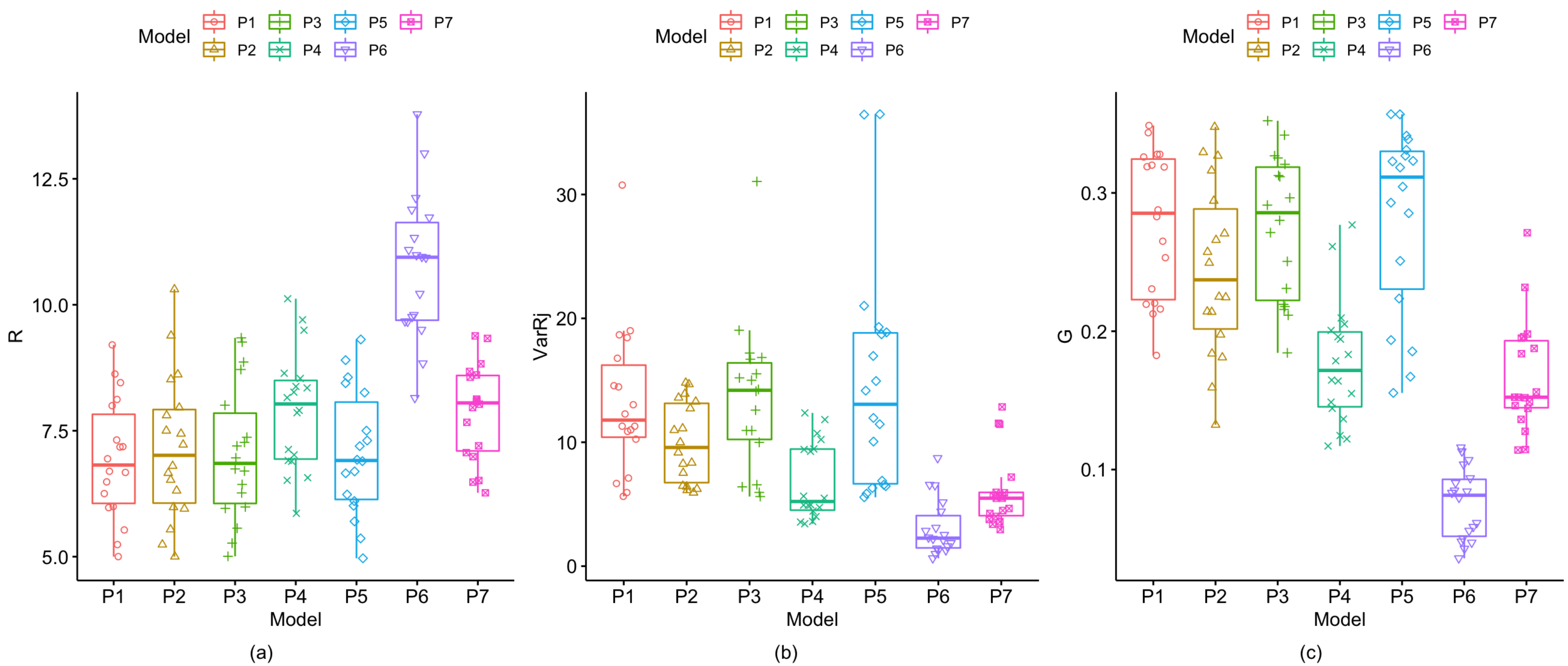


Figure 3.4: R, $VarR_j$ and G results for the best solution for models in 18 instances

where $\vec{O}_{P_m}$ is a row vector of $(\bar{R}, \bar{G})$, $\bar{R}$ is the average mean response time and $\bar{G}$ is average Gini coefficient under P_m over all instances and S is the covariance matrix for the measures R and G.

As clusters emerge, single linkage is used to define the distance between the clusters. In single linkage, the distance between clusters is taken as the distance between the nearest neighbors of these clusters. Then, the two clusters with the smallest distance are merged. The clusters are shown on a dendrogram which visualizes the merge of clusters through the iterations with the corresponding dissimilarity values. The models are clustered, starting with each model as a cluster itself, and merged with the most similar ones until all models merged into one final cluster. The resulting dendrogram for clustering of the models is presented in Figure 3.5.

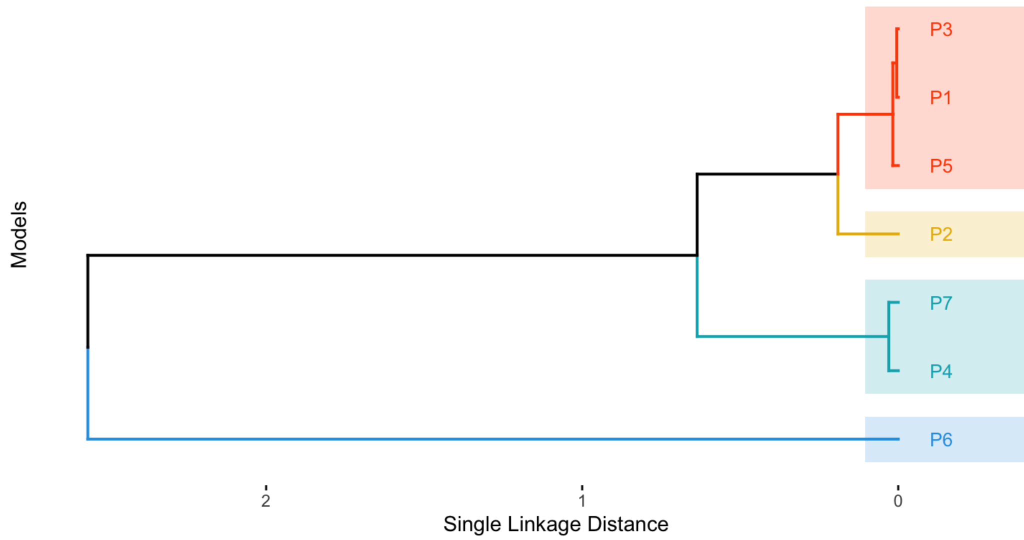


Figure 3.5: Dendrogram with Mahalanobis distance

Initially, each model is a single cluster. $\boldsymbol{P_1}$ and $\boldsymbol{P_3}$ is the most similar pair to each other in comparison to the other pairs and these models are clustered firstly. Then, $\boldsymbol{P_5}$ is grouped with the cluster of ($\boldsymbol{P_1}$, $\boldsymbol{P_3}$). The next closest clusters are the clusters of $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$, so they are clustered next. Then, $\boldsymbol{P_2}$ is clustered with the cluster of ($\boldsymbol{P_1}$, $\boldsymbol{P_3}$, $\boldsymbol{P_5}$). Later, all models except $\boldsymbol{P_6}$ are clustered together before the final single cluster. The most dissimilar model among all is $\boldsymbol{P_6}$, similar to our observation from Figure 3.4.

To check the statistical significance of the difference between performance measures obtained under the best solution for models clustered together, we use full factorial multivariate analysis of variance (MANOVA) since we have multiple response variables as R and G.

The mathematical model, distribution pattern of regions, number of vehicles, and incident handling rate are taken as factors. *Model* has a different number of levels based on the comparison. *Pattern* has three levels as uniform, center-accumulated, outer-accumulated. *Number of Vehicles* has three levels as 4,5 and 6, and *Incident Handling Rate* has two levels as 6 and 12. For the statistical tests, we take ten replications for each model's best solution using the simulation model.

We apply MANOVA for four experiments where *Model* factor has levels of ($\boldsymbol{P_1}$, $\boldsymbol{P_3}$), ($\boldsymbol{P_1}$, $\boldsymbol{P_3}$, $\boldsymbol{P_5}$), ($\boldsymbol{P_4}$, $\boldsymbol{P_7}$) as the first three clusters formed, and ($\boldsymbol{P_1}$ - $\boldsymbol{P_7}$) as all models. In all four tests, p-values for all factors and interaction terms are smaller than $2E^{-16}$. This result shows that *Model*, *Pattern* and *Number of Vehicles*, and their interactions affect the performance measures R and G significantly in all MANOVA models. Further details of MANOVA, and the test results for the experiment with *Model* factor level ($\boldsymbol{P_1}$, $\boldsymbol{P_3}$) as an example are given in Appendix A.

In order to quantify the difference of models from the base model $\boldsymbol{P_1}$, we provide the performance measures obtained under the best solutions in relative to the ones obtained under the best solution for $\boldsymbol{P_1}$. The mean absolute percent deviation of measures of the models, $MAPD$, from base model $\boldsymbol{P_1}$ is calculated for each measure. Let k represents k^{th} instance in the set of instances, K, then $MAPD$ between model P_m and base model $\boldsymbol{P_1}$ is found as:

$$MAPD = \frac{100}{|K|} \sum_{k \in K} \left| \frac{O_{P_m}(k) - O_{P_1}(k)}{O_{P_1}(k)} \right| \tag{3.38}$$

where $O_{P_1}(k)$ is the batch mean of measure O for the best solution in instance k under $\boldsymbol{P_1}$.

We report the average positive percent deviations of the mean response time, R, of the best solution in the models from $\boldsymbol{P_1}$. This measure for a model P_m is

$$avgpd(\%) = \frac{100}{|K'|} \sum_{j \in K'} \frac{O_{P_m}(j) - O_{P_1}(j)}{O_{P_1}(j)} \tag{3.39}$$

where $K' = \{k \in K : O_{P_m}(k) - O_{P_1}(k) > 0\}$.

The percent of instances with positive deviation, $ppd(\%)$, in R is also reported to show the fraction of instances with positive deviation. $ppd(\%)$ for P_m is calculated as:

$$ppd(\%) = \frac{100\,|K'|}{|K|} \tag{3.40}$$

where $K' = \{j \in K : O_{P_m}(j) - O_{P_1}(j) > 0\}$.

For $VarR_j$ and G, the average negative percent deviation, $avgnd(\%)$, and the percent of instances with negative deviation, $pnd(\%)$, is reported instead since equity gets better as $VarR_j$ and G decrease. In Table 3.1, these statistics are reported to show the changes in R, $VarR_j$ and G.

Table 3.1: Comparison of Models P_2 to P_7 with P_1 in performance measures

Measure	Statistics	P_2	P_3	P_4	P_5	P_6	P_7
	$MAPD(\%)$	2.86	1.77	14.62	1.85	57.10	14.79
R	$ppd(\%)$	83.33	72.22	100.00	83.33	100.00	100.00
	$avgpd(\%)$	3.34	2.21	14.62	2.10	57.10	14.79
	$MAPD(\%)$	18.89	7.00	45.37	26.21	69.77	51.73
$VarRj$	$pnd(\%)$	77.78	44.44	100.00	44.44	100.00	100.00
	$avgnd(\%)$	23.24	4.64	45.37	16.71	69.77	51.73
	$MAPD(\%)$	12.68	2.60	35.73	10.99	72.63	39.05
G	$pnd(\%)$	77.78	61.11	100.00	50.00	100.00	100.00
	$avgnd(\%)$	15.76	2.67	35.73	9.38	72.63	39.05

As seen in Table 3.1, $\boldsymbol{P_6}$ yields the biggest changes in all three measures where $\boldsymbol{P_7}$ follows as the second. Compared to the best solutions of $\boldsymbol{P_1}$, the mean response time (R) worsens in all the instances, and equity measures $VarR_j$ and G better off in all instances under the best solutions of $\boldsymbol{P_6}$. $\boldsymbol{P_6}$ has the worst R values with more than 30% deviation in each instance and an overall average of 57%. This result is attributed to the objective function being analogous to minimizing the variance of region-wise mean response time since we minimize the sum of positive deviation of region-wise mean response time from its average. This objective function results in the best solution with minimum $VarR_j$ in each but five instances, minimum G and maximum R results in each instance among models. The statistics in Table 3.1 show that the best solutions under $\boldsymbol{P_3}$ yields very small $MAPD$ in R ($\boldsymbol{P_5}$ follows it) where others show significant deviations from $\boldsymbol{P_1}$.

Based on the results in Table 3.1, we present our findings as observations and give additional plots in the rest of this section.

Observation 1 *Coverage maximization could be an alternative objective when mean response time is harder to derive.*

Having mean response time values R almost inline under $\boldsymbol{P_1}$ and $\boldsymbol{P_5}$ implies that the coverage maximization objective is comparable to mean response time minimization in the resulting mean response time values under the optimal solutions.

Observation 2: *Deterministic coverage constraints do not deliver as expected if used to narrow down the solution space, while chance constraints are better at capturing the minimum possible mean response time.*

In ten out of eighteen instances, the best solutions under $\boldsymbol{P_3}$ are the same with the ones under $\boldsymbol{P_1}$. On the other hand, $\boldsymbol{P_2}$, which utilizes deterministic coverage constraints, evaluates the same best solutions in five instances, and the average deviation (3.34%) in the mean response time from $\boldsymbol{P_1}$ for other instances is slightly higher than $\boldsymbol{P_3}$. Therefore, chance constraints for coverage as in $\boldsymbol{P_3}$ could be useful in narrowing down the solution space if the objective is to achieve the minimum possible mean response time.

Another measure that we are interested in is the variance of region-wise mean response time. It could be considered a Rawlsian equity measure where one is interested in response time differences in among regions.

In terms of $VarR_j$, the models $\boldsymbol{P_2}$, $\boldsymbol{P_4}$, $\boldsymbol{P_6}$ and $\boldsymbol{P_7}$ stand out in decreasing the variance among regions in comparison to $\boldsymbol{P_1}$.

Observation 3: *Minimizing total positive deviation of region-wise mean response times from the average region-wise mean response time of all regions worsens mean response time, R, while it provides better equity among the regions.*

$\boldsymbol{P_6}$ has the smallest variance among instances except for five instances. However, it has the largest (worst) R in all instances. $\boldsymbol{P_6}$ has 70% decrease in variance whereas $\boldsymbol{P_4}$ follows with 45%, $\boldsymbol{P_7}$ with 52% and $\boldsymbol{P_2}$ with 23% decrease in variance on the average.

Observation 4: *Deterministic coverage constraints impose better equity among regions than chance constraints, with a small negative effect on overall mean response time.*

$\boldsymbol{P_2}$ has a slight effect on R as 3% increase; however, it helps to decrease the variance of region-wise mean response time by 23% on the average in fourteen out of eighteen instances. Chance constraints in $\boldsymbol{P_3}$ could be used to narrow down the solution space (in the expense of 2% increase in R) since its effect on equity is substantial with an average decrease of 5% in $VarR_j$ in eight instances. So, incorporating chance constraints in assessing coverage works towards the overall system performance rather than seeking the performance from an equity perspective.

In accordance with $VarR_j$ results, models $\boldsymbol{P_2}$, $\boldsymbol{P_4}$, $\boldsymbol{P_6}$ and $\boldsymbol{P_7}$ have smaller Gini coefficients than other models have in comparison to $\boldsymbol{P_1}$ according to Table 3.1. $\boldsymbol{P_6}$ is significantly better than most of the models, whereas $\boldsymbol{P_2}$ has a closer Gini coefficient to the base model $\boldsymbol{P_1}$ than others, but it still has better equity. The average percent negative deviation $avgnd\%$ and $MAPD$ for $\boldsymbol{P_2}$, $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ are slightly smaller than in $VarR_j$. However, this does not change the general behaviour of the models in terms of equity.

Observation 5: *Minimizing maximum region-wise mean response time increases equity but worsens overall service quality.*

$\boldsymbol{P_4}$ worsens R by about 15% whereas G decreases by 36% on the average with respect to $\boldsymbol{P_1}$. $\boldsymbol{P_7}$ has a greater effect on G (39% decrease) while it results in higher R values than $\boldsymbol{P_1}$ has by 15% on the average. Focusing on the least well-off region as in $\boldsymbol{P_4}$ does not result in better equity among regions than $\boldsymbol{P_7}$, where the focus in $\boldsymbol{P_7}$ is on total region-wise positive deviation from a threshold.

Observation 6: *Minimizing total region-wise positive deviation from a threshold results in a more equitable system, and its effect on system-wide performance is limited in comparison to its effect on equity measures.*

Although $\boldsymbol{P_7}$ puts an indirect bound on mean response time while minimizing the positive deviation from a threshold, it provides a better equity than $\boldsymbol{P_4}$ which is a model focusing on the region with the worst mean response time. Besides, $\boldsymbol{P_7}$ has a similar system-wide performance measure as $\boldsymbol{P_4}$. $\boldsymbol{P_7}$ also balances out the shortcoming of $\boldsymbol{P_6}$ where $\boldsymbol{P_6}$ ends up putting all the emphasis on the variance of region-wise mean response time.

The observations up to this point coincides with the results of the hierarchical clustering analysis that ($\boldsymbol{P_1}$, $\boldsymbol{P_3}$, $\boldsymbol{P_5}$) show similar performances, and ($\boldsymbol{P_4}$, $\boldsymbol{P_7}$) are the other models which behave similarly.

In addition to the comparison with $\boldsymbol{P_1}$, main effect and interaction effects of factors are also explored to gather insight into the changes of performance measures. In Figure 3.6, main effect plots for factors *Model*, *Distribution Pattern*, *Number of Vehicles* and *Incident Handling Rate* are given for three performance measures.

In the first plot of each row in Figure 3.6, the main effect of *Model* on the performance measures over all instances are consistent with the previous analysis on the similarities and dissimilarities between the models.

Observation 7: *Distribution of the demand regions over the plain significantly affects system-wide average performance and equity.*

Concerning the results shown in Figure 3.6 (f) & (j), it is seen that equity among regions is better when demand regions are uniformly distributed over the plain. In the case of accumulation of regions in some part of the plain, equity tends to worsen, especially in *Center-accumulated* instances.

Observation 8: *Increasing number of vehicles does not necessarily better off equity while decreasing mean response time.*

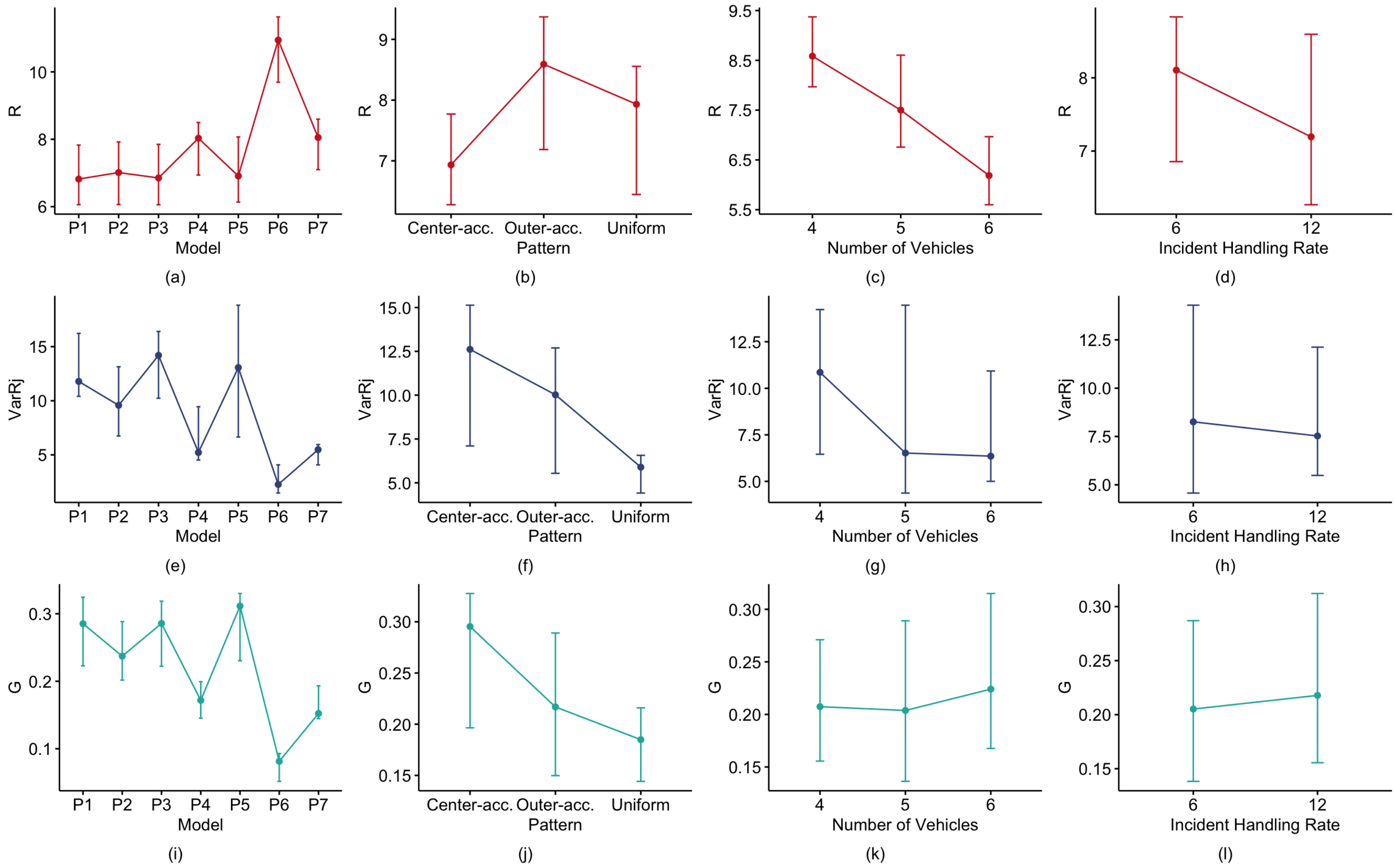


Figure 3.6: Main effect plots for R, $VarR_j$ and G

Mean response time, R, decreases in the number of vehicles, see Figure 3.6 (c). This result is intuitive as more vehicles are available, the average travel time to demand regions from vehicle locations could decrease. However, this does not directly result in the increase of equity among regions as the change in G does not show a regular pattern while $VarR_j$ decreases, see Figure 3.6 (g) & (k).

Observation 9: *Increasing incident handling rate does not necessarily better off equity while decreasing mean response time.*

The average mean response time over instances is decreased as the incident handling rate increases as in Figure 3.6 (d). The incident handling time is one of the three components of service time (sum of travel time to the region, incident handling time, and travel time back to vehicle location), so an increase in incident handling rate makes service time decrease. Due to this decrease, regions would be served by the closest vehicle more, which would decrease the mean response time. However, the change in the equity measures is not consistent as the average of $VarR_j$ decreases while G increases.

To better understand the effect of network specifications on the behavior of different models, we present another set of analysis in Figure 3.7. In this figure, the interaction of network features with modeling approaches in performance measures can be seen.

Observation 10: *The choice of modeling approach has greater importance in the existence of accumulation of demand regions over the plain.*

The change in equity measures, $VarR_j$ and G, are more prominent in patterns *Center-accumulated* and *Outer-accumulated* with respect to the results in Figure 3.7 (d) & (g) while the line connecting medians of each model under *Uniform* pattern is flatter. However, mean response time R is affected similarly in all three patterns. This shows us that the distribution of demand regions would change the scale of the effect of the models on equity, and the preference of a mathematical model over another could be more desirable under certain demand distribution patterns.

Only in model $\boldsymbol{P_3}$, the effect in equity has a similar behavior as in $\boldsymbol{P_1}$, irrespective of the distribution of regions and the number of vehicles. This shows that $\boldsymbol{P_3}$ behaves very similarly to the base model $\boldsymbol{P_1}$ under different network specifications

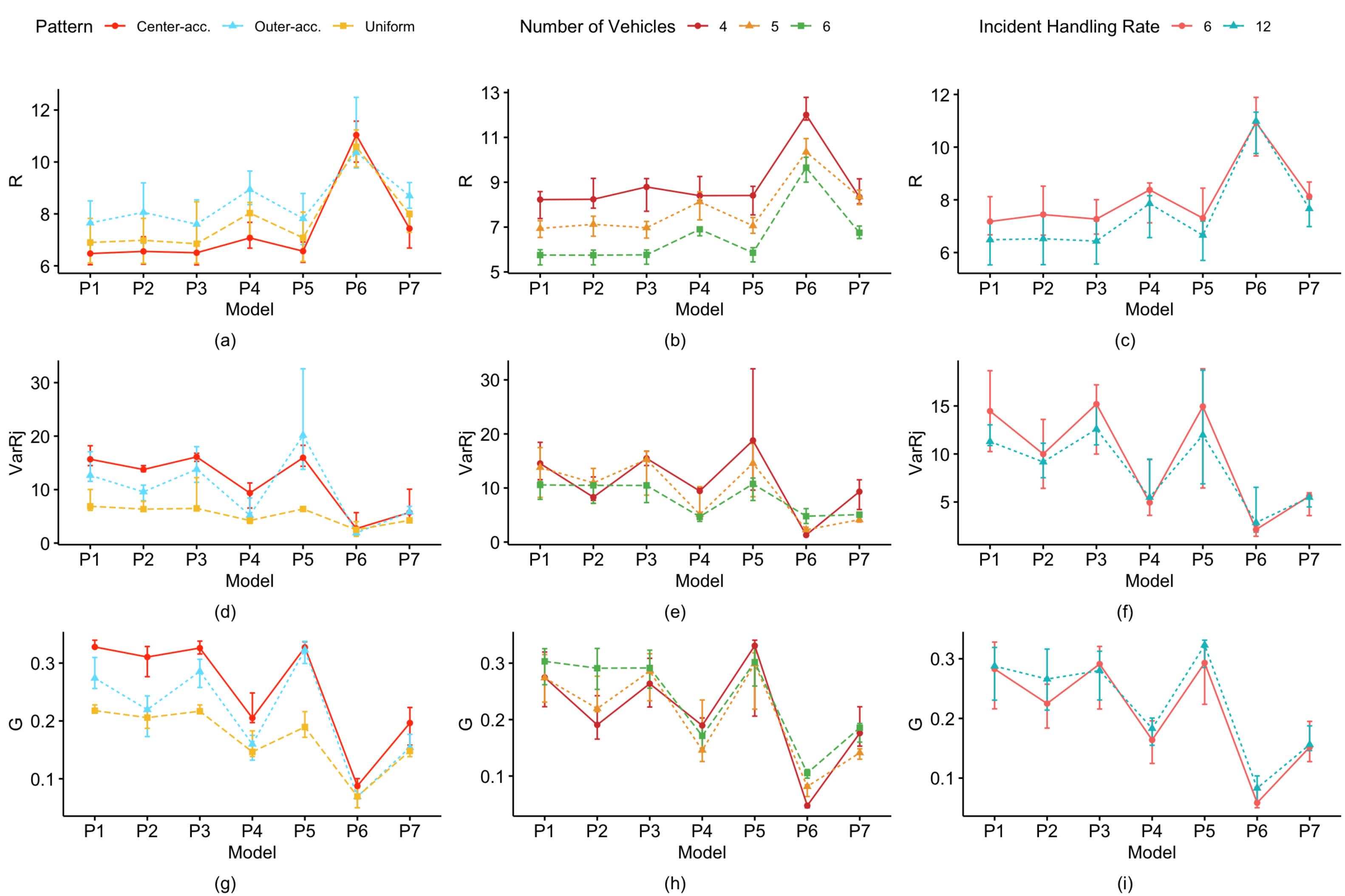


Figure 3.7: Interaction of R, $VarR_j$ and G with models and network specifications

and reinforces the previous inference.

Observation 11: *Minimizing variance results in similar equity irrespective of the distribution of regions.*

The performance measures $VarR_j$ and G are very close for $\boldsymbol{P_6}$ in all patterns in Figure 3.7. This might result from the fact that $\boldsymbol{P_6}$ enforces the best obtainable equity by minimizing the total positive deviation from the average in the objective function.

In Figure 3.8, interaction between factors are given in a different setting where this time *Distribution Pattern* and *Number of Vehicles* are on the x-axis.

The scale of change in equity measures under $\boldsymbol{P_4}$, $\boldsymbol{P_5}$ and $\boldsymbol{P_7}$ are easier to observe in Figure 3.8 (c) & (e). The difference between models is more prominent in the patterns having an accumulation of demand regions in some parts of the plain.

Following the second plot in each row, it is seen that an increase in the number of regions consistently decreases mean response time. However, this does not translate into a consistent behavior in equity measures, as it is stated in Observation 8. It is seen that equity worsens in our instances under $\boldsymbol{P_2}$ and $\boldsymbol{P_6}$ as the number of vehicles increases, while in other models, it tends to better off. Therefore, the change in $VarR_j$ and G in the number of vehicles depends on the model.

Observation 12: *Increasing incident handling rates worsens equity in models with the objective of minimizing inequality among regions.*

The change in equity measures is different under different incident handling rates It is seen that the change in $VarR_j$ is dependent on the model when the incident handling rate increases. For $\boldsymbol{P_1}$, $\boldsymbol{P_2}$, $\boldsymbol{P_3}$, $\boldsymbol{P_5}$, there is a decrease in variance from rate 6 to 12 (a slight decrease for $\boldsymbol{P_2}$), while in the other models there seem to be slight increases. It is also seen that G values increase in $\boldsymbol{P_4}$, $\boldsymbol{P_6}$ and $\boldsymbol{P_7}$ with rate increase. These models are similar to each other in the objective function where the objective is to minimize inequality among regions in general. This shows that increasing incident handling rates worsens equity for models having this type of objective function. We would like to point out that the incident handling time is the smaller portion of the overall service time in the best solutions of instances under all models for both incident handling

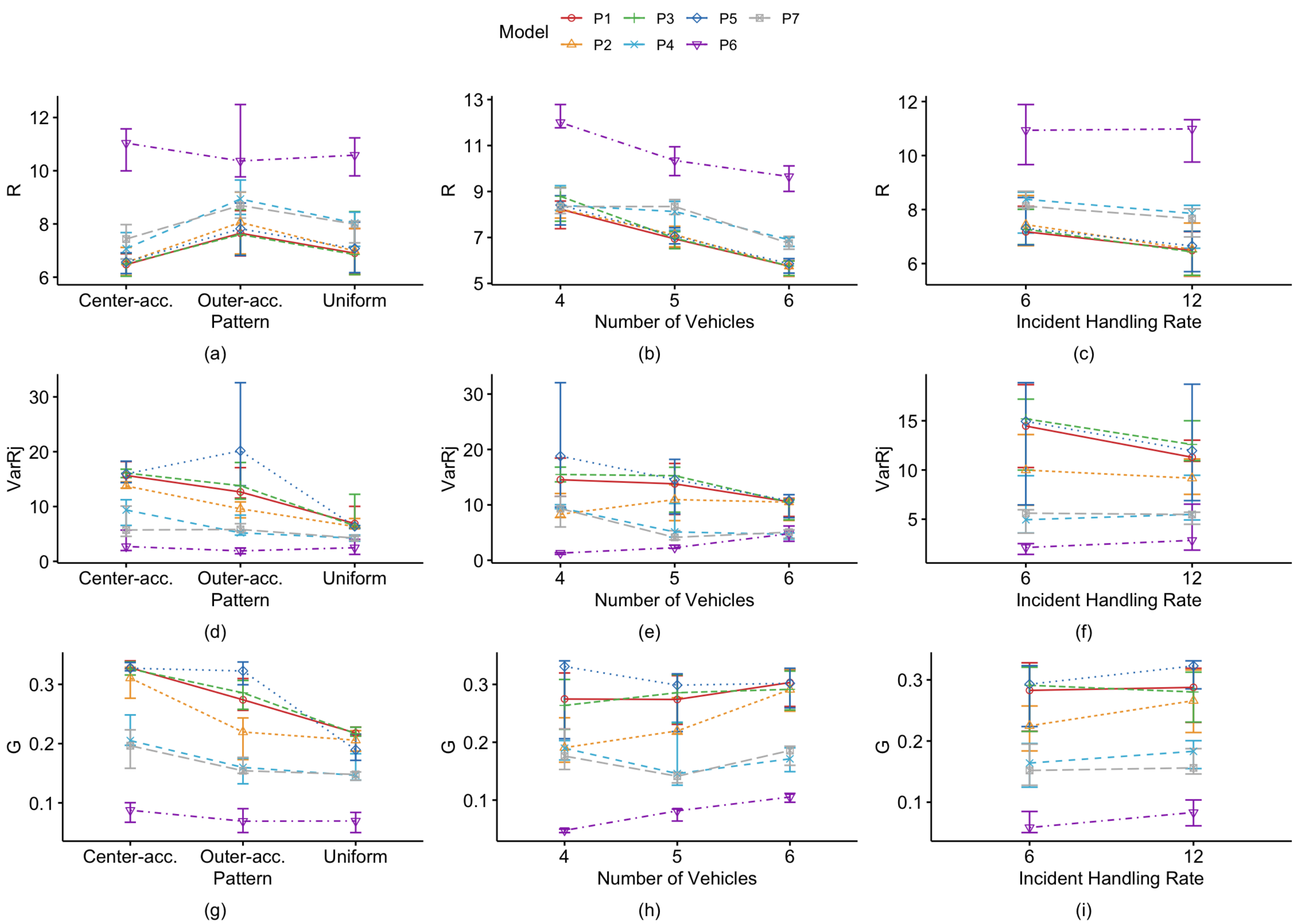


Figure 3.8: Interaction of R, $VarR_j$ and G with models and network specifications

rates.

The change in incident handling rates affects service time differently based on the travel times for a location solution. If the average travel time is a smaller component of the service time, the effect of increasing incident handling rate on average service time would be more significant. We realize that average service time depends on the location solution; however, we construct the problem instances in a way that the distribution of demand regions over the plain is an indicator of the average service time, where we control demand rates, incident handling rates and number of vehicles systematically. This is also justified by the interaction of R with distribution patterns in Figure 3.8 (a). The average pairwise distances between regions are 15.31, 12,2, and 13.58 minutes in instances with outer-accumulated, center-accumulated and uniform patterns, respectively. It is seen that mean response time, R, has the highest values for outer-accumulated and lowest values in center-accumulated instances for the best solution for a given model except $\boldsymbol{P_6}$ as anticipated. Therefore, another plot is presented in Figure 3.9 to show the interaction effect of incident handling rate and distribution of regions over the plain on the behavior of the models.

The results in Figure 3.9 show that the scale of change in R with increasing incident handling rates is directly proportional to average pairwise distance in instances where we observe the most significant decreases in outer-accumulated pattern.

Observation 13: *The difference between modeling approaches in terms of equity lessens as the traffic intensity of the system decreases.*

The effect of incident handling rate on the equity measures changes with respect to the distribution pattern and the models. It is seen that $VarR_j$ values are very close to each other in uniform pattern for a given model. For the other patterns, equity is affected based on the model of choice. As the incident handling rate increases (which would decrease the system's traffic intensity), the difference in $VarR_j$ among models tends to get less prominent where the line connecting the medians gets flatter. This change is more prominent in outer-accumulated pattern. However, G results are similar irrespective of the incident handling rate, which could be due to the Gini coefficient's sensitivity to outliers.

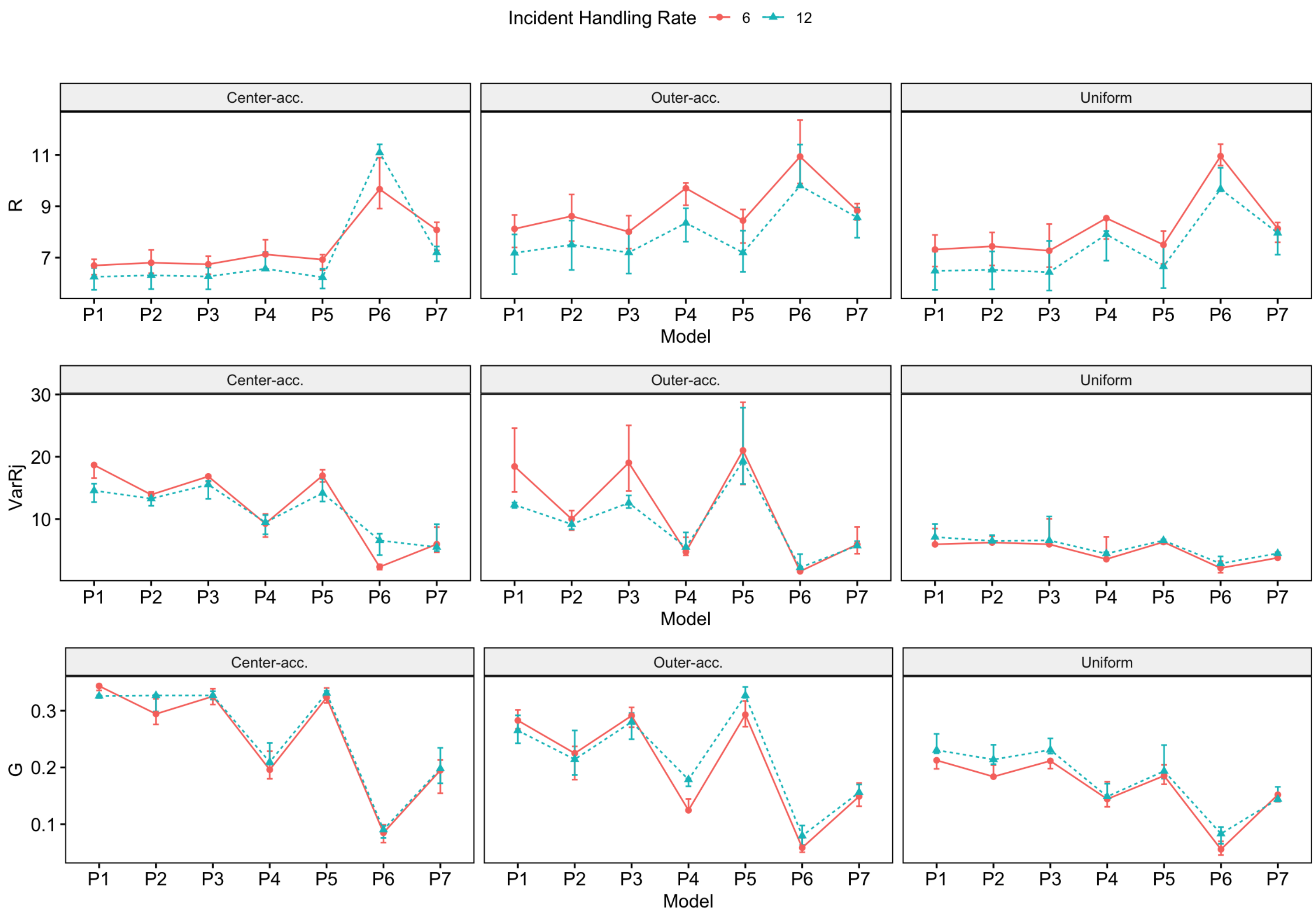


Figure 3.9: Interaction of R, $VarR_j$ and G with models and incident handling rate under different distribution patterns

Lastly, we reinforce Observation 10 with the help of Figure 3.9. It could be seen that the change in equity measures with respect to models is more prominent in patterns with an accumulation of demand regions (center-accumulated and outer-accumulated).

3.4 Conclusion

In this chapter, we work on various models for EMS vehicle location problem in an effort to address issues related to equity. The motivation of this chapter is that EMS vehicle location problem requires a more comprehensive approach with multiple important criteria. Otherwise, decision-makers sacrifice from equity where emergency care should be reasonably available to every member of society.

It is seen that models focusing on overall performance, such as minimization of mean response time without any further considerations, do not perform well in terms of equity among regions. Deterministic coverage constraints are more promising in decreasing the disparities between regions than chance constraints. It is also possible to improve equity by focusing on region-wise performance measures such as mean region-wise response time. On the other hand, the use of chance constraints is found to be more promising in narrowing down the feasible region than deterministic constraints. In the results, it is shown that network specifications affect the equity enforced by the models. Therefore, it is better to understand the network specifications and incorporate them in the process of choosing the mathematical model to be used.

In this chapter, we stick to a simulation study to evaluate the EMS systems' performance measures since the environment involves uncertainty and the exact queueing model is computationally expensive. One important research direction is to obtain the performance measure via deterministic approximation models/algorithms rather than simulation study. In the next chapter, we work on a decomposition method to approximate the performance measure of the exact queueing model and evaluate the performance of an EMS system.

CHAPTER 4

DECOMPOSITION METHODS FOR ESTIMATING THE PERFORMANCE MEASURES OF STOCHASTIC EMS SYSTEMS

In Chapter 3, we work on emergency vehicle location problem form an equity perspective rather than developing a mathematical model for the problem. The evaluation of alternative solutions for an EMS system is done based on discrete event simulation which is computationally expensive. In this chapter, we study on developing a method to evaluate performance measures of an EMS system based on stochastic processes, which could be used to locate emergency vehicles.

The exact queueing model constructed for the problem is computationally cumbersome as the size of the state space increases exponentially in the number of regions and vehicles to be located. Accordingly, we work on a decomposition method to approximate the performance measures of an EMS system under the exact queueing model in this chapter. Instead of working on a single queueing model as HQM by Larson (1974), the exact queueing model is decomposed into interdependent partial queueing models whose balance equations form a non-linear set of simultaneous equations. This decomposition enables us to work with service rates that depend on the server location and the region. The decomposition methods proposed are also capable of representing solutions with multiple vehicles at a single location. Different than Budge et al. (2009), we distinguish the vehicles at a single location by assigning priorities in dispatching, which allows the decision-maker to observe the change in the busyness of a vehicle when an additional vehicle is located at the exact location. We present an approximation method to solve the resulting set of equations and estimate the steady-state probabilities. Once the steady-state probabilities are evaluated, several performance measures could be obtained based on those.

We present two optimization problems and use the decomposition methods in the evaluation of the objective function value. An analysis of the performance of the decomposition methods under optimization setting is presented, which is not shown in studies of Budge et al. (2009) and Toro-Díaz et al. (2015), by employing a ranking selection algorithm based on simulation to obtain best performing location solutions. Since the mathematical models have no closed-form formulation, a genetic algorithm is presented to find near-optimal solutions.

In the rest of the chapter, the problem environment is introduced in Section 4.1, and the decomposition of the exact queueing model is explained in Section 4.2. A meta-heuristic solution algorithm for the mathematical model is given in Section 4.3 The experimental results for the performance of decomposition methods and the meta-heuristic algorithm are given in Section 4.4. Lastly, the chapter is concluded in Section 4.5.

4.1 Problem Definition

We construct the mathematical model with only one set of decision variables, $x_i, \forall i \in I$ which is the number of vehicles located in vehicle location i. Let $\vec{x}$ be the vector of decision variables, x_i's, representing a solution. The objective function, $R(\vec{x})$, is the mean response time of the system under solution $\vec{x}$. It is one of the most critical measures that is taken into account for the evaluation of emergency systems and defined as the time passed between a demand call and the arrival of a vehicle at the corresponding demand location. As we define our problem environment based on stochastic processes, there is no closed-form formulation of the performance measures as a function of $\vec{x}$; therefore, the objective function is defined as a function of a candidate solution without introducing further notations.

We define two mathematical models P_S and P_M where we allow only a single and multiple vehicles to be located at a given location in P_S and P_M, respectively.

P_S and P_M are as follows:

$$(\boldsymbol{P_S}) \quad \text{Minimize} \quad R(\vec{x}) \tag{4.1}$$

$$\text{subject to:} \quad \sum_{i \in I} x_i = N \tag{4.2}$$

$$x_i \in \{0, 1\}, \qquad \forall i \in I. \tag{4.3}$$

$$(\boldsymbol{P_M}) \quad \text{Minimize} \quad R(\vec{x}) \tag{4.4}$$

$$\text{subject to:} \quad \sum_{i \in I} x_i = N \tag{4.5}$$

$$x_i \geq 0, integer, \qquad \forall i \in I. \tag{4.6}$$

In the next section, we propose decomposition algorithms to evaluate the objective function values of these problems based on queueing theory.

4.2 Decomposing the Exact Queueing model

Recall the exact queueing model introduced in Chapter 1 which could be used to evaluate the performance measures of an EMS system.

$\{B_t, t \geq T\}$ is a continuous time Markov chain with state space L, and B_t is an N-dimensional state variable defining the underlying queueing system as follows:

$$B_t = (b_1, b_2, ..., b_N), \qquad t \geq T$$

where b_k represents the status of the k^{th} vehicle at time t and denotes a 3-tuple as:

$$b_k = (i_k, s, m), \; s \in J \cup \{0\}, \; m = 0, 1, 2, 3$$

where i_k represents the location that the vehicle is located, s stands for "idle" or the region being served, and m represents the mode of the service at time t.

If the vehicle is not busy, s and m are set to 0. Then, $b_k = (i_k, 0, 0)$ represents the state that the vehicle at location i_k is free. If the vehicle is busy serving demand region $j \in J$, s is set to j. When the vehicle is busy, m is set to $1, 2$ or 3 if the vehicle is en-route to demand region j, handling the incident or en-route to the vehicle location, respectively.

Under this definition, the size of the set of the possible statuses for every $b_k \in B_t,\ k = 1, \ldots, N$ is equal to $(3\,|J| + 1)$. Hence, the state definition results in a state space L with $(3\,|J| + 1)^N$ many states, and the size of the state space increases exponentially with the increasing number of vehicles. Therefore, we propose an approximation method to estimate the performance measures that would result from the exact queueing model.

The exact queueing model is decomposed into N interdependent queueing models, each representing a single vehicle. Furthermore, we assume that the service time for a demand call from region j by a vehicle from location i is exponentially distributed with mean that is equal to the sum of mean travel time to demand region, mean incident handling time and mean travel time back to vehicle location ($\omega_{ij} + \phi_j + \omega_{ij}$).

In the exact queueing model, the closest available vehicle is assigned to a demand call. To approximate it with interdependent multiple queueing models, we need to incorporate the busyness of the vehicles (sum of the steady-state probabilities of the states where the vehicle is busy) into the queueing models. So, the demand arrival rate from a certain region for a specific vehicle should differ concerning the rank of this vehicle among other vehicles in proximity to that region. This means that the demand arrival rate from certain region for a specific queueing model depends on the sum of a set of the steady-state probabilities of other queueing models where the corresponding vehicles are closer to that region.

We present three variations of the decomposition method. In the first method, at most one vehicle is allowed at a single location. In the second method, we improve the calculation of demand rates by addressing the dependency between busy probabilities of vehicles in the exact queueing model while allowing again at most one vehicle at a single location. In the third method, multiple vehicles are allowed at a single location and dependency between vehicles is addressed in calculating demand rates.

4.2.1 DM-S : single vehicle only case

We first focus on the case where at most one vehicle is allowed to be located at a single location as in mathematical model P_S. Let $I^{'}$ be the set of vehicle locations with one vehicle located under the given solution $\vec{x}$, $I^{'} = \{i \in I | x_i > 0\}$. We construct N interdependent queueing models, denoted by QM_i, for $\forall i \in I^{'}$.

Let $\{b_t, t \geq T\}$ be a continuous time Markov chain with state space L_i for QM_i. A state $b_t \in L_i$, $t \geq T$ is 0 when the vehicle is free and j when it is busy serving region j at time t. Then, $L_i = J \cup \{0\}$ for QM_i, $\forall i \in I^{'}$.

Let π^i_j be the steady state probability for state $j \in L_i$ of QM_i. Then, π^i_0 denotes the probability that vehicle i is free.

Let S^i_j be the set of vehicles that are closer to region j than i, $S^i_j = \{k \in I' | \omega_{kj} < \omega_{ij}\}$. We can find the probability that all vehicles closer to region j than vehicle i are busy by using the set S^i_j. Let c^i_j be the probability that all vehicles closer to region j than vehicle i being busy. Then,

$$c^i_j = \begin{cases} 1 & \text{if } S^i_j = \emptyset, \\ \prod_{k \in S^i_j} (1 - \pi^k_0), & \text{otherwise.} \end{cases} \tag{4.7}$$

In the exact queueing model, the probability that all the closer vehicles are busy may not be equal to the multiplication of busy probabilities of individual vehicles. Hence we note that servers are assumed independent in building this probability, c^i_j, whereas this may not be justified depending on the busyness of the system.

The transition from state $j = 0$ to state $j \in L_i \setminus \{0\}$ is realized when a demand call arrives from region j. For QM_i, the transition rate from state i to state j is equal to the demand rate of region j if vehicle i is the closest vehicle to this region. Otherwise, the transition rate is calculated based on both the demand rate of the region and the busy probabilities of vehicles that are closer to this region than vehicle i.

Accordingly, the transition rate from state $j = 0$ to state $j \neq 0$ for QM_i where a demand call arrives is $\lambda_j c^i_j$.

The service rate for state $j \neq 0$ for QM_i, i.e. for region j, when served by vehicle i, is

$$\mu_{ji} = \begin{cases} \dfrac{60}{\phi_j} & \text{if } j = i, \\ \dfrac{60}{2\omega_{ij} + \phi_j}, & \text{otherwise.} \end{cases} \tag{4.8}$$

Example: Assume $\vec{x} = (1, 0, 1)$ where $I = \{1, 2, 3\}$, $J = \{1, 2, 3\}$ and $N = 2$. We construct a queueing model for the vehicle in Region 1 as QM_1 and another for the one in Region 3 as QM_3. Then, $L_1 = \{0, 1, 2, 3\}$ for QM_1 and $L_3 = \{0, 1, 2, 3\}$ for QM_3.

Let $\pi^1 = (\pi_0^1, \pi_1^1, \pi_2^1, \pi_3^1)$ be the vector of the steady state probabilities for QM_1, and $\pi^3 = (\pi_0^3, \pi_1^3, \pi_2^3, \pi_3^3)$ for QM_3 .

Assume that Region 2 is closer to Region 1 than Region 3. So, a demand call from Region 2 is served by vehicle at Region 3 only if the vehicle at Region 1 is busy. Then, the queueing models could be constructed as in Figure 4.1 where $(1 - \pi_0^1)$ and $(1 - \pi_0^3)$ are the probabilities that the vehicle located at Region 1 and Region 3 are busy, respectively.

The balance equations for QM_1 are as follows:

$$\pi_0^1\lambda_1 = \pi_1^1\mu_{11},$$
$$\pi_0^1\lambda_2 = \pi_2^1\mu_{21},$$
$$\pi_0^1\lambda_3(1 - \pi_0^3) = \pi_3^1\mu_{31}.$$

The balance equations for QM_3 are as follows:

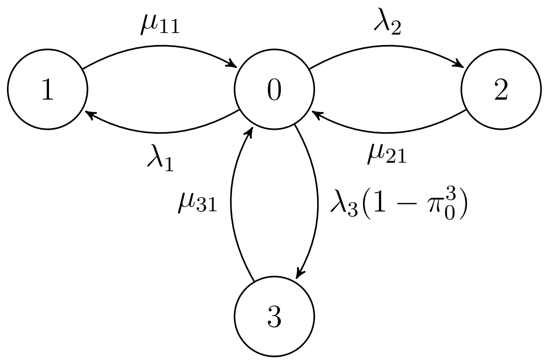
$$\pi_0^3\lambda_1(1 - \pi_0^1) = \pi_1^3\mu_{13},$$
$$\pi_0^3\lambda_2(1 - \pi_0^1) = \pi_2^3\mu_{23},$$
$$\pi_0^3\lambda_3 = \pi_3^3\mu_{33}.$$

By using balance equations for QM_1 and $\pi_0^1 + \pi_1^1 + \pi_2^1 + \pi_3^1 = 1$, we reformulate steady state probabilities as follows:

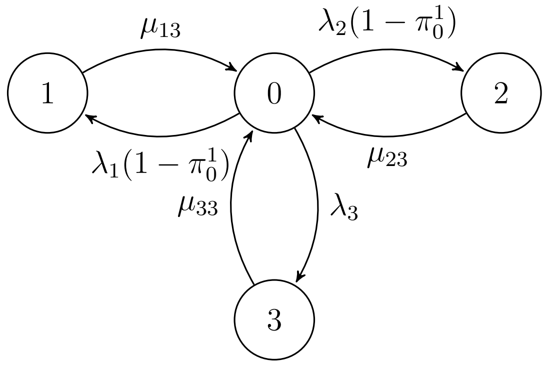
$$\pi_0^1 + \pi_0^1\frac{\lambda_1}{\mu_{11}} + \pi_0^1\frac{\lambda_2}{\mu_{21}} + \pi_0^1\frac{\lambda_3(1 - \pi_0^3)}{\mu_{31}} = 1$$

which is equal to

$$\pi_0^1\left(\mu_{11}\mu_{21}\mu_{31} + \lambda_1\mu_{21}\mu_{31} + \lambda_2\mu_{11}\mu_{31} + \lambda_3(1 - \pi_0^3)\mu_{11}\mu_{21}\right) = \mu_{11}\mu_{21}\mu_{31}.$$



(a) Rate diagram for QM_1



(b) Rate diagram for QM_3

Figure 4.1: Rate diagrams for the queueing models of the given solution $\vec{x}$

Then, the probability that the vehicle at Region 1 is free is found as

$$\pi_0^1 = \frac{\prod_{j\in\{1,2,3\}} \mu_{j1}}{\sum_{j\in\{1,2,3\}} c_j^1 \lambda_j \prod_{k\in\{1,2,3\},k\neq j} \mu_{k1} + \prod_{j\in\{1,2,3\}} \mu_{j1}} \tag{4.9}$$

where $c_j^1 = (1-\pi_0^3)$ is the probability that all vehicles closer to region i than location 1 is busy if S_j^1 is not an empty set.

By dividing numerator and denominator of (4.9) by $\prod_{j\in L_1\setminus\{0\}} \mu_{j1}$, we have

$$\pi_0^1 = \frac{1}{\sum_{j\in\{1,2,3\}} \frac{c_j^1\lambda_j}{\mu_{j1}} + 1}.$$

Then, (4.10) and (4.11) give the steady state probabilities of states of QM_i in general form:

$$\pi_0^i = \frac{1}{\sum\limits_{j \in L_i \backslash \{0\}} \frac{c_j^i \lambda_j}{\mu_{ji}} + 1}, \tag{4.10}$$

$$\pi_j^i = \frac{c_j^i \lambda_j}{\mu_{ji} \left(\sum\limits_{k \in L_i \backslash \{0\}} \frac{c_k^i \lambda_k}{\mu_{ki}} + 1 \right)}, j \in L_i \backslash \{0\}. \tag{4.11}$$

Notice that (4.10) and (4.11) have the term c_j^i which includes the steady state probabilities of QM_3. Therefore, the set of non-linear equations for the example that needs to be solved to find the steady-state probabilities is:

$$\pi_0^1 + \frac{\lambda_1}{\mu_{11}} \pi_0^1 + \frac{\lambda_2}{\mu_{21}} \pi_0^1 + \frac{\lambda_3 (1 - \pi_0^3)}{\mu_{31}} \pi_0^1 = 1, \tag{4.12}$$

$$\pi_0^3 + \frac{\lambda_1 (1 - \pi_0^1)}{\mu_{13}} \pi_0^3 + \frac{\lambda_2 (1 - \pi_0^1)}{\mu_{23}} \pi_0^3 + \frac{\lambda_3}{\mu_{33}} \pi_0^3 = 1, \tag{4.13}$$

where $0 \leq \pi_0^1 \leq 1$ and $0 \leq \pi_0^3 \leq 1$.

It is seen that, the number of unknowns and non-linear equations that needs to be solved are equal to the number of vehicles located. (4.12) and (4.13) could be written in general form as follows:

$$\pi_0^i + \sum_{j \in L_i \backslash \{0\}} \frac{\lambda_i}{\mu_{ji}} c_j^i \pi_0^i = 1, \qquad \forall i \in I'. \tag{4.14}$$

In the next section, the calculation of demand rates are improved by changing the estimation of the busy probability of vehicles closer than vehicle i to region j, c_j^i.

4.2.2 DM-S-CF: single vehicle only case with correction factors

To estimate the EMS system's performance measures, we decompose the exact queueing model into separate queueing models, one for each vehicle, in DM-S. In order to imitate the exact queueing model under solution $\vec{x}$, we use busy probabilities of other vehicles in the demand rate of the region j for QM_i since vehicle i responds to a demand call only if all vehicles that are closer to region j than vehicle i are busy.

We calculate the probability that all vehicles closer to region j than vehicle i is busy by $\prod_{k \in S_j^i} (1 - \pi_0^k)$. This assumes that servers are independent; hence we could multiply the busy probabilities of vehicles to find the probability that these vehicles are busy at the same time. However, in the exact queueing model, this probability is not equal to the multiplication of individual busy probabilities since servers are not independent.

To overcome this bias in the calculation of this probability, Larson (1975) proposes correction factors based on M/M/N/(.) queueing systems.

Assume the general M/M/N/∞ queueing system with demand rate λ and service rate μ per server and $\rho = \lambda/N\mu < 1$. Let U_k indicate the state where k servers are busy.

For M/M/N/∞ queueing system, we know that

$$\begin{aligned}
P(U_0)\ &= P_0 = \frac{1}{\sum\limits_{i=0}^{N-1} \frac{N^i\rho^i}{i!} + \frac{N^N\rho^N}{N!(1-\rho)}}, \\
P(U_k)\ &= P_k = \frac{N^k\rho^k}{k!}P_0, \qquad k = 1, 2, \ldots, N-1, \\
P(U_N) &= P_N = \frac{N^N\rho^N}{N!(1-\rho)}P_0.
\end{aligned}$$

Suppose, we randomly sample servers until we find the first server which is free. Let W_j be the event that j^{th} selected server is busy and $A_j = W_j^c$ be the event that j^{th} selected server is available. We derive an expression for $P(W_1W_2...W_jA_{j+1})$.

By conditioning,

$$P(W_1W_2...W_jA_{j+1}) = \sum_{k=0}^{N} P(W_1W_2\ldots W_jA_{j+1}|U_k)P_k \tag{4.15}$$

and

$$\begin{aligned}
&P(W_1W_2\ldots W_jA_{j+1}|U_k) \\
&\qquad = P(A_{j+1}|W_1W_2\ldots W_jU_k)P(W_j|W_1W_2...W_{j-1}U_k)...P(W_1|U_k).
\end{aligned} \tag{4.16}$$

We know that $P(W_1|U_k) = \frac{k}{N}$, $P(W_2|W_1U_k) = \frac{k-1}{N-1}$.

In general,

$$P(W_i|W_1W_2\ldots W_{i-1}U_k) = \frac{k-(i-1)}{N-(i-1)},\ i = 1, 2, ..., k+1. \tag{4.17}$$

Similarly,

$$P(A_{j+1}|W_1W_2\ldots W_jU_k) = \frac{N-k}{N-j},\ j = 0,1,2,...,k. \tag{4.18}$$

Then, we get the following equation by inserting (4.16), (4.17) and (4.18) into (4.15):

$$P(W_1W_2\ldots W_jA_{j+1}) = \sum_{k=j}^{N-1}\left[\frac{k}{N}\right]\left[\frac{k-1}{N-1}\right]...\left[\frac{k-(j-1)}{N-(j-1)}\right]\left[\frac{N-k}{N-j}\right]P_k \tag{4.19}$$

When we insert P_k into (4.19) and reorder the equation, we have

$$P(W_1W_2\ldots W_jA_{j+1}) = \left[\sum_{k=j}^{N-1}\left(\frac{(N-j-1)!(N-k)}{(k-j)!}\right)\left(\frac{N^k}{N!}\right)\rho^{k-j}\right]\boldsymbol{\rho^j(1-\rho)}\left[\frac{P_0}{(1-\rho)}\right]. \tag{4.20}$$

As it is seen, we now have $\rho^j(1-\rho)$ in (4.20) and other terms in the form a multiplier which reflects the correction factor for the probability that $j+1^{st}$ vehicle is available when first j is busy where independence among servers is not implied. Hence, this multiplier, $C(N,\rho,j)$, could be used as a correction factor in the calculation of the probability of the event that all j vehicles which are closer to a region than the vehicle in QM_i are busy.

$$P(W_1W_2\ldots W_jA_{j+1}) = C(N,\rho,j)\boldsymbol{\rho^j(1-\rho)}$$

where

$$C(N,\rho,j) = \left[\sum_{k=j}^{N-1}\left(\frac{(N-j-1)!(N-k)}{(k-j)!}\right)\left(\frac{N^k}{N!}\right)\rho^{k-j}\right]\left[\frac{P_0}{(1-\rho)}\right],\ j = 0,1,...,N-1.$$

Until now, we work on an M/M/N/∞ system. For the system M/M/N/N (no queueing), the actual fraction of time $\rho^{'}$ that a server is busy is smaller than $\rho = \lambda/N\mu$ since calls are lost if all servers are busy. For M/M/N/N,

$$\rho^{'} = \rho(1-P_N^{'})$$

where

$$P(U_0)= P_0^{'} = \frac{1}{\sum_{i=0}^{N} \frac{N^i \rho^i}{i!}},$$
$$P(U_k)= P_k^{'} = \frac{N^k \rho^k}{k!} P_0^{'}, \qquad k = 0, 1, ..., N.$$

We can derive the new correction factor as $C^{'}(N, \rho, j) = P(W_1 W_2 \ldots W_j A_{j+1})/(\rho^{'})^j (1 - \rho^{'})$ by substituting ρ with $\rho^{'}/(1 - P_N^{'})$ to obtain the term $(\rho^{'})^j (1 - \rho^{'})$ in $P(W_1 W_2 \ldots W_j A_{j+1})$. Then,

$$P(W_1 W_2 \ldots W_j A_{j+1}) = \left[\sum_{k=j}^{N-1} \frac{(N - j - 1)!(N - k)}{(k - j)!} \left(\frac{N^k}{N!} \right) \rho^{k-j} \right] \left(\frac{1}{1 - P_N^{'}} \right)^j \left(\frac{P_0}{1 - \rho(1 - P_N^{'})} \right) \boldsymbol{(\rho^{'})^j (1 - \rho^{'})}.$$

Hence, the correction factor $C^{'}(N, \rho, j)$ can be set as follows:

$$C^{'}(N, \rho, j) = \left[\sum_{k=j}^{N-1} \frac{(N - j - 1)!(N - k)}{(k - j)!} \left(\frac{N^k}{N!} \right) \rho^{k-j} \right] \left(\frac{1}{1 - P_N^{'}} \right)^j \left(\frac{P_0}{1 - \rho(1 - P_N^{'})} \right).$$

Eventually, the correction factor $C^{'}(N, \rho, j)$ is used in the calculation of the probability c_j^i to overcome the bias due to the assumption of servers being independent. Let $c_j^{'i}$ be an alternative to c_j^i in (4.7) as follows:

$$c_j^{'i} = \begin{cases} 1 & \text{if } S_j^i = \emptyset, \\ C^{'}(N, \rho, |S_j^i|) \prod_{k \in S_j^i} (1 - \pi_0^k), & \text{otherwise.} \end{cases} \tag{4.21}$$

where $|S_j^i|$ is the size of set S_j^i , and ρ is equal to the total demand rate over total average service rate μ_{avg} (which is calculated from QM_i's): $\frac{\sum_{j \in J} \lambda_j}{N \mu_{avg}}$.

Hence, we rewrite the set of equations in (4.14) that need to be solved simultaneously to find the steady-state probabilities of QM_i's according to DM-S-CF decomposition method as follows:

$$\pi_0^i + \sum_{j \in L_i \backslash \{0\}} \frac{\lambda_i}{\mu_{ji}} c_j^{'i} \pi_0^i = 1, \qquad \forall i \in I^{'}. \tag{4.22}$$

4.2.3 DM-M-CF: multiple vehicle case with correction factors

In DM-S-CF method, at most one vehicle is allowed in a single location. We extend the DM-S-CF decomposition method in a way to handle the case where multiple vehicles to be located at a single location. This would change the calculation of the probability that all vehicles closer than the vehicle in QM_i to a region are busy.

In this method, the interdependent queueing models are constructed differently. In the previous methods, each QM_i is constructed for only one vehicle. Since we now allow more than one vehicle at a single location, if we continue with one queueing model for each vehicle, some queueing models would be very similar in transition rates. We could allow some models to include more than one server; however, this increases the size of the state space of the model exponentially similar to the exact queueing model. Instead of constructing queueing models with more than one vehicle, we stick to the previous structure by giving pseudo orders to the vehicles in the same location. So, an interdependent queueing model is constructed for each vehicle while paying attention to the order of vehicles at the same locations. The representation of multiple vehicles in this method also allows one to observe the change in the busyness of a vehicle when an additional vehicle located at the same location.

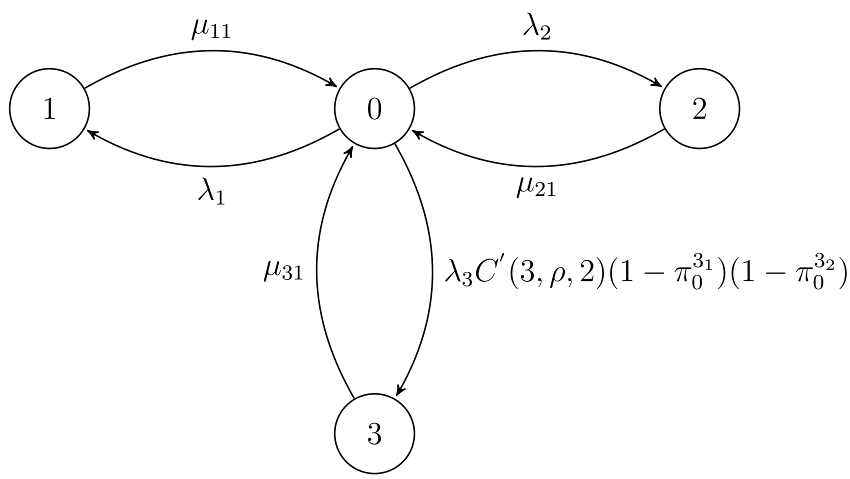
Following from the previous example, we locate another vehicle at Region 3 and change $\vec{x}$ as follows: $\vec{x} = (1, 0, 2)$ where $I = \{1, 2, 3\}$, $J = \{1, 2, 3\}$ and $N = 3$. The definition of set $I^{'}$ is redefined for DM-M-CF as follows:

$$I^{'} = \{i_k | i \in I, x_i > 0, k = 1, \ldots, x_i\} .$$

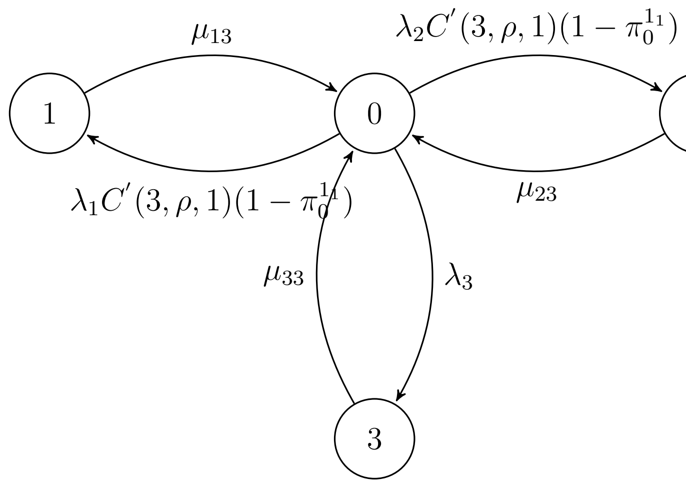
Then, $I^{'} = \{1_1, 3_1, 3_2\}$ under $\vec{x}$. We construct a queueing model for each vehicle in $I^{'}$: QM_{1_1} for the vehicle at Region 1 and QM_{3_1}, QM_{3_2} for the vehicles in Region 3. Then, the state space for all 3 queueing models are the same and equal to $\{0, 1, 2, 3\}$.

Let π^i_j, $\forall i \in I^{'}$, $\forall j \in L_i$ be the steady-state probabilities for QM_i's and assume that Region 2 is closer to Region 1 than Region 3. Then, the queueing models could be constructed as in Figure 4.2.

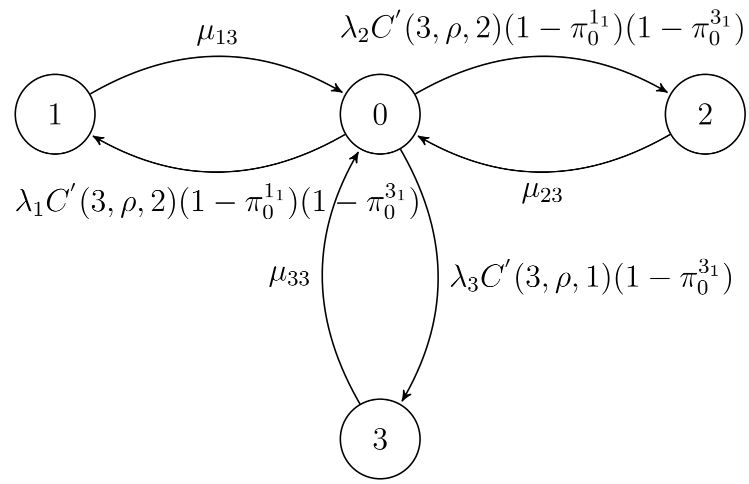
As seen in the figure, the calculation of the probability that all vehicles closer to region j than vehicle i are busy is changed since we have more than one vehicle



(a) Rate diagram for QM_{1_1}



(b) Rate diagram for QM_{3_1}



(c) Rate diagram for QM_{3_2}

Figure 4.2: Rate diagrams for the queueing models of the given solution $\vec{x}$

at a location now. For the vehicle locations with more than one vehicle, the index of the vehicle location in set $I^{'}$ defines the preference order of the vehicles in this location. Following the example, 3_2 responds a demand call from Region 2 only if vehicle 1_1 and 3_1 is busy. Hence, the transition rate from state $j = 0$ to state $j = 2$ is $\lambda_2 C^{'}(3, \rho, 2)(1 - \pi_0^{1_1})(1 - \pi_0^{3_1})$ for QM_{3_2}.

Accordingly, we redefine the set S_j^i as the set of vehicle that are closer to region j than vehicle i or the vehicles that are preferred to vehicle i if there are more than one vehicle at the location of i. Let $Z(k)$ is equal to l which is the pseudo order of $k \in \{i_l | i \in I, x_i > 0, l = 1, ..., x_i\}$. Then, we define set S_j^i as follows: $S_j^i = \{k \in I' | \omega_{kj} < \omega_{ij} \vee Z(k) < Z(i)\}$.

The set of non-linear equations that need to be solved simultaneously to find the steady-state probabilities of QM_i's is the same as the one for DM-S-CF in (4.22) where the definition of $c_j^{'i}$ is now based on the set $S_j^i = \{k \in I' | \omega_{kj} < \omega_{ij} \vee Z(k) < Z(i)\}$.

$$\pi_0^i + \sum_{j \in L_i \setminus \{0\}} \frac{\lambda_i}{\mu_{ji}} c_j^{'i} \pi_0^i = 1, \qquad \forall i \in I^{'} \tag{4.23}$$

where $I^{'} = \{i_k | i \in I, x_i > 0, k = 1, \ldots, x_i\}$.

4.2.4 Approximation method for the steady-state probabilities

The proposed decomposition methods require a set of non-linear equations to be solved to find the steady-state probabilities of the queueing models constructed for vehicles.

Let a_{ji} be equal to $\frac{\lambda_i}{\mu_{ji}}$ and $\mathbf{\Pi}$ be the vector of π_0^i's. By inserting a_{ji} in 4.23, we write N functions that needs to solved simultaneously with N unknowns for DM-M-CF which is the most general decomposition method allowing multiple vehicles in a single location and incorporating correction factors for the busy probabilities as follows:

$$f_i(\mathbf{\Pi}) = \pi_0^i + \sum_{j \in L_i \setminus \{0\}} a_{ji} c_j^{'i} \pi_0^i - 1, \qquad \forall i \in I^{'} \tag{4.24}$$

where $I^{'} = \{i_k | i \in I, x_i > 0, k = 1, \ldots, x_i\}$

Then, we write N equalities that defines a system of non-linear equations for DM-M-CF decomposition method:

$$f_i(\mathbf{\Pi}) = 0, \qquad \forall i \in I^{'}, \quad \mathbf{\Pi} \in [0,1]^N . \tag{4.25}$$

Let $\mathbf{\Pi}^*$ be the solution to the system of non-linear equations defined in (4.25). To prove that such $\mathbf{\Pi}^*$ exists, we use Poincaré-Miranda Theorem by Kulpa (1997).

Theorem 1 (Poincaré-Miranda Theorem)**.** *Let $I^n := [0,1]^n$ be the n-dimensional cube of the Euclidean space R^n and let δI^n be its boundary. For each $i \leq n$ let us denote*

$$I_i^- = \{x \in I^n : x(i) = 0\}, I_i^+ = \{x \in I^n : x(i) = 1\} \tag{4.26}$$

the i-th opposite faces.

Let $f : I^n \to R^n$, $f = (f_1, ..., f_n)$, be a continuous map such that for each $i \leq n$, $f_i(I_i^-) \subset (-\infty, 0]$ and $f_i(I_i^+) \subset [0, \infty)$. Then, there exists a point $c \in I^n$ such that $f(c) = 0$.

Proof. We show the properties of the function f_i introduced in (4.24).

Let $\mathbf{0}$ and $\mathbf{1}$ be the vector of zeroes and ones which correspond to I_i^- and I_i^+, respectively.

By (4.24), we have

$$f_i(\mathbf{0}) = -1, \qquad \forall i \in I^{'}, \tag{4.27}$$

and,

$$f_i(\mathbf{1}) = \sum_{j \in L_i \backslash \{0\}} a_{ji} c_j^{'i}, \qquad \forall i \in I^{'}. \tag{4.28}$$

We know that $a_{ji} = \frac{\lambda_j}{\mu_{ji}}$, making $0 < a_{ji}$, $\forall j \in L_i \backslash \{0\}$, $\forall i \in I^{'}$ by definition.

Recall

$$c_j^{'i} = \begin{cases} 1 & \text{if } S_j^i = \emptyset, \\ C^{'}(N, \rho, |S_j^i|) \prod_{k \in S_j^i} (1 - \pi_0^k), & \text{otherwise} \end{cases} \tag{4.29}$$

where $S_j^i = \{k \in I' | \omega_{kj} < \omega_{ij} \vee Z(k) < Z(i)\}$.

By (4.29), $c_j^{'i} = 1, \; \forall (i,j) \in \left\{(k,l) | k \in L_i \backslash \{0\}, \forall l \in I', S_k^l = \emptyset\right\}$ and $c_j^{'i} = 0, \; \forall (i,j) \in \left\{(k,l) | k \in L_i \backslash \{0\}, \forall l \in I', S_k^l \neq \emptyset\right\}$ since $\pi_0^i = 1, \; \forall i \in I'$. Hence, the summation in (4.28) is positive and $f_i(\mathbf{1}) \geq 0, \forall i \in I'$.

Therefore, we show that $f_i(I_i^-) \subset (-\infty, 0]$ and $f_i(I_i^+) \subset [0, \infty)$, $\forall i \in I'$ where $I_i^- = \mathbf{0}$ and $I_i^+ = \mathbf{1}$.

Then, there exist a solution $\mathbf{\Pi}^*$ such that $\mathbf{f}(\mathbf{\Pi}^*) = \mathbf{0}$ by Poincaré-Miranda Theorem.

□

$\mathbf{\Pi}^*$ is said to be a nonsingular if the associated Jacobian matrix for the set of functions in (4.24) is nonsingular at $\mathbf{\Pi}^*$: det $\mathbf{f}'(\mathbf{\Pi}^*) \neq 0$. However, to be able to say that there exist only one solution to this system of equations, one needs to show that det $\mathbf{f}'(\mathbf{\Pi}) \neq 0,$ for all $\mathbf{\Pi} \in [0,1]^N$.

As we show that there exists at least one solution to the non-linear system of equations, we propose an iterative algorithm to solve the system of equations. In every iteration of the algorithm, we construct the queueing models by using the busy fractions $(1 - \pi_0^i)$ from the previous iteration. Then, we solve the balance equations for each queueing model separately and get closer to the solution of the non-linear set of equations in (4.24). The algorithm continues until a convergence criterion is satisfied.

In the approximation of $\mathbf{\Pi}^*$, the correction factor $C'(N, \rho, .)$ is set to 1 initially for DM-S-CF and DM-M-CF, then it is updated after each iteration based on the average service rate, μ_{avg}, found from the steady-state probabilities of QM_i's. For the DM-S method, the algorithm could be used by just disregarding the correction factors.

The pseudo code in Algorithm 3 summarizes how the steady-state probabilities, $\mathbf{\Pi}^*$, are approximated.

Algorithm 3 Algorithm to approximate steady-state probabilities with correction factors

initialize iteration counter: $iter = 0$
initialize $_{iter}C^{'}(N, \rho, \left|S_j^i\right|) = 1,\ \forall i \in I^{'},\ \forall j \in L_i \backslash \{0\}$
initialize $_{iter}P_0^i = 0,\ \forall i \in I^{'}$
construct individual queueing models, $QM_i,\ \forall i \in I^{'}$ by using $_{iter}C^{'}(N, \rho, \left|S_j^i\right|)$ and using $(1 - {}_{iter}P_0^i)$ as the busy probability for vehicle i
$iter = 1$
solve balance equations and find $\pi_j^i,\ \forall j \in L_i$ for each QM_i
set $_{iter}P_0^i = \pi_0^i,\ \forall i \in I^{'}$
while $\max_{i \in I'} \left\{ \left| {}_{iter}P_0^i - {}_{(iter-1)}P_0^i \right| \right\} >= 0.001$ **do**
    calculate ρ and $_{iter}C^{'}(N, \rho, \left|S_j^i\right|)$ based on $_{iter}P_0^i,\ \forall i \in I^{'},\ \forall j \in L_i \backslash \{0\}$
    construct individual queueing models, $QM_i,\ \forall i \in I^{'}$ by using $_{iter}P_0^i$
    solve balance equations and find $\pi_j^i,\ \forall j \in L_i$ for each QM_i
    $iter = iter + 1$
    set $_{iter}P_0^i = \pi_0^i$
end while
return $\pi_j^i,\ \forall j \in L_i,\ \forall i \in I^{'}$

4.2.5 Calculation of the performance measures

Once steady-state probabilities of QM_i's are evaluated, several performance measures for an EMS system could be estimated based on those. The mean response time for the system is used as the objective function value in the mathematical models in Section 4.1. In addition to this, three other performance measures that could be obtained through steady-state probabilities of QM_i's are listed below:

- $ED(\vec{x})$: expected satisfied demand in unit time,
- $CD(\vec{x})$: expected satisfied demand under threshold time τ,
- $R_j(\vec{x})$: mean region-wise response time of region j.

Let $ED(\vec{x})$ be the expected satisfied demand in unit time. $ED(\vec{x})$ is found by using the number of dispatches of vehicles to regions in unit time which is calculated based on steady state probabilities and transition rates of QM_i's.

The number of times vehicle i is dispatched to region j in unit time is equal to the steady state probability of vehicle i being free, π_0^i, multiplied with the transition rate to state j, that is $c_j^i\lambda_j$. So, the number of dispatches in unit time from i to j is $\pi_0^i c_j^i \lambda_j$.

Then, $ED(\vec{x})$ is equal to the number of dispatches from vehicle i to region j summed over all vehicles and regions,

$$ED(\vec{x}) = \sum_{j\in J}\sum_{i\in I'} \pi_0^i c_j^i \lambda_j. \tag{4.30}$$

Similarly, it is possible to estimate the expected covered demand $CD(\vec{x})$ of the system under $\vec{x}$. Recall that T_{ij} is a random variable representing travel time between location i and region j and $F_{T_{ij}}$ is the cumulative distribution function of random variable T_{ij}. The expected number of demand calls satisfied in unit time that is covered under threshold τ from region j by vehicle i is given by $F_{T_{ij}}(\tau)\pi_0^i c_j^i \lambda_j$. Then, the expected covered demand under $\vec{x}$ is found as

$$CD(\vec{x}) = \sum_{j\in J}\sum_{i\in I'} F_{T_{ij}}(\tau)\pi_0^i c_j^i \lambda_j. \tag{4.31}$$

Another measure that can be evaluated based the number of dispatches is mean region-wise response time, $R_j(\vec{x})$, under solution $\vec{x}$. The region-wise mean response time for region j is the weighted average of mean travel times to region j based on the number of dispatches under $\vec{x}$.

$$R_j(\vec{x}) = \frac{\sum\limits_{i\in I'} \pi_0^i c_j^i \lambda_j \omega_{ij}}{\sum\limits_{i\in I'} \pi_0^i c_j^i \lambda_j}, \ \forall j \in J, \tag{4.32}$$

where numerator is the weighted total travel time based on the number of dispatches and the denominator is the total number of dispatches to region j in unit time.

Similar to mean region-wise response time, the mean response time of the system, $R(\vec{x})$, is evaluated as follows:

$$R(\vec{x}) = \frac{\sum\limits_{j\in J}\sum\limits_{i\in I'} \pi_0^i c_j^i \lambda_j \omega_{ij}}{\sum\limits_{j\in J}\sum\limits_{i\in I'} \pi_0^i c_j^i \lambda_j}. \tag{4.33}$$

$R(\vec{x})$ is used in P_S and P_M, however; performance measures can be diversified such as $E(\vec{x})$, $C(\vec{x})$ and $R_j(\vec{x})$ which are estimated based on QM_i's.

In the next section, a meta-heuristic solution algorithm is presented to be able to find near optimal solutions for the mathematical models, P_S and P_M where $R(\vec{x})$ is estimated based on the steady-state probabilities and transition rates of QM_i's by using the approximation algorithm in Section 4.2.4.

4.3 A Genetic Algorithm for P_M

In Section 4.1, we present two mathematical models, P_S and P_M, to locate EMS vehicles whose objective functions are evaluated with the help of decomposition methods proposed in Section 4.2. Since there is no closed-form formulation for the objective function, a genetic algorithm (GA) is constructed to find near-optimal solutions for P_M while evaluating the objective function values with the decomposition methods.

Genetic algorithms use chromosome structure to encode different solutions and compares their fitness function values, i.e. objective function values.They are designed to generate a population of initial solutions and evolve toward better ones in terms of the fitness function value. Evolution is realized through the reproduction of the population using two main genetic operators, crossover and mutation operator, which create the next generation for the algorithm.

A feasible solution to model P_M is the number of vehicles located in given locations. In the chromosome structure, a solution is represented by an N-dimensional array. Each entry of this array is called a gene and shows the location of a vehicle in the solution. Since P_M allows multiple vehicles in one location, genes representing the same location might show up in chromosomes more than once. The number of genes with the same location indicates the number of vehicles located in that location. For an instance where $N = 3$, the solution at hand $\vec{x} = (1, 0, 2)$ and $I = \{1, 2, 3\}$, chromosome is encoded as $\langle 1|3|3 \rangle$. Each chromosome is called an individual, the collection of which makes up the population of the algorithm.

The notation used for the genetic algorithm are given in Table 4.1.

Table 4.1: Notation used for Genetic Algorithm

Parameters	
M	Size of the population and the mating pool
O_i	Fitness value of individual i of the population
$u_s(i)$	Probability of selection of individual i
u_c	Probability of crossover
c	Crossover point
u_m	Probability of mutation

M many random individuals are generated to form the initial population. Each individual represents a feasible solution for the model to be solved. A predefined number of individuals, M, are copied to a mating pool to reproduce the next generation population in the reproduction process. The selection of the individuals is based on their fitness values. A larger probability is assigned to an individual with a smaller fitness value.

The fitness value, O_i, of an individual i is simply the objective function value for the solution encoded in the chromosome. The DM-M-CF decomposition method is used to find the objective function value of individual and is recorded as the fitness value.

The probability of selection for individual i is given by:

$$u_s(i) = \frac{\frac{1}{O_i}}{\sum_{j=1}^{S} \frac{1}{O_j}}.$$

After constructing the mating pool with individuals from the population, parents are selected in pairs for reproduction. This selection is random with equal probabilities for all individuals in the mating pool.

The crossover operator is used to transfer genes from parents to children and applied to selected pairs of individuals (as parents) with a predefined probability u_c. If the probability fails, crossover operation is not applied, the parents are duplicated, i.e., children are the same as the parents. If crossover were applied, one-point crossover is used. A crossover point, c, is selected randomly between 1 and N. The first c genes of Parent 1 are copied to Child 1 and Parent 2 to Child 2. Genes after the crossover point, c, are copied from Parent 1 to Child 2 and from Parent 2 to Child 1.

As an example consider Parent 1, Parent 2 which are given below and $c = 2$. Child 1 and Child 2 are reproduced with the crossover as follows;

Parent 1 Parent 2

$\langle 1|1|3 \rangle$ $\langle 2|3|5 \rangle$

Child 1 Child 2

$\langle 1|1|5 \rangle$ $\langle 2|3|3 \rangle$

After reproducing children from crossover operation, every gene of a child is mutated to diversify the solutions in the population and better search the solution space by not restricting the search to solutions only with genes represented in a generation. The mutation is realized with probability u_m for every gene in a child. If probability succeeds, the child's gene is overridden by a random location from set I with equal probabilities (with locations that do not exist in the child if only single vehicles are allowed in a location).

After crossover and mutation operations, $2 * M$ individuals exist in the mating pool, including the children. The next generation is constructed from the best M feasible individuals in terms of the fitness function value from the mating pool. When the next generation produced from a population of individuals consists of a single chromosome represented M times, it is assumed that the population has converged. This individual is taken as the best solution suggested by the genetic algorithm for P_M , and iteration is terminated.

This genetic algorithm could be used for P_S as well by changing the crossover and mutation operators to maintain feasibility in the generation of the next population. The pseudo-code of the GA for P_M is given in Algorithm 4.

Algorithm 4 Pseudo-code of the genetic algorithm

Generate initial population with M many solutions, $Pop = \{p_1, ..., p_M\}$
Find fitness values of the solutions in initial population
repeat
Initiate mating pool, $Pool = \emptyset$
for $k = 1$ to M **do**
$u = rand(0, 1)$
$Pool = Pool \cup \{p_i\}$ where $p_i \in Pop$ and $\sum_{j=1}^{i-1} u_s(j) \leq u \leq \sum_{j=1}^{i} u_s(j)$
end for
for $k = 1$ to $M/2$ **do**
Generate $i = rand(1, M)$ and $l = rand(1, M)$
$Parent_1 = m_i$ and $Parent_2 = m_l$
$u = rand(0, 1)$
if $u \leq u_c$ **then**
$Child_1 :=$ Crossover$(Parent_1, Parent_2)$
$Child_2 :=$ Crossover$(Parent_1, Parent_2)$
else
$Child_1 := Parent_1$
$Child_2 := Parent_2$
end if
for $j = 1$ to 2 **do**
for $n = 1$ to N **do**
$u = rand(0, 1)$
if $u \leq u_m$ **then**
$Child_j(n) :=$ Mutation$(Child_j(n))$
end if
end for
end for
Choose the best M solutions from $(Parent_1, ..., Parent_M, Child_1, ..., Child_M)$ to the population Pop
end for
until Termination condition is satisfied **return** The best individual

4.4 Experimental Study

An experimental study is conducted to see the performance of decomposition methods in evaluating objective function values of each feasible solution of the mathematical models, P_S and P_M, given in Section 4.1. The optimal solutions for P_S and P_M based on complete enumeration with decomposition methods are compared to the best solution found based on simulation as well to check the performance under optimization problems. The proposed decomposition methods are used to evaluate two different objective functions from Chapter 3. Lastly, the performance of the proposed genetic algorithm is checked in terms of the quality of solutions reported at the end of runs.

Various problem instances having different network configurations are used in the experiments as in Chapter 3. In addition to toy data, decomposition methods are tested on a real-life data set as well to check the performance of the methods under larger size problems. The analysis performed are explained in detail below.

Three decomposition methods are used to evaluate the objective function value of a feasible solution of the mathematical models, P_S and P_M, given in Section 4.1. A discrete event simulation model is constructed and coded in Matlab environment to simulate the emergency medical systems and evaluate the mean response time of a location solution since the exact queueing model is computationally expensive. Then, the quality of approximation of the decomposition methods is checked by comparing the objective function values of each feasible solution of the problem evaluated with the simulation model and the decomposition methods based on complete enumeration.

Differently from one-to-one comparison of feasible solutions, the quality of the optimal solutions of the mathematical models under decomposition methods found by complete enumeration is also analyzed by comparing them to the best solutions obtained from a ranking selection algorithm that employs the simulation model. Hence, the performance of the decomposition methods under optimization problems is tested.

In addition, the proposed decomposition methods are used to evaluate the objective function values of the models P_4 and P_7 from Chapter 3 which are the models found performing well in terms of equity. The optimal solutions for the problems are found with complete enumeration. Then, the changes in the mean response time, variance of the region-wise mean response time and Gini coefficient are checked similarly for the optimal solutions of the models in order to observe whether similar equity results are achieved when feasible solutions are evaluated with the decomposition methods.

Lastly, a design of experiments is conducted for the genetic algorithm which is proposed to find near-optimal solutions for the mathematical models.

4.4.1 Test bed

For the experiments, the same sets of regions from Chapter 3 are used, which include three forms with 15 demand regions as seen in Figure 4.3.

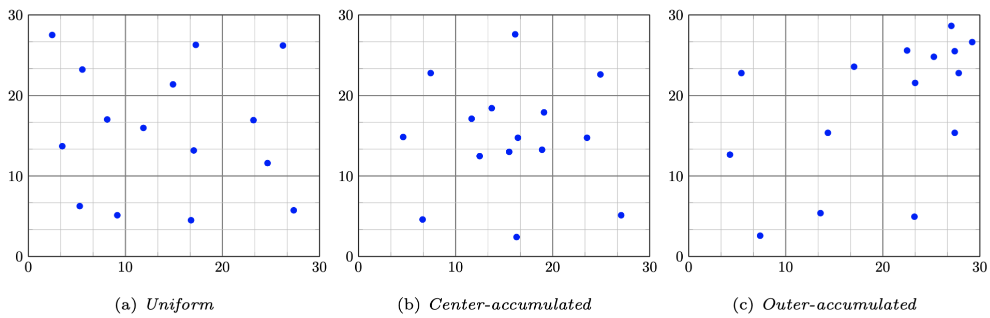


Figure 4.3: Distribution of demand regions

The number of vehicles is set to 4, 5 or 6. The demand rate is assumed to be the same for every region as 0.5 units per hour. The incident handling rate is set to 3, 6, 9 and 12.

In total, 36 instances are generated based on a full factorial design using three factors: *Form* with three levels, *Number of Vehicles* with three levels (4,5 and 6) and *Incident Handling Rate* with four levels (3, 6, 9 and 12).

4.4.2 Selection of the best solution and evaluation of performance measures

Since there is no closed-form formulation for the models that includes stochastic processes in the study, the discrete event simulation model constructed is again used to simulate the emergency medical systems. In order to show the performance of the decomposition methods under optimality, the best solution for each problem instance for models P_S and P_M are found using *KN*++ algorithm by Kim and Nelson (2006) based on this discrete event simulation model as in the previous chapter.

For *KN*++ algorithm, first-stage sample size n_0 is set to 100,000 demand calls. The initial number of batches, b_0, is set to 10. The confidence level γ is set to 0.05, indifference-zone parameter δ to 0.01, and parameter c to 1. *KN*++ algorithm is stopped when the best performing alternative has a 0.1 % difference in the objective function value from the worst performing among the remaining alternatives. Then, the best solution is selected randomly from the remaining alternatives.

4.4.3 Performance of the decomposition methods

The performance of the proposed models is explored in two steps, by analyzing the goodness of the approximation of the objective function value for each feasible solution, and by comparing the optimal solutions of the models obtained using decomposition methods based on complete enumeration to the best solutions obtained with *KN*++.

The mean response times of all feasible solutions for mathematical models are evaluated both with the decomposition methods and the simulation model for each instance. Due to the computational burden, the batch means method is used to estimate the performance measures with a total of 550,000 demand calls where the warm-up period is set to 50,000 calls and ten batches of size 50,000 are used.

Let $O_{(.)}(k)$ be the objective function value evaluated by the decomposition method (.), where (.) is replaced by DM-S, DM-S-CF and DM-M-CF, and $O_{Sim}(k)$ be the mean of the confidence interval of the objective function value evaluated by the simulation model for the solution k of a problem instance under any mathematical model. We find the mean absolute percent deviation (MAPD) of $O_{(.)}(k)$ from $O_{Sim}(k)$ over all feasible solutions of a problem instance as follows:

$$MAPD = \frac{100}{K} \sum_{k=1}^{K} \left| \frac{O_{(.)}(k) - O_{Sim}(k)}{O_{Sim}(k)} \right| \tag{4.34}$$

where K is the total number of feasible solutions compared.

In Table 4.2, we present MAPD results for 36 instances under models using DM-S, DM-S-CF and DM-M-CF methods. Recall that DM-S method is used to evaluate performance measures of EMS systems with at most one vehicle per location, DM-S-CF is an extension of DM-S where we use correction factors in the calculation of the demand rates of the queueing models, and DM-M-CF incorporates correction factors in evaluating the performance measures of EMS systems where multiple vehicles at a single location are allowed . Hence, we use DM-S and DM-S-CF methods to evaluate the objective function values of the solutions of model P_S and DM-M-CF for P_M.

Table 4.2: MAPD of the objective function value under decomposition methods from simulation evaluation overall feasible solutions in 36 instances

			Model P_S		Model P_M
Form	Nb. of Vehc.	Inc. Hand. Rate	DM-S	DM-S-CF	DM-M-CF
Uniform	4	3	6.68	0.36	0.38
		6	8.32	0.48	0.42
		9	8.92	0.63	0.42
		12	9.22	0.76	0.43
	5	3	9.76	0.41	0.57
		6	10.50	0.57	0.70
		9	9.28	0.66	0.74
		12	7.95	0.70	0.76
	6	3	12.00	0.51	0.93
		6	5.48	0.68	1.31
		9	1.35	0.91	1.50
		12	1.33	1.11	1.60
Center-Acc.	4	3	5.83	0.36	0.37
		6	7.17	0.48	0.40
		9	7.54	0.60	0.43
		12	7.65	0.69	0.43

Continued on next page

Table 4.2 – continued from previous page

			Model P_S		Model P_M
Form	Nb.of Vehc.	Incident Hand. Rate	DM-S	DM-S-CF	DM-M-CF
	5	3	8.56	0.43	0.49
		6	8.72	0.61	0.61
		9	7.25	0.69	0.66
		12	5.92	0.75	0.69
	6	3	10.47	0.52	0.73
		6	4.05	0.62	0.97
		9	1.43	0.63	1.12
		12	1.83	0.73	1.23
Outer-Acc.	4	3	7.07	0.42	0.47
		6	8.98	0.50	0.53
		9	9.88	0.61	0.53
		12	10.38	0.77	0.53
	5	3	10.29	0.54	0.75
		6	11.73	0.66	0.93
		9	10.85	0.71	0.99
		12	9.52	0.76	1.03
	6	3	12.84	0.80	1.29
		6	5.88	1.33	1.87
		9	1.95	1.90	2.18
		12	2.99	2.25	2.33
Average			7.49	0.73	0.87

The results show that DM-S evaluates the mean response time under a feasible solution with up to an average of 13% deviation in some instances and with an average of 7.5% over all. However, the quality of approximation of the mean response significantly betters off with the inclusion of correction factors, where MAPD is smaller than 1% in most instances under DM-S-CF and DM-M-CF.

The results show that allowing multiple vehicles at a single location increases the average MAPD from 0.73% to 0.87%. But, the overall MAPD is very low under models with correction factors (DM-S-CF and DM-M-CF).

In order to evaluate the decomposition methods' ability to differentiate the feasible solutions in line with the actual performance and deliver an optimal solution performing close to the best solution with respect to *KN*++, the optimal solution of the mathematical models found by complete enumeration based on the evaluation using decomposition methods and the best solution obtained with *KN*++ is compared.

After finding the best solutions with *KN*++ algorithm and the optimal solutions of the models based on complete enumeration for each instance, the objective function values are estimated by running the discrete event simulation model for ten independent replications. A total of 550,000 demand calls is simulated for every solution at hand in the simulation model. The warm-up period is selected as 50,000 calls, and the objective function values are reported accordingly. The analysis of the performances of the decomposition methods concerning the optimal solutions is done based on this experiment.

Let $\vec{x}^*_{(.)}$ be the solution that gives the minimum mean response time with respect to complete enumeration under decomposition method (.) and $O_{Sim}(\vec{x}^*_{(.)})$ denotes the mean of 90% confidence interval of the objective function value for this solution based on 10 independent replications of the simulation model. Subsequently, let $\vec{x}^*_{KN++}$ represent the best solution evaluated by KN++ algorithm and $O_{Sim}(\vec{x}^*_{KN++})$ denote the mean of the 90% confidence interval of the objective function value for this solution based on 10 independent replications.

The percentage deviation from the optimal solution for an instance is denoted by $\%\Delta^*_{Sim}$ and is found by

$$\%\Delta^*_{Sim} = I_{\left\{\vec{x}^*_{(.)} \neq \vec{x}^*_{Sim}\right\}} \left(100 \frac{O_{Sim}(\vec{x}^*_{(.)}) - O_{Sim}(\vec{x}^*_{KN++})}{O_{Sim}(\vec{x}^*_{KN++})} \right),$$

where I is the indicator function being equal to 0 if the solutions are the same.

$\%\Delta^*_{Sim}$ results are presented in Table 4.3 for 36 instances with respect to DM-S, DM-S-CF and DM-M-CF. $\%\Delta^*_{Sim}$ values are reported as zero if the optimal solution under a model is the same with the best solution evaluated by *KN*++. Suppose the difference between objective function values of the optimal solution of a model and the best solution by *KN*++ is not significant, meaning confidence intervals of the objective function values coincide. In that case, the percentage deviation found is marked with an asterisk in the table.

Table 4.3: $\%\Delta^*_{Sim}$ for the best solution for models evaluated with respect to decomposition methods in 36 instances

			Model P_S		Model P_M
Form	Nb. of Vehc.	Inc. Hand. Rate	DM-S	DM-S-CF	DM-M-CF
Uniform	4	3	0	0	-0.19 *
		6	0	0	-1.08
		9	-0.54	-0.44	0
		12	0	0	0
	5	3	0	0	0
		6	-0.03 *	-1.15	0
		9	0	0	0
		12	0	0	0
	6	3	-0.51	-0.31	0
		6	0	0	0
		9	0	-0.32 *	-0.23 *
		12	0	0	0
Center-Acc.	4	3	0	0.20 *	0.14 *
		6	0.54	-0.10 *	-0.30 *
		9	0.36	0	-0.61
		12	0.41	0	0
	5	3	0.44	0	-1.04
		6	-0.14 *	-0.09 *	0
		9	0	0.43	0.41
		12	0.08 *	0.10 *	-0.31
	6	3	0	-0.41	0
		6	0	0	0
		9	0.21 *	0.22 *	0.04 *
		12	0	0	0
Outer-Acc.	4	3	0	-0.22 *	0
		6	0	0	0.03 *
		9	0.01 *	0.04 *	0.17 *
		12	1.53	0	0
	5	3	1.14	0	-0.16 *
		6	0	0	0
		9	0	0	0
		12	0.08 *	0.29 *	0
	6	3	0	0	0
		6	0	0	0

Continued on next page

Table 4.3 – continued from previous page

			Model P_S		Model P_M
Form	Nb.of Vehc.	Incident Hand. Rate	DM-S	DM-S-CF	DM-M-CF
		9	0	0	0
		12	0	0	0

In order to summarize $\%\Delta^*_{Sim}$ results in Table 4.3 , four statistics are computed as follows:

- **Number of instances for which the differences are not statistically significant:** the number of instances where the optimal solution is the same with the best solution for $\boldsymbol{P_S}$ reported by *KN*++ or the optimal solutions are not statistically different in terms of the objective function value,
- **Average $\%\Delta^*_{Sim}$:** the average $\%\Delta^*_{Sim}$ over all instances,
- **Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant:** the average $\%\Delta^*_{Sim}$ over instances having optimal solutions that are statistically different,
- **Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant:** the average absolute $\%\Delta^*_{Sim}$ over instances having optimal solutions that are statistically different.

The summary statistics for $\%\Delta^*_{Sim}$ results are given in Table 4.4.

The results in Table 4.4 show that DM-S performs similar to other models in finding solutions near the best solution by *KN*++. Remember that the average MAPD over instances is 7.49% for DM-S from Table 4.2. Although DM-S evaluates the mean response time of a solution with up to 13% deviation in some instances, it performs as well as DM-S-CF in delivering optimal solutions close to the best solution by *KN*++ where the average deviation of instances with statistically significant solutions is 0.42% for DM-S.

In Table 4.3, there are negative $\%\Delta^*_{Sim}$ for some instance-model pairs. These are acceptable given that *KN*++ is an algorithm that is said to guarantees to find the

Table 4.4: Summary of $\%\Delta^*_{Sim}$ in the corresponding objective function values with respect to *KN*++ best solutions in 36 instances

	Model P_S		Model P_M
Statistics	DM-S	DM-S-CF	DM-M-CF
Number of instances for which the differences are not statistically significant	28	31	31
Average $\%\Delta^*_{Sim}$	0.10	-0.05	-0.09
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	0.42	-0.38	-0.52
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	0.68	0.55	0.69

best solution with a predefined probability. Therefore, there could be solutions that perform better than the best solution evaluated by *KN*++. In Table 4.4 , the average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant is -0.38% for DM-S-CF and -0.52% for DM-M-CF. Although the averages are small, these results show that use of DM-S-CF and DM-M-CF methods in evaluation of the objective function value result in optimal solutions that are statistically significantly better than the best solution obtained with KN++.

The average $\%\Delta^*_{Sim}$ over all instances are very low for all three methods. Therefore, assuming server independence (as in DM-S where correction factors are not used) does not significantly affect the ability of decomposition method in finding close enough solutions to the best solution, although it worsens the approximation of the mean response time where the average MAPD over instances is 7.49% for DM-S from Table 4.2.

In addition to the quality of the approximations, computation times in seconds are reported in Table 4.5. As mentioned earlier, the optimal solutions for P_S and P_M are found by enumerating all feasible solutions and evaluating the objective function value with methods DM-S, DM-S-CF and DM-M-CF. In Table 4.5, the computation times spent on finding the optimal solutions under the models and the best solution with *KN*++ algorithm are presented. The experiments are run on the same workstation with Intel Xeon E2246G, a 3.6 GHz processor and 16 GB RAM.

Table 4.5: Computational time in seconds for models in 36 instances

				Model P_S			Model P_M
Form	Nb. of Vehc.	Inc. Hand. Rate	*KN*++	DM-S	DM-S-CF	*KN*++	DM-M-CF
Uniform	4	3	558	2	8	1150	18
		6	460	2	11	966	22
		9	586	2	11	1156	24
		12	443	2	12	1069	26
	5	3	984	6	30	3736	113
		6	854	7	42	3469	140
		9	871	8	45	3331	155
		12	861	9	47	3248	157
	6	3	1581	16	82	13424	604
		6	1326	21	108	12793	707
		9	1309	20	107	11683	708
		12	1307	19	105	11697	709
Center-Acc.	4	3	526	2	8	1093	17
		6	499	2	9	1004	21
		9	468	2	11	1103	24
		12	575	2	12	925	25
	5	3	991	5	29	3590	112
		6	1006	7	37	3455	134
		9	836	8	42	3290	146
		12	908	8	44	3443	144
	6	3	1561	15	79	12733	579
		6	1556	20	103	12799	651
		9	1597	19	100	12267	1484
		12	1439	18	97	11498	646
Outer-Acc.	4	3	658	3	13	1160	23
		6	570	2	10	1027	21
		9	496	2	12	1103	24
		12	448	2	12	1003	25
	5	3	1168	6	31	3803	119
		6	865	7	39	3497	141
		9	1005	8	47	3444	155
		12	960	9	49	3472	160
	6	3	1168	15	84	12918	604
		6	1528	23	102	12580	728
		9	1586	22	113	12136	745

Continued on next page

Table 4.5 – continued from previous page

				Model P_S			Model P_M
Form	Nb. of Vehc.	Inc. Hand. Rate	*KN*++	DM-S	DM-S-CF	*KN*++	DM-M-CF
		12	1448	21	111	12182	726

In Table 4.5, the computation times for proposed methods are significantly lower than *KN*++ for both single and multi-vehicle models despite complete enumeration of feasible solutions. The required computational effort increases with the inclusion of correction factors from DM-S to DM-S-CF. When multiple vehicles are allowed at a single location (as in P_M), the computation time again increases since the number of feasible solutions increases for a problem instance. However, the computation times required for the proposed models are still significantly less than the time required for *KN*++.

Based on these results, DM-S-CF and DM-M-CF methods are very promising since they approximate the objective function value very close to the simulation model without the need for simulation and deliver optimal solution close to the best solution found by *KN*++. The required computation time is also significantly less than *KN*++ where the best performing solution is obtained through the simulation of the alternative solutions.

In the next section, we use the DM-S-CF method in Model P_S on a real-life data set, checking the performance on problem instances with higher number of vehicles and demand regions.

4.4.4 Performance of the methods on Edmonton Data

The data set of City of Edmonton Emergency Medical Services by Ingolfsson et al. (2003) consists of 180 demand nodes with given demand rates and 16 stations with specified capacities ranging from one to three. The demand rates for the nodes vary from node to node. The means of travel time between stations and nodes are given in the data set. The mean incident handling time is set to 45 minutes for this experiment. Total demand calls per hour is set to 5.

We use P_S model with DM-S-CF method to locate from 8 to 12 emergency medical vehicles in Edmonton City, resulting in five problem instances in total. *KN*++ algorithm is used to select the best solutions for these instances with respect to simulation evaluation. According to the results, the best solution (according to *KN*++) in all five instances is found when DM-S-CF method is used to evaluate the objective function value of a solution. Hence, the number of instances for which the differences are not statistically significant is 5 out of 5 for this data set. This shows that DM-S-CF is a good method to find near-best solutions for instances with a higher number of vehicles as well, where the number of individual queueing models constructed to approximate the objective function value is also higher.

The computation times required to solve the problem instances are also reported in Table 4.6 for Edmonton data.

Table 4.6: Computational time in seconds

Nb. of Vehc.	*KN*++	P_S with DM-S-CF
8	15234	4270
9	14736	4864
10	11505	4175
11	6977	2642
12	3025	1240

It is seen that computation time for complete enumeration of P_S with DM-S-CF is less than the time required for *KN*++ algorithm to evaluate the best solution. As the number of feasible solutions under an instance increases (from 12 vehicles to 8 vehicles), the decrease in computation time in comparison to *KN*++ increases. This is due to the increasing computational burden of simulation of all feasible solutions in the initial stage of the *KN*++ and the increasing number of solutions that needs to be eliminated in the iterations when the number of feasible solution increase.

4.4.5 Performance of DM-S-CF under $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$

In addition to the models P_S and P_M, we use the decomposition models to evaluate different objective functions. For this purpose, models $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ are used from Chapter 3. From the experiments in the Chapter 3, it is seen that $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ improves equity in resulting optimal solutions in comparison to minimizing mean response time. Therefore, decomposition methods are used to find optimal solutions for these models with different objective functions from P_S and P_M.

Recall that $\boldsymbol{P_4}$ minimizes the maximum mean region-wise response time while $\boldsymbol{P_7}$ minimizes the total positive deviation of mean region-wise response time from a threshold travel time τ.

$$(\boldsymbol{P_4}) \qquad \text{Minimize} \qquad \max_{j \in J}(R_j(\vec{x})) \tag{4.35}$$

$$\sum_{i \in I} x_i = N \tag{4.36}$$

$$x_i \in \{0,1\}, \qquad \forall i \in I. \tag{4.37}$$

$$(\boldsymbol{P_7}) \qquad \text{Minimize} \qquad \sum_{j \in J} [R_j(\vec{x}) - \tau]_+ \tag{4.38}$$

$$\sum_{i \in I} x_i = N \tag{4.39}$$

$$x_i \in \{0,1\}, \qquad \forall i \in I, \tag{4.40}$$

where $[\,.\,]_+ = \max\{0, .\}$.

We use DM-S-CF method to evaluate the objective function value of feasible solutions of $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$, and find the optimal solutions for each model under each problem instance by complete enumeration. By using *KN*++, the best solutions according to the simulation study are found as well. After finding the optimal solution by using DM-S-CF and the best solution by using *KN*++, the performance measures for these solutions are evaluated from a separate simulation run.

Firstly, $\%\Delta^*_{Sim}$ are found for each problem instances by comparing the objective function values of solution found by DM-S-CF and *KN*++. In order to summarize

$\%\Delta^*_{Sim}$ results, four statistics are computed as in Chapter 4.4.3 and given in Table 4.7.

Table 4.7: Summary of $\%\Delta^*_{Sim}$ in the corresponding objective function values with respect to *KN*++ best solutions in 36 instances

Statistics	$\boldsymbol{P_4}$	$\boldsymbol{P_7}$
Number of instances for which the differences are not statistically significant	35	33
Average $\%\Delta^*_{Sim}$	-0.03	0.53
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	-0.79	5.04
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	0.79	7.32

It is seen that DM-S-CF evaluates the same solution with *KN*++ best or a solution that the difference is not statistically significant in thirty-five instances for $\boldsymbol{P_4}$ and thirty-three instances for $\boldsymbol{P_7}$. The average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant are -0.79 % and 5.04 % for $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$, respectively. The average $\%\Delta^*_{Sim}$ is relatively high for $\boldsymbol{P_7}$ where this can be attributed to the convolution of estimation errors due to the summation of region-wise measures in the objective function. However, DM-S-CF still evaluates the same solution or a solution that the difference is not statistically significant in thirty-three instances. Then, the performance of DM-S-CF for these two measures is similar to its performance in mean response time measure assessed using models P_S and P_M.

Secondly, the effect of evaluating the objective function with DM-S-CF on equity measures of the resulting solution of $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ is also checked. The optimal solutions of $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ found by using DM-S-CF and the best solutions found by using *KN*++ are compared to the best solution for P_S in terms the mean response time, the variance of region-wise mean response time and the Gini coefficient as in Chapter 3.

In order to quantify the difference of the models, the mean absolute percent deviation of measures for the optimal solution of the models, $MAPD$, from P_S is calculated for each measure over all instances. We report the average percent positive deviations, $avgpd(\%)$, of the mean response time of the best solution in the models from $\boldsymbol{P_S}$. The percent of instances with positive deviation, $ppd(\%)$, in mean response time is also reported in order to show the fraction of instances with positive deviations. One can refer to (3.38),(3.39) and (3.40) for the details of the statistics, $MAPD$, $avgpd(\%)$ and $ppd(\%)$.

For $VarR_j$ and G, the average percent negative deviation, $avgnd(\%)$, and the percent of instances with negative deviation, $pnd(\%)$, are reported instead since equity gets better as $VarR_j$ and G decrease.

In Table 4.8, these statistics are reported to show the changes in R, $VarR_j$ and G.

Table 4.8: Comparison of Models P_4 and P_7 with P_S in performance measures

		P_4		P_7	
Measure	Statistics	DM-S-CF	*KN*++	DM-S-CF	*KN*++
R	$MAPD(\%)$	14.10	14.74	11.23	11.51
	$ppd(\%)$	100	100	100	100
	$avgpd(\%)$	14.10	14.74	11.23	11.51
$VarRj$	$MAPD(\%)$	49.15	49.97	48.31	51.24
	$pnd(\%)$	100	100	91.67	97.22
	$avgnd(\%)$	49.15	49.97	51.59	51.80
G	$MAPD(\%)$	37.76	38.59	34.47	36.67
	$pnd(\%)$	100	100	97.22	97.22
	$avgnd(\%)$	37.76	38.59	35.29	37.48

The results show that the effect of $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ on the mean response time, the variance of mean region-wise response time and the Gini coefficient does not significantly change when DM-S-CF is used to evaluate the objective function value of the feasible solutions. This result is said to be expected since DM-S-CF performs very well in finding the best solutions reported by *KN*++ for $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ as mentioned earlier.

4.4.6 Performance of the proposed genetic algorithm

In this section, the performance of the genetic algorithm proposed is studied in terms of finding the optimal solution for the mathematical models and the deviations of the best solutions from the optimal.

A design of experiments is constructed to see the effect of different algorithm parameters on the performance of the genetic algorithm. These parameters are determined as population size (M), probability of crossover (u_c) and probability of mutation (u_m). For each parameter, two levels are defined as in Table 4.9.

Table 4.9: GA parameter levels for the design of experiments

Parameter	M	u_c	u_m
Levels	50	0.80	0.05
	100	0.90	0.10

Thirty-six problem instances of P_M are solved using GA where the number of feasible solutions is higher than P_S. DM-M-CF method are used to evaluate the objective function value of a solution. Each problem instance is solved by starting GA ten times. The best solutions, $\vec{x}^*_{KN++}$, under *KN*++ and optimal solutions, $\vec{x}^*_{P_M}$ for Model P_M with DM-M-CF method for these instances are known from the previous analysis.

Two measures are reported to show the performance of the GA under different settings. The first one is the percent of the replications of GA with different initial populations, AOF(%), that GA found the optimal solution $\vec{x}^*_{P_M}$ of an instance out of ten independent replications. The second measure is the percent of time, OF(%), that GA finds the optimal solution for a problem instance in at least one of the ten replications. For these measures, solutions are not compared in terms of the objective function values but in vehicle locations only. The averages of these measures overall problem instances are given in Table 4.10 with respect GA parameters.

In addition, the solutions reported by GA is compared to the best solution obtained with *KN*++ and previous measures are given for this comparison in Table 4.10 as well.

According to the results, GA always finds the optimal solution, $\bar{x}^*_{P_M}$, among at least one of the ten independent replications when population size is set to 100. This population size increases the likelihood of finding the best solution, $\vec{x}^*_{KN++}$, in a single run (AOF for *KN*++) as well. More detailed results are given under

Table 4.10: Average performance of the GA according to P_M and *KN*++ solutions

			P_M Optimal		*KN*++ Best	
S	u_c	u_m	AOF(%)	OF(%)	AOF(%)	OF(%)
50	0.8	0.05	45	94	28	78
		0.1	57	100	34	83
	0.9	0.05	43	94	25	72
		0.1	56	97	34	78
100	0.8	0.05	70	100	38	67
		0.1	79	100	43	69
	0.9	0.05	71	100	40	75
		0.1	81	100	45	67

different network specifications (form, number of vehicles and incident handling rate) in Appendix B.

In addition to the performance of GA in finding optimal or near-best solutions, we also report the percent deviation $\%\Delta^*_{Sim}$ and mean absolute percent deviation of the objective function value of GA solutions from $\vec{x}^*_{P_M}$ and $\vec{x}^*_{KN++}$. Every solution found by GA is evaluated under the simulation model with ten independent replications, and a 90% confidence interval on the objective function value is constructed to compare objective function values.

Let $\vec{x}^*_{(.)}$ be the solution that gives the minimum mean response time under model (.) and $O_{Sim}(\vec{x}^*_{(.)})$ denotes the mean of 90% confidence interval of the objective function value for the solution based on 10 independent replications of the simulation model.

$$\%\Delta^*_{Sim} = \frac{100}{K}\sum_{k=1}^{K} I_{\left\{\vec{x}^*_{(.)}\neq\vec{x}^*_{GA,k}\right\}}\left(\frac{O^*_{Sim}(\vec{x}^*_{GA,k}) - O_{Sim}(\vec{x}^*_{(.)})}{O_{Sim}(\vec{x}^*_{(.)})}\right)$$

where I is the indicator function, K is the total number of GA replications for an instance and equal to ten, and k indexes the corresponding statistics for the k^{th} run of the GA. (.) is replaced by P_M and *KN*++ for Tables 4.11 and 4.12, respectively.

$$MAPD = \frac{100}{K}\sum_{k=1}^{K} I_{\left\{\vec{x}^*_{(.)}\neq\vec{x}^*_{GA,k}\right\}}\left|\frac{O^*_{Sim}(\vec{x}^*_{GA,k}) - O_{Sim}(\vec{x}^*_{(.)})}{O_{Sim}(\vec{x}^*_{(.)})}\right|$$

where K is the total number of GA replications for an instance and equal to 10 and

$(.)$ is replaced by P_M and *KN*++ for Tables 4.11 and 4.12, respectively.

In Tables 4.11 and 4.12, the averages of those measures over all solutions and the averages over only the solutions statistically different from the optimal or best solution are reported under the columns *Overall* and *Statistically Significant*, respectively.

Table 4.11: Average $\%\Delta^*_{Sim}$ and MAPD of GA results from P_M Optimal Solution

			Overall		Statistically Significant	
S	u_c	u_m	$Avg\ \%\Delta^*_{Sim}$ (%)	MAPD(%)	$Avg\ \%\Delta^*_{Sim}$ (%)	MAPD(%)
50	0.8	0.05	0.77	0.79	1.64	1.68
		0.1	0.39	0.42	1.21	1.25
	0.9	0.05	0.80	0.81	1.48	1.50
		0.1	0.48	0.51	1.27	1.31
100	0.8	0.05	0.23	0.26	1.16	1.21
		0.1	0.15	0.18	0.92	0.98
	0.9	0.05	0.22	0.25	1.10	1.15
		0.1	0.13	0.17	0.95	1.03

From the results in Table 4.11, it is seen that GA has the smallest $Avg\ \%\Delta^*_{Sim}$ and MAPD(%) from the P_M optimal solution under the setting $S = 100$, $u_c = 0.90$, $u_m = 0.10$ over all solutions.

Table 4.12: Average Deviation and MAPD of GA results from KN++ Best Solution

			Overall		Statistically Significant	
S	u_c	u_m	$Avg\ \%\Delta^*_{Sim}$ (%)	MAPD(%)	$Avg\ \%\Delta^*_{Sim}$ (%)	MAPD(%)
50	0.8	0.05	0.67	0.80	1.45	1.62
		0.1	0.29	0.48	0.95	1.22
	0.9	0.05	0.69	0.83	1.26	1.45
		0.1	0.37	0.55	1.03	1.26
100	0.8	0.05	0.14	0.35	0.80	1.09
		0.1	0.06	0.29	0.63	1.00
	0.9	0.05	0.14	0.34	0.74	1.05
		0.1	0.04	0.28	0.57	0.97

In Table 4.12, the smallest values for the measures are realized under the same setting

($S = 100$, $u_c = 0.90$, $u_m = 0.10$) for both of the cases: over all solutions and statistically different solutions. Hence, parameters could be set as $S = 100$, $u_c = 0.90$, $u_m = 0.10$ for the GA.

In addition to quality of the solution reported by GA, the computation time required for GA to report a solution under the setting ($S = 100$, $u_c = 0.90$, $u_m = 0.10$), next to solution times for *KN*++ and complete enumeration of feasible solutions of P_M with DM-M-CF method are presented in Table 4.13

Table 4.13: Computational time in seconds for KN++, complete enumeration and GA in 36 instances

				P_M with DM-M-CF	
Form	Nb. of Vehc.	Inc.Hand. Rate	*KN*++	Compl. Enum.	GA
Uniform	4	3	1150	18	5
		6	966	22	5
		9	1156	24	7
		12	1069	26	6
	5	3	3736	113	11
		6	3469	140	14
		9	3331	155	18
		12	3248	157	17
	6	3	13424	604	22
		6	12793	707	28
		9	11683	708	27
		12	11697	709	26
Center-Acc.	4	3	1093	17	6
		6	1004	21	6
		9	1103	24	6
		12	925	25	7
	5	3	3590	112	11
		6	3455	134	14
		9	3290	146	14
		12	3443	144	14
	6	3	12733	579	22
		6	12799	651	26
		9	12267	1484	28
		12	11498	646	24

Continued on next page

Table 4.13 – continued from previous page

				P_M with DM-M-CF	
Form	Nb. of Vehc.	Inc. Hand. Rate	*KN*++	Compl. Enum.	GA
Outer-Acc.	4	3	1160	23	4
		6	1027	21	5
		9	1103	24	5
		12	1003	25	6
	5	3	3803	119	9
		6	3497	141	11
		9	3444	155	12
		12	3472	160	13
	6	3	12918	604	18
		6	12580	728	19
		9	12136	745	21
		12	12182	726	22

Together with the previous results, it is seen that the proposed genetic algorithm finds a solution under an average of thirty seconds in all instances with very small deviations from the objective function value of the optimal solution of P_M and of the best solution obtained with *KN*++.

4.5 Conclusion

In this chapter, the exact queueing model is decomposed into interdependent queueing models to assess the EMS system's performance measures due to the exponentially increasing size of the exact queueing model with the number of vehicles and demand regions. The decomposition method proposed results in a set of interdependent balance equations, forming a non-linear set of simultaneous equations. An approximation method is proposed to solve the resulting set of equations and estimate the steady-state probabilities.

The proposed decomposition methods work on cases where multiple vehicles are allowed at a single location and service rates are specific to vehicle location - demand region pairs. Differently from the studies of Budge et al. (2009), the vehicles at a single location are differentiated by prioritizing the dispatches among them.

Furthermore, decomposition methods are analyzed under an optimization setting to reveal the ability in differentiating alternative solutions under a mathematical model in line with the simulation model. A meta-heuristic algorithm to solve the mathematical models since the mathematical models have no closed-form formulation.

An extensive experimental study is conducted on both toy and real-life data to assess the performance of the methods. The experiments have shown that DM-S-CF and DM-M-CF methods are very good at approximating the mean response time of the system by about %2.5 deviation at the most and about 0.9% deviation on average in the problem instances. It is also shown that near-best solutions are found by about 0.5% deviation from the best solution obtained from a simulation study. In addition, DM-S-CF are tested on two models $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ from Chapter 3 with two different objective functions: maximum mean region-wise response time and the total positive deviation of mean region-wise response time from a threshold. It is seen that DM-S-CF are good at finding optimal solutions for those models where it finds the optimal solution in thirty-five and thirty-three instances out of thirty-six for $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$, respectively.

The genetic algorithm proposed is also analyzed in terms of the solutions found. It is seen that GA always finds the optimal solution of the mathematical model and finds the best solution obtained by simulation in 67% of instances in one of the ten independent runs for the toy data. Although GA could not find the best solution with respect to simulation study in some instances, the average deviation of GA solutions from the best solution of the simulation study is very low at around 0.57%.

In this chapter, an algorithmic approach is proposed for the exact queueing model to assess the performance measures of the EMS system. Working with queueing models in the proposed decomposition methods allows the decision-maker to evaluate various measures, including mean region-wise response time, coverage or lost demand. However, the method requires algorithmic approaches in assessing the measures due to set of nonlinear equations and in finding the optimal solutions for a mathematical model such as meta-heuristics.

In the next chapter, we work on closed form formulations to approximate the exact queueing model which can be solved with package solvers, enabling one to construct optimization problems with various performance measures.

CHAPTER 5

MATHEMATICAL MODELS BASED ON CLOSED-FORM APPROXIMATIONS OF PERFORMANCE MEASURES FOR STOCHASTIC EMS VEHICLE LOCATION PROBLEM

In this chapter, mathematical models which can be solved with commercial solvers are proposed for the EMS vehicle location problem. Closed-form formulations to approximate the performance measures of the EMS system are developed based on queueing models similar to the decomposition methods in Chapter 4. Hence, objective functions and constraints are expressed as a function of decision variables which enables one to construct several mathematical models for the EMS vehicle location problem.

In the literature, there are various probabilistic models that can be solved with commercial such as MEXLCP by Daskin (1983), PLSCP by ReVelle and Hogan (1988) and MALP by ReVelle and Hogan (1989). Later, several studies extend those models which includes local reliability contraints for service or multi-objective problems, previously mentioned in Chapter 2. However, aforementioned models and extensions require estimation of the fundamental parameter, busy probability of vehicles, *a priori*. In this chapter, we propose mathematical models where the busy probabilities of vehicles are expressed in the form of decision variables based on queueing models. Therefore, we embed the estimation of the busy probabilities in the models which makes the application easier for the decision makers. Another contribution is that the busy probabilities are estimated with respect to feasible solutions instead of using pre-computed constant busy probabilities which are not affected by the location decisions.

In Chapter 4, a decomposition method with various variants is proposed to evaluate the performance measures of an EMS system. The decomposition methods relies on the queueing models to evaluate the measures. EMS system is decomposed into interdependent queueing models, and performance measures of the EMS system are estimated based on the steady-state distributions of those interdependent models. The decomposition methods proposed requires algorithmic solution approaches in approximating the steady-state distributions. Therefore, the decomposition methods are more computationally burdensome than closed-form mathematical models, where commercial solvers are easily exploited to solve the models.

In this chapter, the interdependence among vehicles in serving the demand calls is ignored. Separate queueing models are constructed for each vehicle and treated as independent systems. Since those models are independent, the balance equations for a model do not incorporate nonlinear terms. Then, the busy probability of each vehicle is easily found based on the steady-state distribution of each model. The busy probabilities of models are used to estimate several performance measures of the EMS systems. The estimation of busy probabilities and performance measures are expresses in the form of decision variables and embedded in mathematical models.

Several mathematical models are proposed based on the same constraints by changing objective functions such as maximizing expected satisfied demand, maximizing expected covered demand or minimizing mean response time. Two different way of estimating busy probabilities are proposed. The performance of models are tested under optimality. The models' ability in finding near-best solution in comparison to *KN++* algorithm are evaluated. The quality of the optimal solution of the models are evaluated with respect to solutions obtained by complete enumeration. In addition, a commercial solver is used to solve the mathematical models and the optimality gap is reported for the models. The experiments are done on both toy data and the real life data set from Chapter 4.

The rest of the chapter is as follows. In Section 5.1, the problem environment, construction of the separate queueing models and estimation of the performance measures are explained. In Section 5.2, mathematical models incorporating closed-form formulations are presented. Experimental study is given in Section 5.3 and this chapter is concluded in Section 5.4.

5.1 Modeling Approach and Estimation of the Performance Measures

In the decomposition methods discussed in Chapter 4, the exact queueing model is decomposed into N interdependent queueing models, each representing a single vehicle. Furthermore, the service time for a demand call from region j by a vehicle from location i is assumed to be exponentially distributed with a mean that is equal to the sum of mean travel time to demand region, mean incident handling time, and mean travel time back to the vehicle location ($\omega_{ij}+\phi_j+\omega_{ij}$). In the decomposition methods, queueing models are constructed interdependently since the closest available vehicle is assigned to a demand call. This requires the demand rates in the queueing models to change with respect to the busy probability of other vehicles so that the assignment of demand calls in the exact queueing model are represented in the interdependent queueing models. However, this structure results in a set of nonlinear equations which need to be solved simultaneously to find steady state probabilities.

In this chapter, the interdependence among busy probabilities of vehicles is ignored. Hence, separate independent queueing models constructed for each vehicle. Differently from the previous chapter, the closed form formulations are only developed for single-vehicle problems where at most one vehicle is allowed at a single location.

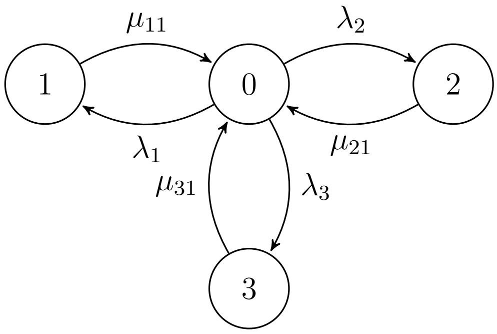
Recall the queueing model for decomposition method DM-S in Chapter 4 where $\{b_t, t \geq T\}$ is a continuous time Markov chain with state space $L_i = J \cup \{0\}$ for $QM_i,\ \forall i \in I'$, $I' = \{i \in I | x_i > 0\}$ under given solution $\vec{x}$. A state $b_t \in L_i,\ t \geq T$ is 0 when the vehicle is free and j when it is busy serving region j at time t. The transition from state $j = 0$ to state $j \neq 0$ is realized when a demand call arrives and its rate for QM_i is $\lambda_j c^i_j$ where c^i_j is the probability that all vehicles closer to region j than vehicle i are busy if there is any, and 1 otherwise.

Since interdependence between vehicles is ignored in this chapter, the transition rate from state $j = 0$ to state $j \neq 0$ for QM_i becomes λ_j. The service rate μ_{ji} for state $j \neq 0$ for QM_i is kept the same as in (4.8).

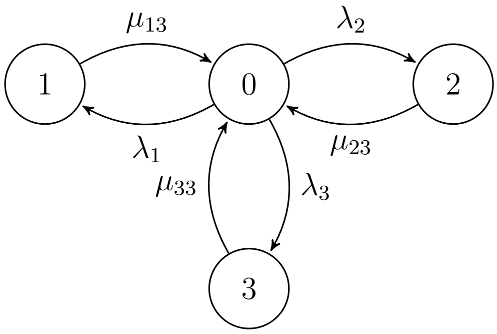
Recall the example from Chapter 4 .

Example: Assume $\vec{x} = (1,0,1)$ where $I = \{1,2,3\}$, $J = \{1,2,3\}$ and $N = 2$. Assume that Region 2 is closer to Region 1 than Region 3. We construct a queueing model for the vehicle in Region 1 as QM_1 and another for the one in Region 3 as QM_3. Then, L_1 and L_3 are the same and equal to $\{0,1,2,3\}$ for both models.

Then, the queueing models could be constructed as in Figure 5.1. Now, none of the transition rates from state $j = 0$ includes the term c^i_j differently from decomposition method DM-S since queueing models are mutually independent now.



(a) Rate diagram for QM_1



(b) Rate diagram for QM_3

Figure 5.1: Rate diagrams for the queueing models of the given solution $\vec{x}$

Recall that π^i_j is the steady state probability for state $j \in B_i$ of QM_i. Then, the probability vehicle i being free, π^i_0 , is written as follows :

$$\pi^i_0 = \frac{1}{\sum\limits_{j \in L_i \setminus \{0\}} \frac{\lambda_j}{\mu_{ji}} + 1}. \tag{5.1}$$

Notice that the computation of π^i_0 in (5.1) is different from (4.10) where the former lacks c^i_j as a multiplier of λ_j.

Based on the steady-state probabilities, the busy probability of vehicle i under $\vec{x}$ is equal to $(1-\pi^i_0)$. Now, these busy probabilities of vehicles derived from the queueing models are used to construct closed-form formulations for performance measures of the EMS system.

Remark. *In the exact system, a vehicle responds to a demand call only if all closer vehicles are busy. Therefore, the busy probability found by $(1-\pi^i_0)$ is an upper bound for the busy probability of vehicle i in the exact queueing system since all vehicles are* *assumed to be mutually independent and serve the demand calls from all regions without considering closer vehicles being available.*

Following the assumption of independence, several performance measures could be approximated by using busy probabilities of vehicles, such as expected satisfied demand or mean response time of the system.

For the sake of brevity, let $p_i = (1 - \pi_0^i)$ be the busy probability of vehicle i under $\vec{x}$ based on the steady state probabilities of the states of the separate queueing models. Let v_{jk} be the vehicle location that is the k^{th} closest to region j, v_{j1} being the closest and v_{jN} being the farthest location to region j with a vehicle located.

The probability that a demand call from region j is satisfied by the closest vehicle is

$$P\Big(\text{a dem. call from reg. } j \text{ is satisfied by the closest vehc.}\Big) = (1 - p_{v_{j1}}). \quad (5.2)$$

Based on the assumption of independence among vehicles, the probability that a demand call from region j is satisfied by at most the second closest vehicle is

$$\begin{aligned} &P\big(\text{ a dem. call from reg. } j \text{ is satf. by the cls. or the second cls. vehc.}\big) \\ &\qquad = (1 - p_{v_{j1}}) + (1 - p_{v_{j2}})p_{v_{j1}}. \end{aligned} \quad (5.3)$$

Following (5.2) and (5.3), the probability that a demand call from region j is satisfied is calculated as follows:

$$\begin{aligned} &P\big(\text{a demand call from region } j \text{ is satisfied}\big) \\ &\qquad = (1 - p_{v_{j1}}) + \sum_{t=2}^{N}(1 - p_{v_{jt}}) \prod_{k=1}^{t-1} p_{v_{jk}}. \end{aligned} \quad (5.4)$$

By using the probability given in (5.4), the expected amount of demand satisfied in unit time under solution $\vec{x}$, $ED(\vec{x})$ is estimated as follows:

$$ED(\vec{x}) = \sum_{j \in J} \lambda_j \left((1 - p_{v_{j1}}) + \sum_{t=2}^{N}(1 - p_{v_{jt}}) \prod_{k=1}^{t-1} p_{v_{jk}} \right). \quad (5.5)$$

Similar to $ED(\vec{x})$, the expected covered demand $CD(\vec{x})$ under $\vec{x}$ could be estimated. The probability that a demand call form region j is covered under threshold τ is

$$\begin{aligned} &P\big(\text{a demand call from region } j \text{ is covered under threshold } \tau\big) \\ &\qquad = (1-p_{v_{j1}})F_{T_{jv_{j1}}}(\tau) + \sum_{t=2}^{N}(1-p_{v_{jt}})F_{T_{jv_{jt}}}(\tau)\prod_{k=1}^{t-1}p_{v_{jk}} \end{aligned} \tag{5.6}$$

where $F_{T_{ij}}$ is the cumulative distribution function of the travel time between region j and vehicle i.

Then, the expected covered demand in unit time under $\vec{x}$ is found as

$$CD(\vec{x}) = \sum_{j\in J}\lambda_j\left((1-p_{v_{j1}})F_{T_{jv_{j1}}}(\tau) + \sum_{t=2}^{N}(1-p_{v_{jt}})F_{T_{jv_{jt}}}(\tau)\prod_{k=1}^{t-1}p_{v_{jk}}\right). \tag{5.7}$$

Another measure that can be calculated based on busy probabilities is region-wise mean response time, $R_j(\vec{x})$, under solution $(\vec{x})$.

$$R_j(\vec{x}) = \frac{(1-p_{v_{j1}})\omega_{jv_{j1}} + \sum_{t=2}^{N}(1-p_{v_{jt}})\omega_{jv_{jt}}\prod_{k=1}^{t-1}p_{v_{jk}}}{1-\prod_{k=1}^{N}p_{v_{jk}}} \tag{5.8}$$

where $\omega_{jv_{jt}}$ is the mean travel time between region j and vehicle location v_{jt}. The numerator in (5.8) is the weighted total travel time based on the probabilities that the vehicle in the corresponding location (v_{jt}) is assigned to the demand call form region j while the denominator is the probability that there is at least one vehicle that could be assigned to this demand call.

In addition, the mean response time $R(\vec{x})$ could be calculated based on the region-wise mean response time $(R_j(\vec{x}))$ as follows:

$$R(\vec{x}) = \frac{\sum_{j\in J}\lambda_j\left((1-p_{v_{j1}})\omega_{jv_{j1}} + \sum_{t=2}^{N}(1-p_{v_{jt}})\omega_{jv_{jt}}\prod_{k=1}^{t-1}p_{v_{jk}}\right)}{\sum_{j\in J}\lambda_j\left((1-p_{v_{j1}}) + \sum_{t=2}^{N}(1-p_{v_{jt}})\prod_{k=1}^{t-1}p_{v_{jk}}\right)} \tag{5.9}$$

where the numerator is the demand weighted mean region-wise response time and the denominator is the expected demand satisfied in unit time.

Following the formulations for the performance measures, several mathematical models are constructed by defining the busy probabilities of vehicles as decision variables associated with the location solution in the next section.

5.2 Mathematical Models with Closed-form Formulations

In this section, we construct mathematical models for EMS vehicle location problem by expressing the performance measures developed in the previous section based on decision variables. The notation used for the mathematical models is given in Table 5.1.

Table 5.1: Notation used for the mathematical models

Sets	
I	the set of vehicle locations
J	the set of demand regions
Parameters	
N	the number of vehicles to be located
ω_{ij}	mean travel time between vehicle location i and demand region j
λ_j	demand rate of region j per unit time
μ_{ji}	service rate for region j when served by a vehicle from location i

The mathematical models are constructed with the following decision variables in Table 5.2.

The relation between decision variables are shown with an example. Assume that $x_i = 1$ for given i. $a_{ij1} = 1$ if it is the farthest vehicle to region j, and then $y_{j1} = \omega_{ij}$. Similarly, $a_{ijN} = 1$ if it is the closest vehicle to region j, and $y_{jN} = \omega_{ij}$.

Recall that the probability that a demand call from region j is satisfied by the closest vehicle is defined as $(1 - p_{v_{j1}})$ where $p_{v_{j1}}$ is the busy probability of the vehicle which is the closest to region j and located at v_{j1}.

In terms of the decision variables p_i and a_{ijk}, the probability that a demand call from

Table 5.2: Decision variables used in the mathematical models

x_i	a binary variable being equal to 1 if a vehicle is located at vehicle location i and 0, otherwise
p_i	the busy probability of vehicle located in region i
a_{ijk}	a binary variable which takes the value of 1 if vehicle location i is the k^{th} farthest server location (where a vehicle is located) to region j and 0, otherwise
y_{jk}	mean travel time from the k^{th} farthest vehicle to region j

region j is satisfied by the closest vehicle can be expressed as

$$P\begin{pmatrix}\text{a demand call from region } j \text{ is satisfied}\\ \text{by the closest vehicle}\end{pmatrix} = 1 - \sum_{i\in I} a_{ijN}p_i \tag{5.10}$$

where a_{ijN} takes the value of 1 only for $i = v_{j1}$ for a given j by definition. Hence, $\sum_{i\in I} a_{ijk}p_i$ computes the busy probability of the k^{th} farthest vehicle to region j.

Based on the notation and decision variables, several mathematical models are proposed with different objective functions as maximizing expected satisfied demand, maximizing expected covered demand and minimizing mean response time.

5.2.1 Maximizing Expected Satisfied Demand (MESD) Model

The first model constructed is based on the measure ED, which is the expected amount of demand satisfied in unit time.

MESD which maximizes expected satisfied demand in unit time is as follows:

$$\begin{aligned}\text{Max} \quad & \sum_{j\in J}\lambda_j\Bigg((1-\sum_{i\in I}a_{ijN}p_i) \\ & \qquad + \sum_{t=1}^{N-1}\Big((1-\sum_{i\in I}a_{ijt}p_i)\prod_{k=t+1}^{N}\sum_{i\in I}a_{ijk}p_i\Big)\Bigg)\end{aligned} \tag{5.11}$$

$$\text{s. to:} \quad p_i = \left(1 - \frac{1}{\sum\limits_{j\in J}\frac{\lambda_j}{\mu_{ji}} + 1}\right)x_i, \qquad \forall i \in I, \tag{5.12}$$

$$\sum_{i \in I} x_i = N, \tag{5.13}$$

$$y_{jk} = \sum_{i \in I} a_{ijk} \omega_{ij}, \qquad \forall j \in J, k = 1, \ldots, N, \tag{5.14}$$

$$y_{j(k+1)} \leq y_{jk}, \qquad \forall j \in J, k = 1, \ldots, N-1, \tag{5.15}$$

$$\sum_{k=1}^{N} a_{ijk} \leq x_i, \qquad \forall i \in I, \forall j \in J, \tag{5.16}$$

$$\sum_{i \in I} a_{ijk} = 1, \qquad \forall j \in J, k = 1, \ldots, N, \tag{5.17}$$

$$x_i \in \{0, 1\}, \qquad \forall i \in I, \tag{5.18}$$

$$p_i \geq 0, \qquad \forall i \in I, \tag{5.19}$$

$$a_{ijk} \in \{0, 1\}, \qquad \forall i \in I, \forall j \in J, k = 1, \ldots, N, \tag{5.20}$$

$$y_{jk} \geq 0, \qquad \forall j \in J, k = 1, \ldots, N. \tag{5.21}$$

The objective function in (5.11) maximizes expected satisfied demand in unit time. Recall (5.5),

$$ED(\vec{x}) = \sum_{j \in J} \lambda_j \left((1 - p_{v_{j1}}) + \sum_{t=2}^{N} (1 - p_{v_{jt}}) \prod_{k=1}^{t-1} p_{v_{jk}} \right),$$

where $p_{v_{j1}}$ is the busy probability of the vehicle which is the closest to region j and located at v_{j1}. The term $(1 - \sum_{i \in I} a_{ijN} p_i)$ in (5.11) corresponds to the term $(1 - p_{v_{j1}})$ in (5.5) and $\sum_{i \in I} a_{ijk} p_i$ to the busy probability of the k^{th} farthest vehicle to region j. Hence, the objective function evaluates the expected satisfied demand in unit time under solution $\vec{x}$.

(5.12) in **MESD** is constructed based on the steady state probability of vehicle at location i being busy, $(1 - \pi_0^i)$ from QM_i. The multiplier of x_i in right-hand side of (5.12) is equal to $(1 - \pi_0^i)$ where π_0^i is substituted with (5.1). Hence, (5.12) enforces p_i to be equal to the busy probability of vehicle at location i if a vehicle is located and to 0 otherwise. (5.14) and (5.15) sorts the vehicle locations having a vehicle located with respect to their mean travel time (ω_{ij}) to region j. (5.16) ensures that only location with a vehicle is included in the sorting and (5.17) allows only one location is sorted as the k^{th} farthest location to region j.

Notice that the objective function in (5.11) includes nonlinear terms that can be linearized. The term $a_{ijk}p_i$ is the multiplication of a binary and a continuous decision variable. Let $s_{ijk} = a_{ijk}p_i$. Then, the following constraints is added to the model to replace $a_{ijk}p_i$ with continuous decision variable s_{ijk}.

$$0 \leq s_{ijk} \leq a_{ijk}, \qquad \forall i \in I, \forall j \in J, k = 1, \ldots, N, \tag{5.22}$$

$$p_i - (1 - a_{ijk}) \leq s_{ijk} \leq \ p_i, \qquad \forall i \in I, \forall j \in J, k = 1, \ldots, N. \tag{5.23}$$

Following, **MESD** is rewritten in the form of a mixed integer nonlinear programming model without any nonlinear terms in the constraints as follows:

$$\text{Max} \quad \sum_{j \in J} \lambda_j \left((1 - \sum_{i \in I} s_{ijN}) + \sum_{t=1}^{N-1} \left((1 - \sum_{i \in I} s_{ijt}) \prod_{k=t+1}^{N} \sum_{i \in I} s_{ijk} \right) \right) \tag{5.24}$$

$$\text{s. to:} \quad (5.12) - (5.23).$$

Although the term $a_{ijk}p_i$ is linearized, the objective function in (5.24) still includes nonlinear terms due to the multiplication of decision variables s_{ijk}. Hence, the model is a mixed integer nonlinear program.

5.2.2 Maximizing Expected Covered Demand (MECD) Model

Another model is constructed to maximize expected covered demand, $CD(\vec{x})$. Recall that $CD(\vec{x})$ in (5.7) is estimated based on the busy probabilities as follows:

$$CD(\vec{x}) = \sum_{j \in J} \lambda_j \left((1 - p_{v_{j1}}) F_{T_{jv_{j1}}}(\tau) + \sum_{t=2}^{N} (1 - p_{v_{jt}}) F_{T_{jv_{jt}}}(\tau) \prod_{k=1}^{t-1} p_{v_{jk}} \right)$$

where $p_{v_{jk}}$ is the busy probability of k^{th} closest vehicle location, v_{jk}, to region j.

Similar to the representation of $p_{v_{jk}}$ based on decision variables a_{ijk} and p_i, $F_{T_{jv_{j1}}}(\tau)$ is replaced with $\sum_{i \in I} a_{ijN} F_{T_{ij}}(\tau)$ in the objective function. a_{ijN} takes the value of one for the closest vehicle location i to the region j (N^{th} farthest) where a vehicle is located, making $F_{T_{jv_{j1}}}(\tau) = \sum_{i \in I} a_{ijN} F_{T_{ij}}(\tau)$.

The following model is constructed similar to **MESD**, having expected covered demand in the objective function.

Model **MECD** is as follows:

$$\begin{aligned} \text{Max} \quad & \sum_{j\in J}\lambda_j\Bigg((1-\sum_{i\in I}s_{ijN})\sum_{i\in I}a_{ijN}F_{T_{ij}}(\tau) \\ & \qquad +\sum_{t=1}^{N-1}\Big((1-\sum_{i\in I}s_{ijt})\sum_{i\in I}a_{ijt}F_{T_{ij}}(\tau)\prod_{k=t+1}^{N}\sum_{i\in I}s_{ijk}\Big)\Bigg) \end{aligned} \tag{5.25}$$

$$\text{s. to:} \quad (5.12)-(5.23).$$

Similar to **MESD**, **MECD** is a mixed integer nonlinear program due to the nonlinear terms in the objective function.

5.2.3 Minimizing Mean Response Time (MMRT) Model

In this section, a new model that minimizes mean response time is constructed. Recall the approximation of the mean response time based on busy probabilities are

$$R(\vec{x}) = \frac{\sum\limits_{j\in J}\lambda_j\left((1-p_{v_{j1}})\omega_{jv_{j1}}+\sum\limits_{t=2}^{N}(1-p_{v_{jt}})\omega_{jv_{jt}}\prod\limits_{k=1}^{t-1}(p_{v_{jk}})\right)}{\sum\limits_{j\in J}\lambda_j\left((1-p_{v_{j1}})+\sum\limits_{t=2}^{N}(1-p_{v_{jt}})\prod\limits_{k=1}^{t-1}(p_{v_{jk}})\right)}. \tag{5.26}$$

where $\omega_{jv_{jt}}$ is the mean travel time between region j and vehicle location v_{jt}.

Similar to the expression of $F_{T_{jv_{j1}}}(\tau)$ in (5.25) in terms of the decision variables, $\omega_{jv_{j1}}$ in (5.26) is expresses as $\sum\limits_{i\in I}a_{ijN}\omega_{ij}$ in the objective function in (5.27) and Model **MMRT** which minimizes mean response time of the system is constructed as follows:

$$\text{Min} \quad \frac{\begin{aligned}&\sum_{j\in J}\lambda_j\Bigg((1-\sum_{i\in I}s_{ijN})\sum_{i\in I}a_{ijN}\omega_{ij} \\ &\qquad +\sum_{t=1}^{N-1}\Big((1-\sum_{i\in I}s_{ijt})\sum_{i\in I}a_{ijt}\omega_{ij}\prod_{k=t+1}^{N}\sum_{i\in I}s_{ijk}\Big)\Bigg)\end{aligned}}{\sum\limits_{j\in J}\lambda_j\left((1-\sum\limits_{i\in I}s_{ijN})+\sum\limits_{t=1}^{N-1}\left((1-\sum\limits_{i\in I}s_{ijt})\prod\limits_{k=t+1}^{N}\sum\limits_{i\in I}s_{ijk}\right)\right)} \tag{5.27}$$

$$\text{s. to:} \quad (5.12)-(5.23).$$

Model **MMRT** has nonlinear terms in the objective function similar to **MESD** and **MECD**. Notice that (5.27) now includes division of decision variables in addition to the multiplication. Hence, Model **MMRT** might be harder to evaluate with package solvers.

For the models **MESD**, **MECD** and **MMRT**, the objective functions are constructed based on the assumption of vehicles' being independent of each other. This assumption results in busy probabilities such that they are upper bounds for the busy probabilities of vehicles in the exact system. In the next section, we propose another approach for the approximation of busy probabilities and estimation of performance measures of the EMS system in the mathematical models in order to the address the estimation of busy probabilities at the upper bounds.

5.2.4 Order of districting for the approximation of busy probabilities and performance measures

In the approximation of the busy probabilities, p_i's, in Section 5.1, it is assumed that a vehicle responds to all regions in line with their demand rates without any prioritization such as the closeness to the regions in regard to other vehicles. It is previously mentioned in Section 5.1 that this is an upper-bound on the busy probability of this vehicle under the exact queueing model since the queueing models are constructed assuming that the vehicle at location i is the only vehicle responding to the demand calls from all regions. However, another vehicle would respond to the call in the exact system if it is closer and available at the time of the call.

Order of districting approach is utilized for the calculation of busy probabilities which would better approximate the busy probabilities of the vehicles instead of estimating the upper bounds. Let d be a parameter stating the order of districting level to be used. Enforcing order of districting level of d assumes that a vehicle serves the regions only if it is at most the d^{th} closest vehicle to a demand region.

The order of districting approach could be applied only on the approximation of the busy probabilities or applied on both the approximation of the busy probabilities and the estimation of performance measures such as expected satisfied demand in unit time or mean response time. We first start with applying order of districting on the approximation of busy probability.

Let f_i be a decision variable that is equal to the probability that the vehicle located at location i is free. This probability is estimated from the queueing models as $f_i = \pi_0^i$ for QM_i. Then, f_i could be written as follows:

$$f_i = \frac{1}{\sum\limits_{j \in J} \frac{\lambda_j}{\mu_{ji}} + 1}, \qquad \forall i \in I.$$

In order to apply order of districting on the estimation of the probability that a vehicle is free, decision variable a_{ijk} is used in the formulation. When the order of districting level is set to d, f_i is estimated as

$$f_i = \frac{1}{\sum\limits_{j \in J} \sum\limits_{k=N-d+1}^{N} a_{ijk} \frac{\lambda_j}{\mu_{ji}} + 1}, \qquad \forall i \in I, \tag{5.28}$$

where vehicle i is assumed to only serve regions where it is the d^{th} closest vehicle at the most. With an arithmetic operation, (5.28) could be expressed as

$$f_i + \sum_{j \in J} \sum_{k=N-d+1}^{N} f_i a_{ijk} \frac{\lambda_j}{\mu_{ji}} = 1, \qquad \forall i \in I, \tag{5.29}$$

where a nonlinear term, $f_i a_{ijk}$, appears in the equation. This term could be linearized by defining a new decision variable, $v_{ijk} = f_i a_{ijk}$. The following set of constraints are used to replace $f_i a_{ijk}$ with the new decision variable v_{ijk}.

$$0 \leq v_{ijk} \leq a_{ijk}, \qquad \forall i \in I, \forall j \in J, k = 1, \ldots, N, \tag{5.30}$$

$$f_i - (1 - a_{ijk}) \leq v_{ijk} \leq \; f_i, \qquad \forall i \in I, \forall j \in J, k = 1, \ldots, N, \tag{5.31}$$

Replacing $f_i a_{ijk}$ with v_{ijk} in (5.29) would make the definition of f_i as follows:

$$f_i + \sum_{j \in J} \sum_{k=1}^{3} v_{ijk} \frac{\lambda_j}{\mu_{ji}} = x_i, \qquad \forall i \in I. \tag{5.32}$$

Note that the right-hand side in (5.29) is replaced with binary decision variable x_i since f_i should be assigned a positive value only if there is a vehicle located at location i.

Since we have $p_i = 1 - f_i$, (5.12) is rewritten as follows:

$$p_i = (1 - f_i) x_i, \qquad \forall i \in I$$

which could be expressed in the form of two linear inequalities in the mathematical model as

$$0 \leq p_i \leq \quad x_i, \qquad \forall i \in I, \tag{5.33}$$

$$(1-f_i)-(1-x_i) \leq p_i \leq (1-f_i), \qquad \forall i \in I. \tag{5.34}$$

Eventually, **MESD-P** model where order of districting is applied only on the estimation of busy probabilities is constructed with the addition of (5.30), (5.31), (5.32), (5.33) and (5.34) as follows:

$$\begin{aligned} \text{Max} \quad & \sum_{j \in J} \lambda_j \left((1 - \sum_{i \in I} s_{ijN}) + \sum_{t=1}^{N-1} \left((1 - \sum_{i \in I} s_{ijt}) \prod_{k=t+1}^{N} \sum_{i \in I} s_{ijk} \right) \right) \qquad (5.35) \\ \text{s. to:} \quad & (5.13) - (5.23), \\ & (5.30) - (5.34). \end{aligned}$$

Similarly, order of districting could be applied on the estimation of the performance measures in the objective function as well. The objective function in (5.35) which computes the expected satisfied demand in unit time (ED) is reformulated incorporating the order of districting level as follows:

$$\sum_{j \in J} \lambda_j \left((1 - \sum_{i \in I} s_{ijN}) + \sum_{t=N-d+1}^{N-1} \left((1 - \sum_{i \in I} s_{ijt}) \prod_{k=t+1}^{N} \sum_{i \in I} s_{ijk} \right) \right). \tag{5.36}$$

Then, the expected satisfied demand now is estimated based on the assumption that at most the d^{th} closest vehicle could satisfy a demand call from a region.

Assume that the order of districting level is set to 3. The expression in (5.36) would be as follows:

$$\sum_{j \in J} \lambda_j \left((1 - \sum_{i \in I} s_{ijN}) + \sum_{t=N-2}^{N-1} \left((1 - \sum_{i \in I} s_{ijt}) \prod_{k=t+1}^{N} \sum_{i \in I} s_{ijk} \right) \right)$$

which is equal to

$$\begin{aligned} \sum_{j \in J} \lambda_j \Bigg(& (1 - \sum_{i \in I} s_{ijN}) + (1 - \sum_{i \in I} s_{ij(N-1)}) \sum_{i \in I} s_{ijN} \\ & + (1 - \sum_{i \in I} s_{ij(N-2)}) \sum_{i \in I} s_{ijN} \sum_{i \in I} s_{ij(N-1)} \Bigg). \end{aligned}$$

When arithmetic operations are performed, the expression above is equal to

$$\sum_{j\in J}\lambda_j\Bigg(1-\sum_{i\in I}s_{ijN}+\sum_{i\in I}s_{ijN}-\sum_{i\in I}s_{ij(N-1)}\sum_{i\in I}s_{ijN} \\ +\sum_{i\in I}s_{ijN}\sum_{i\in I}s_{ij(N-1)}-\sum_{i\in I}s_{ijN}\sum_{i\in I}s_{ij(N-1)}\sum_{i\in I}s_{ij(N-3)}\Bigg)$$

which is further reduced to

$$\sum_{j\in J}\lambda_j\left(1-\sum_{i\in I}s_{ijN}\sum_{i\in I}s_{ij(N-1)}\sum_{i\in I}s_{ij(N-2)}\right).$$

Hence, (5.36) boils down to the following expression without loss of generality when order of districting level is set to d:

$$\sum_{j\in J}\lambda_j\left(1-\prod_{k=N-d+1}^{N}\sum_{i\in I}s_{ijk}\right).$$

Then, **MESD-PO** where order of districting is applied for both estimating the busy probabilities and the expected satisfied demand is written in explicit form as follows

$$\begin{aligned}
\text{Max}\quad & \sum_{j\in J}\lambda_j\left(1-\prod_{k=N-d+1}^{N}\sum_{i\in I}s_{ijk}\right) & \\
\text{s. to:}\quad & \sum_{i\in I}x_i=N & \\
& f_i+\sum_{j\in J}\sum_{k=N-d+1}^{N}v_{ijk}\frac{\lambda_j}{\mu_{ji}}=x_i, & \forall i\in I,\\
& f_i-(1-a_{ijk})\le v_{ijk}\le f_i, & \forall i\in I,\forall j\in J,k=1,\dots,N,\\
& 0\le v_{ijk}\le a_{ijk}, & \forall i\in I,\forall j\in J,k=1,\dots,N,\\
& 0\le p_i\le x_i, & \forall i\in I,\\
& (1-f_i)-(1-x_i)\le p_i\le(1-f_i), & \forall i\in I,\\
& y_{jk}=\sum_{i\in I}a_{ijk}\omega_{ij}, & \forall j\in J,k=1,\dots,N,\\
& y_{j(k+1)}\le y_{jk}, & \forall j\in J,k=1,\dots,N-1,\\
& \sum_{k=1}^{N}a_{ijk}\le x_i, & \forall i\in I,\forall j\in J,\\
& \sum_{i\in I}a_{ijk}=1, & \forall j\in J,k=1,\dots,N,
\end{aligned}$$

$$
\begin{aligned}
&0 \le s_{ijk} \le a_{ijk}, && \forall i \in I, \forall j \in J, k = 1, \ldots, N,\\
&p_i - (1 - a_{ijk}) \le s_{ijk} \le p_i, && \forall i \in I, \forall j \in J, k = 1, \ldots, N\\
&x_i \in \{0, 1\}, && \forall i \in I,\\
&p_i \ge 0, && \forall i \in I,\\
&a_{ijk} \in \{0, 1\}, && \forall i \in I, \forall j \in J, k = 1, \ldots, N,\\
&y_{jk} \ge 0, && \forall j \in J, k = 1, \ldots, N.
\end{aligned}
$$

Therefore, two variants of the model **MESD** are constructed by applying order of districting, **MESD-P** and **MESD-PO**. The motivation for this is to check for the effect of approximating only the busy probabilities with order of districting, and the effect of approximating busy probabilities and estimating the expected satisfied demand with order of districting on the quality of the best solution, independently.

Similarly, the variants of models **MECD** and **MMRT** with order of districting applied both on the approximation of the busy probabilities and estimation of the objective function are constructed as **MECD-PO** and **MMRT-PO**.

Model **MECD-PO** where order of districting is applied on approximation of busy probabilities and estimation of the expected covered demand in unit time is:

$$
\begin{aligned}
\text{Max} \quad & \sum_{j \in J} \lambda_j \Bigg((1 - \sum_{i \in I} s_{ijN}) \sum_{i \in I} a_{ijN} F_{ij}(\tau) \\
& \quad + \sum_{t=N-d+1}^{N-1} \Bigg((1 - \sum_{i \in I} s_{ijt}) \sum_{i \in I} a_{ijt} F_{ij}(\tau) \prod_{k=t+1}^{N} \sum_{i \in I} s_{ijk} \Bigg) \Bigg)
\end{aligned} \tag{5.37}
$$

$$
\begin{aligned}
\text{s. to:} \quad & (5.13) - (5.23),\\
& (5.30) - (5.34).
\end{aligned}
$$

Model **MMRT-PO** where order of districting at level d is applied on both approximating the busy probabilities and estimating the mean response time of the system is as follows:

$$\text{Min} \quad \frac{\displaystyle\sum_{j\in J}\lambda_j\Bigg((1-\sum_{i\in I}s_{ijN})\sum_{i\in I}a_{ijN}\omega_{ij} + \sum_{t=N-d+1}^{N-1}\Bigg((1-\sum_{i\in I}s_{ijt})\sum_{i\in I}a_{ijt}\omega_{ij}\prod_{k=t+1}^{N}\sum_{i\in I}s_{ijk}\Bigg)\Bigg)}{\displaystyle\sum_{j\in J}\lambda_j\Bigg((1-\sum_{i\in I}s_{ijN})+\sum_{t=N-d+1}^{N-1}\Bigg((1-\sum_{i\in I}s_{ijt})\prod_{k=t+1}^{N}\sum_{i\in I}s_{ijk}\Bigg)\Bigg)} \tag{5.38}$$

s. to: $(5.13) - (5.23)$,

$(5.30) - (5.34)$.

5.3 Experimental Study

An extensive experimental study is conducted with various problem instances having different network configurations as in Chapter 3 and 4.

The purpose of the experimental study is to check the quality of the optimal solutions of the proposed models on toy data and real-life data, models' performance under package solvers, models' performance with respect to the decomposition methods proposed in 4 and a well-known model, MEXCLP by Daskin (1983), from literature. The modeling approach is tested on two additional objective functions as well, different from the ones constructed in Section 5.2. The details of the experimental study are as follows.

In the proposed models in Section 5.2, three performance measures are used in the objective function: expected satisfied demand in unit time (ED), expected covered demand under threshold travel time in unit time (CD) and mean response time of the system (R). In order to check the quality of the best solutions of the proposed models, three models having these measures in the objective function are constructed.

Let $x_i, \forall i \in I$ be a binary decision variable stating whether a vehicle located in vehicle location i or not and $\vec{x}$ be the vector of decision variables, x_i's, representing a solution.

Model $\boldsymbol{P_E}$ which maximizes the expected satisfied demand in unit time is as follows:

$$
\begin{aligned}
(\boldsymbol{P_E}) \qquad & \text{Maximize} \quad && ED(\vec{x}) && (5.39)\\
& \text{subject to:} \quad && \sum_{i \in I} x_i = N && (5.40)\\
& && x_i \in \{0,1\}, \qquad \forall i \in I. && (5.41)
\end{aligned}
$$

Model $\boldsymbol{P_C}$ which maximizes the expected covered demand in unit time is as follows:

$$
\begin{aligned}
(\boldsymbol{P_C}) \qquad & \text{Maximize} \quad && CD(\vec{x}) && (5.42)\\
& \text{subject to:} \quad && \sum_{i \in I} x_i = N && (5.43)\\
& && x_i \in \{0,1\}, \qquad \forall i \in I. && (5.44)
\end{aligned}
$$

Model $\boldsymbol{P_R}$ which minimizes the mean response time of the system is:

$$
\begin{aligned}
(\boldsymbol{P_R}) \qquad & \text{Minimize} \quad && R(\vec{x}) && (5.45)\\
& \text{subject to:} \quad && \sum_{i \in I} x_i = N && (5.46)\\
& && x_i \in \{0,1\}, \qquad \forall i \in I. && (5.47)
\end{aligned}
$$

A discrete event simulation model is constructed and coded in Matlab environment to simulate the emergency medical systems and evaluate the objective function value of a location solution for models $\boldsymbol{P_E}$, $\boldsymbol{P_C}$ and $\boldsymbol{P_R}$ since the exact queueing model is computationally expensive.

The optimal solution of **MESD** and **MMRT** models (including variants) are compared with the best solution for $\boldsymbol{P_E}$ or $\boldsymbol{P_R}$ in order to see the performance of models in comparison to simulation. The optimal solutions of **MESD** and **MMRT** are found by complete enumeration of the feasible solutions. To select the best solutions for $\boldsymbol{P_E}$ and $\boldsymbol{P_R}$, *KN*++ algorithm which utilizes the discrete event simulation model is used. To compare the optimal solutions of the models with the best solutions, the objective function values are estimated from a separate simulation run.

The best solutions of the proposed models from complete enumeration are also compared against the decomposition method DM-S-CF which is showed to perform very promising in finding near best optimal solution for the objective of minimizing mean response time in Chapter 4. In the experiments, DM-S-CF is used to evaluate the objective function value of a feasible solution for models $\boldsymbol{P_E}$ and $\boldsymbol{P_R}$. With this analysis, the performance of mathematical models with closed-form expressions of performance measures are compared to an computationally cumbersome algorithmic approach.

MECD variants are compared with $\boldsymbol{P_C}$ and MEXCLP by Daskin (1983) where busy probabilities of facilities are taken into account in estimating the performance measure used in the objective function similar to our models. The effect of estimating the busy probabilities specific to a feasible solution is checked in the comparsion with MEXCLP where busy probabilities are calculated in advance and independent of any feasible solutions.

MESD-PO, **MECD-PO** and **MMRT-PO** are also evaluated using package solver BARON in order to report the gap in the performance measures of the optimal solutions from complete enumeration and solver run. This is to show the performance of the models with package solvers where objective function is nonlinear.

Similar to the analysis in Chapter 4, the proposed modeling approach are used to construct the models $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ from Chapter 3 which are the models found performing well in terms of equity. Instead of using objective functions in the form of functions of solutions, closed form formulations are used to approximately represent $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$. The changes in the mean response time, variance of the region-wise mean response time and Gini coefficient are checked similarly for the optimal solutions of the models in order to observe whether similar equity results are achieved when optimal solutions are found based on the models constructed according to the approach proposed in this chapter.

Lasty, **MESD-PO**, **MECD-PO** and **MMRT-PO** models are used to locate emergency vehicles on Edmonton Data which is previously used in Chapter 4 to show the performance of models on larger problems.

The experimental study is summarized under seven main comparisons in Table 5.3 as follows:

Table 5.3: Summary of the experimental study

Aim	Models	Solution Type	Solution Method	Data Set
Performance compared to simulation	**MESD** vs. P_E **MMRT** P_R	Optimal vs. Best	Comp. enum. vs. *KN*++	Toy
Performance compared to decomposition methods	**MESD-PO** vs. P_E **MMRT-PO** P_R	Optimal vs. Best	Comp. enum. vs. DM-S-CF	Toy
Performance compared to MEXCLP	**MECD** vs. P_C MEXCLP	Optimal vs. Best	Comp. enum. vs. *KN*++	Toy
Performance under package solvers compared to optimal solutions	**MESD-PO** vs. **MESD-PO** **MMRT-PO** **MMRT-PO** **MECD-PO** **MECD-PO**	Optimal vs. Best	Comp. enum. vs. BARON solver	Toy
Performance under package solvers compared to simulation	**MESD-PO** vs. P_E **MMRT-PO** P_R **MECD-PO** P_C	Best vs. Best	BARON solver vs. *KN*++	Toy
Performance under real life data compared to simulation	**MESD-PO** vs. P_E **MMRT-PO** P_R **MECD-PO** P_C	Optimal vs. Best	Comp. enum. vs. *KN*++	Edmonton
Performance of different objective func.s compared to simulation	$P_4 - PO$, vs. P_4 $P_7 - PO$ P_7	Optimal vs. Best	Comp. enum. vs. *KN*++	Toy

5.3.1 Test bed

For the experiments, the same sets of regions from Chapter 3 and 4 are used, which include three forms *Uniform*, *Center-accumulated* and *Outer-accumulated* with 15 demand regions.

The number of vehicles is set to 4, 5 or 6. The demand rate is assumed to be the same for every region as 0.5 units per hour. The incident handling rate is set to 3, 6, 9 and 12.

In total, 36 instances are generated based on a full factorial design using three factors: *Form* with three levels, *Number of Vehicles* with three levels (4,5 and 6) and *Incident Handling Rate* with four levels (3, 6, 9 and 12).

For the experiments, order of districting level (d) is set to 3, and threshold travel time (τ) is set to 10.

5.3.2 Selection of the best solution and evaluation of performance measures

Since there is no closed-form formulation for the exact queueing model in the study, a discrete event simulation model is constructed to simulate the emergency medical systems in Chapter 3.

The best solutions for each instance under models $\boldsymbol{P_E}$, $\boldsymbol{P_C}$ and $\boldsymbol{P_R}$ are found using *KN*++ algorithm by Kim and Nelson (2006) which uses the discrete event simulation model to assess the objective function values.

For *KN*++ algorithm, first-stage sample size n_0 is set to 100,000 demand calls. The initial number of batches, b_0, is set to 10. The confidence level γ is set to 0.05, indifference-zone parameter δ to 0.01, and parameter c to 1. *KN*++ algorithm is stopped when the best performing alternative has a 0.1 % difference in the objective function value from the worst performing among the remaining alternatives. Then, the best solution is selected randomly from the remaining alternatives.

5.3.3 Performance of MESD and MMRT models

The performance of **MESD** and **MMRT** models is explored by comparing the optimal solutions of the models to the best solutions of $\boldsymbol{P_E}$ and $\boldsymbol{P_R}$ obtained with *KN*++. The optimal solutions of **MSED** models are compared with the best solutions of $\boldsymbol{P_E}$, and **MMRT** with $\boldsymbol{P_R}$.

In order to evaluate the models' ability to differentiate the feasible solutions in line with the actual performance and deliver an optimal solution performing close to the best solution with respect to *KN*++, the optimal solutions of the mathematical models are found by complete enumeration. After finding the best solutions with *KN*++ algorithm and the optimal solutions of the models based on complete enumeration for each instance, the objective function values are estimated by running the discrete event simulation model for 10 independent replications. A total of 250,000 demand calls is simulated for every solution at hand in the simulation model. The warm-up period is selected as 50,000 calls, and the objective function values are reported accordingly. The analysis of the performances of the models concerning the optimal solutions is done based on this experiment.

Let $\vec{x}^*_{(.)}$ be the optimal solution under model (.) and $O_{Sim}(\vec{x}^*_{(.)})$ denotes the mean of 90% confidence interval of the objective function value for this solution based on 10 independent replications of the simulation model. Subsequently, let $\vec{x}^*_{P_{(.)}}$ represent the best solution evaluated by KN++ algorithm for model $P_{(.)}$ having the same performance measure in the objective function with model (.) and $O_{Sim}(\vec{x}^*_{P_{(.)}})$ denote the mean of the 90% confidence interval of the objective function value for this solution based on 10 independent replications.

The percent deviation from the best solution is denoted by $\%\Delta^*_{Sim}$ and is found by

$$\%\Delta^*_{Sim} = I_{\left\{\vec{x}^*_{(.)} \neq \vec{x}^*_{P_{(.)}}\right\}} \left(100 \frac{O_{Sim}(\vec{x}^*_{(.)}) - O_{Sim}(\vec{x}^*_{P_{(.)}})}{O_{Sim}(\vec{x}^*_{P_{(.)}})} \right),$$

where I is the indicator function being equal to 0 if the solutions are the same.

$\%\Delta^*_{Sim}$ results are presented in Table 5.4 for 36 instances with respect to **MESD**, **MESD-P** and **MESD-PO**. $\%\Delta^*_{Sim}$ values are reported as zero if the optimal solution under a model is same with the best solution for $\boldsymbol{P_E}$ evaluated by *KN*++. Suppose the difference between objective function value of the optimal solution of a model and the best solution by *KN*++ is not significant, meaning confidence intervals of the objective function values coincide. In that case, the percentage deviation found is marked with an asterisk in the table.

Table 5.4: $\%\Delta^*_{Sim}$ in ED for the optimal solution of models based on the best solution for P_E in 36 instances

Form	Nb.of Vehc.	Inc. Hand. Rate	MESD	MESD-P	MESD-PO
Uniform	4	3	-0.10 *	-0.87	0.20 *
		6	-1.00	-1.29	-0.30
		9	-1.49	-1.65	-0.51
		12	-1.75	-1.79	-0.64
	5	3	-0.36	-4.83	0.28
		6	-1.03	-5.19	0.13 *
		9	-1.17	-5.26	0.03 *
		12	-0.97	-4.87	0.17
	6	3	-0.30	-3.36	-0.88
		6	-0.51	-3.07	-0.56
		9	-0.49	-2.66	-0.37
		12	-0.55	-2.43	-0.45
Center-Acc.	4	3	0.17 *	-4.87	0.15 *
		6	-0.17	-4.87	-0.10 *
		9	0.05 *	-4.35	0.15 *
		12	-0.20	-4.30	-0.20
	5	3	-0.04 *	-5.13	0.09 *
		6	0.02 *	-4.34	0.26
		9	0.27	-3.47	0.58
		12	-0.30	-3.68	0.03 *
	6	3	-0.16 *	-3.86	0.08 *
		6	-0.09	-1.87	0.16
		9	-0.08	-1.26	0.08
		12	-0.29	-1.06	-0.09
Outer-Acc.	4	3	-1.53	-4.70	-0.32
		6	-2.02	-7.71	-0.08 *
		9	-2.99	-8.90	-0.79
		12	-3.37	-9.28	-0.87
	5	3	-1.42	-5.27	0.70

Continued on next page

Table 5.4 – continued from previous page

Form	Nb.of Vehc.	Inc. Hand. Rate	MESD	MESD-P	MESD-PO
		6	-3.14	-7.27	0.74
		9	-4.26	-8.20	-0.02 *
		12	-4.37	-8.19	-0.05 *
	6	3	-2.79	-5.74	-0.86
		6	-3.44	-6.18	0.05 *
		9	-3.66	-6.11	-0.19
		12	-3.31	-5.30	0.01 *

In order to summarize the results in Table 5.4 and make comparisons of the models easier, four statistics are computed as in Chapter 4:

- **Number of instances for which the differences are not statistically significant:** the number of instances where **MESD** variants has an optimal solution same with the best solution for $\boldsymbol{P_E}$ reported by *KN*++ or the optimal solutions are not statistically different in terms of the expected satisfied demand (ED),
- **Average** $\%\Delta^*_{Sim}$**:** the average $\%\Delta^*_{Sim}$ over all instances,
- **Average** $\%\Delta^*_{Sim}$ **over instances for which the differences are statistically significant:** the average $\%\Delta^*_{Sim}$ over instances having optimal solutions that are statistically different,
- **Average absolute** $\%\Delta^*_{Sim}$ **over instances for which the differences are statistically significant:** the average absolute $\%\Delta^*_{Sim}$ over instances having optimal solutions that are statistically indifferent.

These statistics are reported in Table 5.5 for **MESD** variants.

According to the results in Table 5.4 and 5.5, **MESD-PO** is the best model in finding solutions that are close to the best solution reported by *KN*++ for the objective of maximizing expected satisfied demand where $\%\Delta^*_{Sim}$ is less than 1% for all instances. In fourteen instances, **MESD-PO** finds a solution that has a statistically insignificant deviation from the best solution. **MESD** is the second best performing model,

Table 5.5: Summary of $\%\Delta^*_{Sim}$ in ED results for the comparison of the optimal solutions with P_E in 36 instances

Statistics	MESD	MESD-P	MESD-PO
Number of instances for which the differences are not statistically significant	6	0	14
Average $\%\Delta^*_{Sim}$	-1.30	-4.53	-0.09
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	-1.56	-4.53	-0.18
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	1.31	4.53	0.45

whereas **MESD-P** has optimal solutions for which the differences are statistically significant in all instances with an average of 4.5%.

The average deviation and average absolute deviation without insignificant results for **MESD-PO** are lower than 0.5%. Therefore, it is seen that order of districting employed in model **MESD-PO** (in estimation of the expected satisfied demand and the busy probabilities) helps the model finding a close enough solution to the best. Since a pairwise comparison is not made for each feasible solution between **MESD-PO** and simulation model, it does not necessarily mean that the estimation of the objective function gets better, however the differentiation of the feasible solutions is better than **MESD**.

It is seen that **MESD-P** performs worse than **MESD** and **MESD-PO** which means using order of districting for only estimation of busy probabilities are not useful.

In summay, **MESD-PO** is a promising model in maximizing expected satisfied demand.

Similar to the previous experiment above, $\%\Delta^*_{Sim}$ results for 36 instances under **MMRT** and **MMRT-PO** are presented in Table 5.6. $\%\Delta^*_{Sim}$ values are reported as zero if the optimal solution under a model is the same with the best solution for $\boldsymbol{P_R}$ evaluated by *KN*++.

The base model $\boldsymbol{P_R}$ and the models **MMRT** and **MMRT-PO** share the same measure in the objective function, mean response time (R). Due to the observations in Chapter 3 where it is seen that maximizing expected satisfied demand and minimizing mean response time are similar objectives in terms of the resulting best solution, **MESD-PO** is also added to this comparison. $\%\Delta^*_{Sim}$ for this model is calculated based on the mean response time value of the optimal solution of **MESD-PO** where it is assessed from 10 independent replications of the simulation model as in the other comparisons.

In Table 5.6, $\%\Delta^*_{Sim}$ in R for **MMRT**, **MMRT-PO** and **MESD-PO** are reported.

Table 5.6: $\%\Delta^*_{Sim}$ in R of the optimal solution of models based on the best solution for P_R in 36 instances

Form	Nb.of Vehc.	Inc. Hand. Rate	MMRT	MMRT-PO	MESD-PO
Uniform	4	3	0	2.30	2.45
		6	0	2.22	2.39
		9	-0.50	2.08	2.01
		12	0	2.61	2.58
	5	3	2.22	6.11	0.14 *
		6	0.49	5.82	0.16 *
		9	3.83	8.71	2.47
		12	4.91	9.52	3.63
	6	3	3.89	1.74	6.46
		6	2.30	2.72	9.35
		9	3.94	3.40	11.45
		12	6.01	3.41	13.16
Center-Acc.	4	3	0	9.43	0
		6	0.65	7.04	0.56
		9	1.64	6.50	1.57
		12	2.04	6.16	2.13
	5	3	0.44 *	8.21	0
		6	2.28	4.64	-0.05 *
		9	3.39	3.22	0.26 *
		12	4.49	2.48	0.66
	6	3	0.98	15.73	-0.39
		6	4.25	8.83	1.93
		9	7.62	0.12 *	4.58
		12	11.05	0.31 *	7.47
Outer-Acc.	4	3	0	4.62	0.72

Continued on next page

Table 5.6 – continued from previous page

Form	Nb.of Vehc.	Inc. Hand. Rate	MMRT	MMRT-PO	MESD-PO
		6	0.99	1.36	1.56
		9	1.45	0	1.83
		12	3.53	0	4.06
	5	3	1.05	7.22	0
		6	7.47	7.16	0
		9	11.93	5.87	0
		12	14.71	5.44	0.13 *
	6	3	7.97	3.67	6.75
		6	21.24	0	3.65
		9	31.75	0	4.00
		12	37.15	0	4.69

Similar to the previous experiment, summary statistics for $\%\Delta^*_{Sim}$ values are reported in Table 5.7.

Table 5.7: Summary of $\%\Delta^*_{Sim}$ in R results for the comparison of the optimal solutions with $\boldsymbol{P_R}$ in 36 instances

Statistics	MMRT	MMRT-PO	MESD-PO
Number of instances for which the differences are not statistically significant	6	7	10
Average $\%\Delta^*_{Sim}$	5.70	4.41	2.84
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	6.82	5.46	3.91
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	6.86	5.46	3.94

The results in Table 5.7 show that **MMRT** and **MMRT-PO** find near optimal solutions in six and seven instances, respectively. The average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant is 6.86% for **MMRT**. When order of districting is applied (**MMRT-PO**), this value decreases to 5.46%.

The deviations of the optimal solutions of **MMRT** and **MMRT-PO** from the best solutions of $\boldsymbol{P_R}$ are more significant than they are in the previous comparison for **MESD** variants. Hence, it is justifiable to state that this modeling approach performs better in terms of differentiating feasible solutions in expected satisfied demand measure than in mean response time. Although estimation of ED and R coincide where R has additional mean travel time multiplier in the numerator, the estimation of R results in models with less power in differentiating feasible solutions, probably due to mean travel time magnifying the estimation errors.

On the other hand, **MESD-PO** which is constructed to maximize expected satisfied demand in unit time proves promising according to the results in Table 5.7. The mean response times of the optimal solutions of **MESD-PO** deviate by 2.84% on the average from the best mean response times obtained with *KN*++ for $\boldsymbol{P_R}$. **MESD-PO** finds solutions for which the differences are not statistically significant in ten instances. This is greater than the ones for **MMRT** and **MMRT-PO**, although they are specifically modeled to minimize mean response time. This observation is particularly important since it shows that maximizing expected satisfied demand (**MESD-PO**) performs well for the objective minimizing mean response time as well.

Overall, maximization of expected satisfied demand results in optimal solutions that are close to the optimal solutions under mean response time minimization similar to the previous observations. Therefore, an analytical proof is provided in Appendix C for the special case of locating one vehicle showing that two objective functions are equivalent. In addition, numerical results are reported for the case of locating two vehicles by solving the balance equations of the queueing model for the EMS system and finding the optimal solutions with complete enumeration.

5.3.4 Comparison of MESD-PO, MMRT-PO with DM-S-CF

In an effort to show the performances of models **MESD-PO** and **MMRT-PO** against a more sophisticated algorithmic approach, the decomposition method DM-S-CF proposed in Chapter 4 is used to estimate the objective function value of models $\boldsymbol{P_E}$ and $\boldsymbol{P_R}$. In Chapter 4, it is shown that DM-S-CF is a well performing estimation method for mean response time (R) and near optimal solutions are found in thirty-one out of thirty-six instances when used with complete enumeration.

In the first set of experiments, the optimal solutions **MESD-PO** and **MMRT-PO** and the optimal solution of $\boldsymbol{P_E}$ evaluated with DM-S-CF are compared to the best solution for $\boldsymbol{P_E}$ obtained with *KN*++ in expected satisfied demand (ED).

In Table 5.8, $\%\Delta^*_{Sim}$ in ED for **MESD-PO**, **MMRT-PO** and $\boldsymbol{P_E}$ with DM-S-CF are reported.

Table 5.8: $\%\Delta^*_{Sim}$ in ED of the optimal solution of models based on the best solution for $\boldsymbol{P_E}$ in 36 instances

Form	Nb.of Vehc.	Inc. Hand. Rate	MESD-PO	MMRT-PO	P_E with DM-S-CF
Uniform	4	3	0.20 *	0.25	0.81
		6	-0.30	-0.22	0.53
		9	-0.51	-0.60	0.08 *
		12	-0.64	-0.56	0
	5	3	0.28	-1.23	0.33 *
		6	0.13 *	-1.08	0.19
		9	0.03 *	-0.85	0.21
		12	0.17	-0.59	0.43
	6	3	-0.88	-0.23	0.36
		6	-0.56	0.00 *	0.11 *
		9	-0.37	-0.05 *	0.12
		12	-0.45	-0.15	-0.28
Center-Acc.	4	3	0.15 *	-2.15	-0.08 *
		6	-0.10 *	-1.80	0.04 *
		9	0.15 *	-1.26	0.24 *
		12	-0.20	-1.39	0.14
	5	3	0.09 *	-1.71	0.01 *
		6	0.26	-0.70	0.31
		9	0.58	0.03 *	0.36
		12	0.03 *	-0.37	-0.05 *
	6	3	0.08 *	-2.61	-0.11 *
		6	0.16	-0.86	0.06 *
		9	0.08	0.08	-0.19
		12	-0.09	-0.05	-0.27
Outer-Acc.	4	3	-0.32	-1.62	0
		6	-0.08 *	0.02 *	0.69
		9	-0.79	0.04 *	0.09 *
		12	-0.87	0.26	0.27
	5	3	0.70	-1.27	0.78

Continued on next page

Table 5.8 – continued from previous page

Form	Nb.of Vehc.	Inc. Hand. Rate	MESD-PO	MMRT-PO	P_E with DM-S-CF
		6	0.74	-0.73	0.75
		9	-0.02 *	-1.11	0.01 *
		12	-0.05 *	-0.85	0.10 *
	6	3	-0.86	-0.30	0.49
		6	0.05 *	0.48	0.38
		9	-0.19	0.06 *	0.03 *
		12	0.01 *	0.16	0.19

The results in Table 5.8 show that **MESD-PO**, **MMRT-PO** and DM-S-CF are good at finding near optimal solutions in ED in almost all instances. $\%\Delta^*_{Sim}$ values are smaller than 1% in all instances for **MESD-PO** and DM-S-CF. **MMRT-PO** has upto 2.6% deviation in expected satisfied demand in some instance. In Table 5.9, summary statistics for $\%\Delta^*_{Sim}$ in ED are reported.

Table 5.9: Summary of $\%\Delta^*_{Sim}$ in ED results for the comparison of the optimal solutions with $\boldsymbol{P_E}$ in 36 instances

Statistics	MESD-PO	MMRT-PO	P_E with DM-S-CF
Number of instances for which the differences are not statistically significant	14	6	16
Average $\%\Delta^*_{Sim}$	-0.09	-0.64	0.20
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	-0.18	-0.77	0.31
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	0.45	0.85	0.39

The statistics in Table 5.9 show that **MESD-PO** is as good as DC-S-MF where the average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant are 0.45 and 0.39, respectively. DC-S-MF method is a decomposition method that approximates the steady state probabilities of the exact queueing system to estimate the objective function value. It requires an algorithmic approach to solve nonlinear set of equations in approximating the steady state probabilities. Considering this, the performance of **MESD-PO** is very promising as a mixed integer nonlinear program with linear constraints.

In another set of experiments, the optimal solutions of **MESD-PO** and **MMRT-PO** based on complete enumeration, and the optimal solution of $\boldsymbol{P_R}$ evaluated with DM-S-CF are compared to the best solution for $\boldsymbol{P_R}$ obtained with *KN*++ in mean response time (R).

In Table 5.8, $\%\Delta^*_{Sim}$ in R for **MESD-PO**, **MMRT-PO** and $\boldsymbol{P_R}$ with DM-S-CF are reported.

Table 5.10: $\%\Delta^*_{Sim}$ in R of the optimal solution of models based on the best solution for $\boldsymbol{P_R}$ in 36 instances

Form	Nb.of Vehc.	Inc. Hand. Rate	MESD-PO	MMRT-PO	P_R with DM-S-CF
Uniform	4	3	2.45	2.30	0
		6	2.39	2.22	0
		9	2.01	2.09	-0.44
		12	2.58	2.61	0
	5	3	0.14 *	6.12	0
		6	0.16 *	5.81	-1.16
		9	2.47	8.70	0
		12	3.62	9.52	0
	6	3	6.46	1.75	-0.31
		6	9.34	2.72	0
		9	11.45	3.40	-0.32
		12	13.16	3.39	0
Center-Acc.	4	3	0	9.44	0.20 *
		6	0.56	7.04	-0.11 *
		9	1.57	6.50	0
		12	2.13	6.16	0
	5	3	0	8.21	0
		6	-0.05 *	4.64	-0.09 *
		9	0.26 *	3.22	0.43
		12	0.67	2.49	0.11 *
	6	3	-0.38	15.73	-0.40
		6	1.93	8.83	0
		9	4.59	0.13 *	0.23 *
		12	7.46	0.30 *	0
Outer-Acc.	4	3	0.72	4.62	-0.22 *
		6	1.55	1.36	0
		9	1.82	0	0.03 *

Continued on next page

Table 5.10 – continued from previous page

Form	Nb.of Vehc.	Inc. Hand. Rate	MESD-PO	MMRT-PO	P_R with DM-S-CF
		12	4.06	0	0
	5	3	0	7.22	0
		6	0	7.16	0
		9	0	5.88	0
		12	0.13 *	5.44	0.29 *
	6	3	6.75	3.67	0
		6	3.65	0	0
		9	4.00	0	0
		12	4.69	0	0

As stated previously, near optimal solutions are found in most of the instances when DM-S-CF is used to estimate the objective function value of $\boldsymbol{P_R}$. It is seen in Table 5.10 that DM-S-CF reports the same best solution with *KN*++ in twenty-two and solutions that the difference is not statistically significant in eight instances. According to the summary statistics for $\%\Delta^*_{Sim}$ in Table 5.11, neither **MESD-PO** nor **MMRT-PO** performs close enough to DM-S-CF in resulting mean response time of the optimal solutions. This shows that the mean response time is harder to estimate considering the simplifications in the estimation of busy probabilities of vehicles. However, **MESD-PO** is still preferable to **MMRT-PO** if the objective is to minimize mean response time, since maximization of expected satisfied demand with **MESD-PO** results in optimal solutions that have less deviation in mean response time than the deviation that the optimal solutions obtained by minimizing mean response time with **MMRT-PO** have.

5.3.5 Performance of MECD models compared to MEXCLP

Another set of models proposed in the Section 5.2 are **MECD** and **MECD-PO** where expected covered demand under a threshold travel time is maximized.

In an effort to compare the proposed models with a model from the literature, MEXCLP by Daskin (1983) is chosen. The study is an extension of classical set covering problems where a demand node is assumed to be covered if its distance to a facility is less than a threshold. MEXCLP takes busy probabilities of facilities into

Table 5.11: Summary of $\%\Delta^*_{Sim}$ in R results for the comparison of the optimal solutions with $\boldsymbol{P_R}$ in 36 instances

Statistics	MESD-PO	MMRT-PO	P_R with DM-S-CF
Number of instances for which the differences are not statistically significant	10	7	30
Average $\%\Delta^*_{Sim}$	2.84	4.41	-0.05
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	3.91	5.46	-0.37
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	3.94	5.46	0.51

account. Hence, there is a need for calculating the probability that a region is covered under particular location configuration of facilities.

Recall that λ_j is the demand rate of region j, ω_{ij} is the mean travel time between location i and region j, τ is the threshold travel time used in **MECD** model.

Let, a_{ji} be a parameter which is equal to 1 if a vehicle at location i covers region j meaning $\omega_{ij} \leq \tau$ and 0 otherwise. p states the probability that a facility is not working, analogous to being busy in our problem setting. Let x_i be a decision variable which is equal to 1 if a vehice is located at region i, and y_{kj} is a decision variable that is equal to 1 if region j is covered by at least k vehicles, 0 otherwise.

Single-vehicle version of MEXCLP where at most one vehicle is allowed at a single location, s-MEXCLP is as follows:

$$\textbf{(s-MEXCLP)} \quad \text{Maximize} \quad \sum_{j \in J} \sum_{k=1}^{N} (1-p)p^{k-1}\lambda_j y_{kj} \tag{5.48}$$

$$\text{subject to:} \quad \sum_{k=1}^{N} y_{kj} - \sum_{i \in I} a_{ji} x_i \leq 0, \qquad j \in J \tag{5.49}$$

$$\sum_{i \in I} x_i \leq N \tag{5.50}$$

$$x_i \in \{0,1\}, \qquad i \in I \tag{5.51}$$

$$y_{kj} \in \{0,1\}, \qquad j \in J, k = 1, ..., N. \tag{5.52}$$

s-MEXCLP maximizes expected demand covered where coverage is defined based on mean of travel time and a threshold. Therefore, **MECD** and s-MEXCLP could be compared since they are constructed in a similar fashion in terms to coverage. The difference in **MECD** is that, all vehicle locations can cover all demand regions but for only the demand calls that travel time realization for the call is under the threshold travel time.

MEXCLP model requires busy probability p as a parameter. The busy probability is defined identical for all facility locations in the model. To compare the effect of modeling approaches of **MECD** and MEXCLP, perfect information is supplied to MEXCLP in terms of the required busy probability parameter for each feasible solution in the experiment. The average busy probability of vehicles are estimated for each feasible solution of every instance from a simulation study with 50,000 demand calls. Then, the objective function values of all feasible solutions of s-MEXCLP model are evaluated by using the corresponding average busy probability, and the optimal solution is selected.

The optimal solutions of **MECD**, **MECD-PO** and s-MEXCLP based on complete enumeration are compared to the best solution for $\boldsymbol{P_C}$ obtained with *KN++* in expected covered demand (CD). In Table 5.12, $\%\Delta^*_{Sim}$ in CD for **MECD**, **MECD-PO** and s-MEXCLP are reported.

Table 5.12: $\%\Delta^*_{Sim}$ in CD of the optimal solution of models based on the best solution for $\boldsymbol{P_C}$ in 36 instances

Form	Nb.of Vehc.	Inc. Hand. Rate	MECD	MECD-PO	s-MEXCLP
Uniform	4	3	0.09 *	-0.77	-1.38
		6	0.48 *	-0.82	-0.86
		9	-0.26	-1.45	-0.89
		12	0.55 *	-0.40 *	0

Continued on next page

Table 5.12 – continued from previous page

Form	Nb.of Vehc.	Inc. Hand. Rate	MECD	MECD-PO	s-MEXCLP
	5	3	-0.55	0.33 *	1.01
		6	-0.70 *	-1.13	-0.65
		9	-1.21	-1.43	-0.56
		12	-1.76	-2.18	-0.67
	6	3	-1.75	0.53	-1.47
		6	-0.71	0.16 *	0.32 *
		9	-1.71	-0.72	-2.11
		12	-2.22	-0.78	-1.50
Center-Acc.	4	3	0	0.04 *	-2.60
		6	-0.31 *	-0.21 *	-2.52
		9	0.68	0.69	-1.48
		12	-0.34 *	-0.39 *	0.47 *
	5	3	0.26 *	0.51	-0.45
		6	-0.44	0.69	-0.64
		9	-1.62	-0.17 *	-2.21
		12	-1.80	-0.43	-2.89
	6	3	0.60	-0.05 *	-2.68
		6	-1.18	-0.01 *	-2.71
		9	-2.46	-0.69	-3.28
		12	-3.11	-1.19	-1.55
Outer-Acc.	4	3	-2.67	1.11	-1.76
		6	-5.88	0	-2.33
		9	-6.92	0.30 *	-1.83
		12	-9.25	-1.19	-2.20
	5	3	-5.45	1.13	-1.33
		6	-11.95	0	-1.05
		9	-12.88	1.04	-0.02 *
		12	-7.01	-0.21 *	-1.02
	6	3	-4.66	-1.79	-0.97
		6	-7.76	0.59 *	-0.91
		9	-9.76	0.64	-1.15
		12	-9.99	1.03	-2.14

According to results, it is seen that **MECD-PO** finds solutions close to the best solution reported by *KN*++ for $\boldsymbol{P_C}$ where expected covered demand is maximized. In Table 5.13, summary statistics for $\%\Delta^*_{Sim}$ show that **MECD** is not good at finding well performing solutions in terms of CD. On the other hand, average $\%\Delta^*_{Sim}$

for **MECD-PO** are better than s-MEXCLP and all of the optimal solutions for s-MEXCLP are statistically different from the best solution by *KN*++. This shows that incorporating location specific busy probabilities and the uncertainty in travel time which is exploited in the estimation of covered demand in **MECD-PO** increase the quality of the optimal solutions. Additionally, average $\%\Delta^*_{Sim}$ under s-MEXCLP are at around 1.5% in spite of the perfect information on the busy probabilities.

Table 5.13: Summary of $\%\Delta^*_{Sim}$ in CD results for the comparison of the optimal solutions with $\boldsymbol{P_C}$ in 36 instances

Statistics	MECD	MECD-PO	s-MEXCLP
Number of instances for which the differences are not statistically significant	8	14	10
Average $\%\Delta^*_{Sim}$	-3.16	-0.20	-1.33
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	-4.06	-0.33	-1.52
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	4.15	1.00	1.59

In addition to the comparison of models in expected conditionally satisfied demand, another comparison is made based on the mean response time (R) values of the optimal solutions reported by the models. In Table 5.14, $\%\Delta^*_{Sim}$ in R for this comparison are reported.

Table 5.14: $\%\Delta^*_{Sim}$ in R of the optimal solution of models based on the best solution for $\boldsymbol{P_R}$ in 36 instances

Form	Nb.of Vehc.	Inc. Hand. Rate	MECD	MECD-PO	s-MEXCLP
Uniform	4	3	0	2.47	3.13
		6	0	2.13	2.65
		9	-0.32 *	2.16	1.53
		12	0	2.44	1.77
	5	3	2.32	2.28	1.01
		6	0.51 *	2.24	2.07
		9	3.80	4.74	4.46
		12	5.30	6.01	3.41

Continued on next page

Table 5.14 – continued from previous page

Form	Nb.of Vehc.	Inc. Hand. Rate	MECD	MECD-PO	s-MEXCLP
	6	3	4.17	1.84	4.76
		6	3.61	3.03	1.46
		9	5.70	2.47	7.40
		12	7.22	3.35	5.85
Center-Acc.	4	3	0	0.37	4.55
		6	0.68	0.50	4.51
		9	1.79	1.30	4.86
		12	2.12	1.83	0.74
	5	3	0.25 *	0	1.53
		6	2.21	-0.14 *	2.40
		9	3.53	0.41	4.70
		12	4.49	0.95	6.46
	6	3	0.68 *	1.33	7.19
		6	4.60	0.89	8.53
		9	7.71	1.88	10.36
		12	11.05	4.05	5.53
Outer-Acc.	4	3	8.89	0.66 *	5.37
		6	13.36	1.39	5.93
		9	16.71	2.11	5.68
		12	19.90	3.69	5.60
	5	3	14.30	0	5.30
		6	28.59	0	0.68
		9	36.32	0	1.07
		12	18.32	0.58 *	2.00
	6	3	12.02	3.87	2.26
		6	27.57	0	4.57
		9	38.93	0	5.18
		12	46.17	0	12.30

The summary statistics in Table 5.15 show that **MECD-PO** also performs well in terms of the mean response time of the system resulting from the optimal solutions. s-MEXCLP finds solutions for which the differences are statistically significant in every instance whereas **MECD-PO** finds solutions for which the differences are not statistically significant in ten instances. The performance of s-MEXCLP is similar to **MESD-PO** (in Table 5.7) where the average absolute deviation from the best mean response time for **MESD-PO** is 3.94%.

Table 5.15: Summary of $\%\Delta^*_{Sim}$ in R results for the comparison of the optimal solutions with $\boldsymbol{P_R}$ in 36 instances

Statistics	MECD	MECD-PO	s-MEXCLP
Number of instances for which the differences are not statistically significant	8	10	-
Average $\%\Delta^*_{Sim}$	9.79	1.69	4.35
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	12.55	2.30	4.35
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	12.55	2.30	4.35

According to the results in Table 5.15 and 5.7, **MECD-PO** outperforms **MESD-PO** in several measures. The average absolute deviation for **MECD-PO** is 2.30% while **MESD-PO** has 3.94%. The maximum deviation in mean response time increases up to 14% in some instances under **MESD-PO** while **MECD-PO** has a more uniform distribution in terms of the deviations in instances and the maximum deviation is around 6%. This shows that **MECD-PO** which is contructed to maximize expected covered demand is even better at finding well performing solutions in mean response time than **MESD-PO** and **MMRT-PO**.

5.3.6 Performance under package solvers

The previous experiments are done based on complete enumeration of the feasible solutions of the models. In an effort to show the performance of the models under package solvers, thirty-six problem instances are solved with **MESD-PO**, **MECD-PO** and **MMRT-PO** by using BARON solver on NEOS server (Czyzyk et al. (1998), Dolan (2001) and Gropp and Moré (1997)). The computation time limit is set to 14400 seconds and the solutions reported by BARON at the end of runs are used in the analysis.

In the experiments, BARON solver reported a solution in seven, one and eighteen out of thirty-six instances before the time limit is invoked for **MESD-PO**, **MMRT-PO** and **MECD-PO**, respectively. Based on the results, the relative optimality gap for each is found by comparing the objective function value of the solution reported by the solver and the solution found by complete enumeration. Average and maximum of the relative optimality gaps are reported in Table 5.16 along with the number of instances that the optimal solution with respect to complete enumeration is found by BARON solver.

Table 5.16: Summary of BARON runs for the models in 36 instances

Statistics	MESD-PO	MMRT-PO	MECD-PO
Number of instances optimal solution found	1	6	11
Average optimality gap	0.84	3.03	1.19
Maximum optimality gap	2.68	12.00	5.30

It is seen that BARON finds the optimal solutions in one, six and eleven instances among thirty-six for **MESD-PO**, **MMRT-PO** and **MECD-PO**. The average optimality gap is 0.84%, and the maximum optimality gap is 2.68% for **MESD-PO**. **MMRT-PO** has the highest average and maximum optimality gap among models although optimal solution is found in six instances. BARON finds the optimal solutions in eleven instances for **MECD-PO**, but the average and maximum optimality gap is higher than **MESD-PO**.

In addition, $\%\Delta^*_{Sim}$ values of the solutions obtained with BARON solver are also checked. $\%\Delta^*_{Sim}$ values are found in expected satisfied demand for **MESD-PO**, in mean response time for **MMRT-PO** and in expected covered demand for **MECD-PO**. The summary statistics for $\%\Delta^*_{Sim}$ are given in Table 5.17 for three models according to BARON results.

It is seen that average absolute $\%\Delta^*_{Sim}$ in expected satisfied demand for **MESD-PO** increases to 0.61% from 0.45% when BARON solver is used. In **MMRT-PO**, the average absolute $\%\Delta^*_{Sim}$ increases to 7.28% from 5.46%. The increase in average absolute $\%\Delta^*_{Sim}$ is more significant from 1.0% to 1.94% for **MECD-PO**, however, the number of solutions for which the differences are not statistically significant is twelve and it is similar to the result of fourteen under complete enumeration.

According to the results, it is seen that package solvers can be used to find good solutions for **MESD-PO** although there exists decreases in the performance of the

Table 5.17: Summary of $\%\Delta^*_{Sim}$ in corresponding measures over 36 instances

Statistics	MESD-PO	MMRT-PO	MECD-PO
Number of instances for which the differences are not statistically significant	7	3	12
Average $\%\Delta^*_{Sim}$	-0.32	6.67	-1.03
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	-0.40	7.28	-1.54
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	0.61	7.28	1.94

solutions in comparison to the solutions found by complete enumeration. The optimality gap is significantly high for **MMRT-PO** and **MECD-PO**. For those models, the computation time limit could be increased or other commercial nonlinear solvers could be used to improve the quality of solutions. As an alternative to package solvers, meta-heuristic algorithms could be used to find solutions for the proposed models as well.

In the next section, we use **MESD-PO**, **MECD-PO** and **MMRT-PO** on a real-life data set, checking the performance on problem instances with higher number of vehicles and demand regions.

5.3.7 Performance of the models on Edmonton Data

The data set of City of Edmonton Emergency Medical Services by Ingolfsson et al. (2003) from the previous chapter is used to test **MESD-PO**, **MECD-PO** and **MMRT-PO**, where there are 180 demand nodes with given demand rates and 16 possible vehicle locations with specified capacities ranging from one to three. The demand rates for the nodes vary from node to node. The mean of the travel time between stations and nodes are given in the data set. The mean of the incident handling time is set to 45 minutes for this experiment. The mean number of total demand calls per hour is set to 5 and 10. We located from 8 to 12 emergency medical vehicles in Edmonton City, resulting in ten problem instances in total.

For all instances, *KN*++ is used to select the best solution and it is compared to the optimal solutions of **MESD-PO**, **MECD-PO** and **MMRT-PO** based on complete enumeration. The comparison is made on a separate simulation run. After finding the best solutions with *KN*++ algorithm and the optimal solutions of the models based on complete enumeration for each instance, three performance measures (expected satisfied demand, expected covered demand and mean response time) are estimated by running the discrete event simulation model for 10 independent replications. A total of 550,000 demand calls is simulated for every solution at hand in the simulation model. The warm-up period is selected as 50,000 calls, and the performance measures are reported accordingly.

Based on the performance measures, $\%\Delta^*_{Sim}$ values are found in expected satisfied demand, expected covered demand and mean response time for the optimal solutions of **MESD-PO**, **MECD-PO** and **MMRT-PO**. $\%\Delta^*_{Sim}$ values in expected satisfied demand (compared to the best solution for P_E) are reported in Table 5.18 for three models.

Table 5.18: $\%\Delta^*_{Sim}$ in ED of the optimal solution of models

Total Demand Rate	Nb.of Vehc.	MESD-PO	MECD-PO	MMRT-PO
5	8	-0.63	-0.65	-0.45
	9	-0.42	-0.42	-0.05*
	10	-0.11	-0.09	-0.03*
	11	0.01*	-0.01*	0.05*
	12	0.02*	0.02*	0.06
10	8	-0.07*	-0.30*	-1.94
	9	-0.62	-0.58	-0.74
	10	-0.63	-0.86	-0.78
	11	0.22*	0.25*	-0.19*
	12	-0.02*	-0.05*	-0.20*

According to the results in Table 5.18, it is seen that the difference between optimal solution of models and the best solution reported by *KN*++ becomes insignificant as the number of vehicles increases. As the number of vehicles increases, all three models find solutions for which the difference with respect to the best solution is not statistically significant. Although this results is valid for both levels of the total demand rate, average $\%\Delta^*_{Sim}$ over for which the differences are statistically significant varies in total demand rate. In Table 5.19, the number of instances for which the differences are not statistically significant and the average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant are reported.

Table 5.19: Summary of $\%\Delta^*_{Sim}$ in ED

	MESD-PO		MECD-PO		MMRT-PO	
Total Demand Rate	5	10	5	10	5	10
Number of instances for which the differences are not statistically significant	2	3	2	3	3	2
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	-0.39	-0.62	-0.39	-0.72	-0.19	-1.15

According to results, it is seen that the average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant increases as the total demand rate increases from 5 to 10. In the meantime, number of instances for which the differences are not statistically significant increases from 2 to 3 for **MESD-PO** and **MECD-PO** while it decreases from 3 to 2 for **MMRT-PO**. Overall, $\%\Delta^*_{Sim}$ values are less than 1% in all instances but one instance for **MMRT-PO**, meaning that all models deliver solutions that are close to the best solution for P_E reported by *KN*++.

The performance of **MESD-PO** in minimizing the expected satisfied demand on Edmonton data is similar to its performance on toy data. The average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant on Edmonton data are -0.39% and -0.62% under total demand rate of 5 and 10, respectively, whereas it is -0.18% on toy data in Table 5.5.

Similar to the previous analysis, $\%\Delta^*_{Sim}$ values in expected covered demand (compared to the best solution for P_C) are reported in Table 5.20 for three models.

The results in Table 5.20 shows that $\%\Delta^*_{Sim}$ in expected covered demand from the best solution of P_C reported by *KN*++ usually decreases as the number of vehicles

Table 5.20: $\%\Delta^*_{Sim}$ in CD of the optimal solution of models

Total Demand Rate	Nb.of Vehc.	MESD-PO	MECD-PO	MMRT-PO
5	8	-3.38	-3.33	-1.23
	9	-4.40	-4.53	-0.23*
	10	-1.79	-1.72	0.02*
	11	-1.61	-1.65	0.47
	12	-1.86	-1.88	0.16*
10	8	0.95	0.73	-3.66
	9	-3.09	-3.08	-2.89
	10	-2.84	-3.22	-2.52
	11	0.07*	-0.03*	-1.53
	12	-0.16*	-0.29*	-0.60

increases given total demand rate. This could be attributed to the use of order of districting which asserts that only the closest three vehicles respond to the demand calls of a region. As the number of vehicles increases, the traffic intensity of the system decrease, hence the use of order districting is further justified.

In Table 5.21, the summary of $\%\Delta^*_{Sim}$ values in expected covered demand is given.

Table 5.21: Summary of $\%\Delta^*_{Sim}$ in CD

	MESD-PO		MECD-PO		MMRT-PO	
Total Demand Rate	5	10	5	10	5	10
Number of instances for which the differences are not statistically significant	-	2	-	2	3	-
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	-2.61	-1.66	-2.62	-1.86	-0.38	-2.24

According to Table 5.21, it is seen that **MESD-PO** and **MECD-PO** has similar $\%\Delta^*_{Sim}$ meaning that **MESD-PO** performs well in delivering good solution in expected covered demand objective although it is formulated to minimize expected satisfied demand. The performance of **MESD-PO** and **MECD-PO** in finding a solution that is close the best solution of P_C reported by *KN++* gets better as the total demand rate increases. This effect of increasing total demand rate could be attributed to overall increase in the busyness of vehicles. As the traffic intensity increases, the busy probabilities of vehicles would get closer to each other since a demand call is assigned a vehicle if there is one available at the time of the call, making the effect of the closeness of vehicle to the regions insignificant, in return making it easier to estimate the expected covered demand.

In summary, **MECD-PO** performs worse on Edmonton data than on toy data where the average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant in CD increases from -0.33% to -2.62 and -1.86 for total demand rates of 5 and 10, respectively.

Lastly, $\%\Delta^*_{Sim}$ values in mean response time (compared to the best solution for P_R) are reported in Table 5.22 for three models.

Table 5.22: $\%\Delta^*_{Sim}$ in R of the optimal solution of models

Total Demand Rate	Nb.of Vehc.	MESD-PO	MECD-PO	MMRT-PO
5	8	6.84	6.79	3.93
	9	7.96	7.95	1.40
	10	4.26	4.18	1.14
	11	4.04	3.93	-0.41
	12	4.30	4.11	0.15*
10	8	0.34	0.21	7.46
	9	3.62	3.43	4.17
	10	4.08	4.00	3.55
	11	-1.48	-1.44	0.99
	12	-0.13*	-0.23*	0.16*

The results in Table 5.22 shows that the performance of the models in mean response time gets better in the number of vehicles similar to the previous measures. Given the number of vehicles, the performance of **MESD-PO** and **MECD-PO** get better in the total demand rate, while **MMRT-PO** performs worse under total demand rate of 10. Differently from the results in Section 5.3.3, **MMRT-PO** performs better on Edmonton data than on toy data, both under total demand rates of 5 and 10. In the previous experiments, **MMRT-PO** model has an average $\%\Delta^*_{Sim}$ of 5.46% from the best solution reported by *KN*++ for P_R.

In Table 5.23, the summary of $\%\Delta^*_{Sim}$ values in expected covered demand is given.

Table 5.23: Summary of $\%\Delta^*_{Sim}$ in R

	MESD-PO		MECD-PO		MMRT-PO	
Total Demand Rate	5	10	5	10	5	10
Number of instances for which the differences are not statistically significant	-	1	-	1	1	1
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	5.48	1.64	5.39	1.55	1.52	4.04

With respect to the results in Table 5.23, it is shown that the performance of **MESD-PO** and **MECD-PO** gets better in finding close enough solution to the best solutions reported by *KN*++ as total demand rate increases similar to their performance in the expected covered demand measure. However, **MMRT-PO** performs worse.

As it is mentioned before, the busy probabilities of vehicles would get closer to each other as the traffic intensity increases since a demand call is assigned a vehicle if there is one available at the time of the call irrespective of its closeness in the exact system. This could make the estimation of expected covered demand easier. However, the composition of the probabilities that a demand region is served by a specific vehicle location under the exact system would change depending on the traffic intensity. This could negate the use of order districting since vehicles from farther locations would respond to the demand calls more. In addition, small differences in the busy probabilities of vehicles would be inflated when multiplied with mean travel times in the objective function in **MMRT-PO**, differently from the expected satisfied demand measure. However, **MMRT-PO** still performs better on Edmonton data than toy data.

Overall,we show that **MESD-PO**, **MECD-PO** performs relatively worse on Edmonton data than on toy data. The performance of **MMRT-PO** on Edmonton data is better than on toy data where the average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant are down to 1.52% and 4.04% from 5.46% for total demand rate of 5 and 10, respectively. With these results, we show that these models can be used on larger data sets as well.

In the next section, we use the modeling approach to estimate two more performance measures and construct two models from Chapter 3.

5.3.8 Performance of modeling approach under $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$

In addition to **MESD** and **MECD** and **MMRT** variants, we use the modeling approach to evaluate two different objective functions. For this purpose, models $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ are used from Chapter 3 where it is seen that these models improve equity in resulting optimal solutions in comparison to minimizing mean response time.

Remember that $\boldsymbol{P_4}$ minimizes the maximum mean region-wise response time while $\boldsymbol{P_7}$ minimizes the total positive deviation of mean region-wise response time from a threshold travel time τ.

$$
\begin{aligned}
(\boldsymbol{P_4}) \qquad \text{Minimize} \quad & \max_{j \in J}(R_j(\vec{x})) && (5.53)\\
\text{subject to:} \quad & \sum_{i \in I} x_i = N && (5.54)\\
& x_i \in \{0,1\}, \qquad \forall i \in I. && (5.55)
\end{aligned}
$$

$$
\begin{aligned}
(\boldsymbol{P_7}) \qquad \text{Minimize} \quad & \sum_{j \in J} [R_j(\vec{x}) - \tau]_+ && (5.56)\\
\text{subject to:} \quad & \sum_{i \in I} x_i = N && (5.57)\\
& x_i \in \{0,1\}, \qquad \forall i \in I, && (5.58)
\end{aligned}
$$

where $[\,.\,]_+ = \max\{0, .\}$.

$\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ use region-wise mean response time in the objective function which can be estimated based on the busy probabilities as it is mentioned in Section 5.1. Remember that region-wise mean response time, $R_j(\vec{x})$, under solution $(\vec{x})$ is found as follows:

$$R_j(\vec{x}) = \frac{(1-p_{v_{j1}})\omega_{jv_{j1}} + \sum_{t=2}^{N}(1-p_{v_{jt}})\omega_{jv_{jt}}\prod_{k=1}^{t-1}p_{v_{jk}}}{(1-p_{v_{j1}}) + \sum_{t=2}^{N}(1-p_{v_{jt}})\prod_{k=1}^{t-1}p_{v_{jk}}} \tag{5.59}$$

where $\omega_{jv_{jt}}$ is the mean travel time between region j and vehicle location v_{jt}.

Similar to the construction of **MMRT**, $\omega_{jv_{j1}}$ is replaced with $\sum_{i\in I} a_{ijN}\omega_{ij}$ in 5.59 and region-wise mean response time is defined as a function of the vector $\vec{s}$of decision variable s_{ijt} as follows:

$$R_j(\vec{s}) = \frac{(1-\sum_{i\in I}s_{ijN})\sum_{i\in I}a_{ijN}\omega_{ij} + \sum_{t=1}^{N-1}\left((1-\sum_{i\in I}s_{ijt})\sum_{i\in I}a_{ijt}\omega_{ij}\prod_{k=t+1}^{N}\sum_{i\in I}s_{ijk}\right)}{(1-\sum_{i\in I}s_{ijN}) + \sum_{t=1}^{N-1}\left((1-\sum_{i\in I}s_{ijt})\prod_{k=t+1}^{N}\sum_{i\in I}s_{ijk}\right)}$$

Then, $\boldsymbol{P_4 - PO}$ and $\boldsymbol{P_7 - PO}$ are constructed based on $R_j(\vec{s})$ with the constraints 5.13) - (5.23) and (5.30) - (5.34) of model **MESD-PO**.

$$\begin{aligned} (\boldsymbol{P_4 - PO}) \qquad & \text{Minimize} \qquad \max_{j\in J}(R_j(\vec{s})) \\ & \text{subject to:} \qquad (5.13)-(5.23), \\ & \qquad\qquad\qquad (5.30)-(5.34). \end{aligned} \tag{5.60}$$

$$\begin{aligned} (\boldsymbol{P_7 - PO}) \qquad & \text{Minimize} \qquad \sum_{j\in J}[R_j(\vec{s})-\tau]_+ \\ & \text{subject to:} \qquad (5.13)-(5.23), \\ & \qquad\qquad\qquad (5.30)-(5.34). \end{aligned} \tag{5.61}$$

where $[\,.\,]_+ = \max\{0, .\}$.

In the experiments, the optimal solutions for $\boldsymbol{P_4 - PO}$ and $\boldsymbol{P_7 - PO}$ under each problem instance are found by complete enumeration. The best solution for $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ are found by using *KN*++. After finding the optimal solutions and the best solutions, the performance measures for these solutions are estimated from a separate simulation run.

$\%\Delta^*_{Sim}$ are found for each problem instances by comparing the objective function values of the optimal solutions of $\boldsymbol{P_4 - PO}$ ($\boldsymbol{P_7 - PO}$) and the best solutions found for $\boldsymbol{P_4}$ ($\boldsymbol{P_7}$) by *KN*++. The summary statistics for $\%\Delta^*_{Sim}$ results are given in Table 5.24.

Table 5.24: Summary of $\%\Delta^*_{Sim}$ in the corresponding objective function values with respect to *KN*++ best solutions in 36 instances

Statistics	$P_4 - PO$	$P_7 - PO$
Number of instances for which the differences are not statistically significant	10	-
Average $\%\Delta^*_{Sim}$	5.12	176.41
Average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	6.90	176.41
Average absolute $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant	7.08	176.54

It is seen that proposed model $\boldsymbol{P_4 - PO}$ finds the same solution with *KN*++ best or a solution for which the difference from *KN*++ is not statistically significant in ten instances for $\boldsymbol{P_4}$. $\boldsymbol{P_7 - PO}$ cannot find a close enough solution in any instances. The average $\%\Delta^*_{Sim}$ over instances for which the differences are statistically significant are 6.90 % and 176.41 % for $\boldsymbol{P_4 - PO}$ and $\boldsymbol{P_7 - PO}$, respectively.

The performance of $\boldsymbol{P_4 - PO}$ is close to **MMRT-PO** where it has an average deviation of 5.46% over instances for which the differences are statistically significant, however it is still the worst performing model among **MESD-PO**, **MECD-PO** and **MMRT-PO**. The average $\%\Delta^*_{Sim}$ is very high for $\boldsymbol{P_7 - PO}$. The objective function of this model is the summation of the deviation of region-wise mean response times from a threshold which means the estimation errors in the region-wise mean response time measure are accumulated. Therefore, it is seen that this modeling approach may not be useful for this type of measure which is not a weighted average of other measures but a summation of them.

In addition to the performance of optimal solution inn comparison to simulation optimization, the effect of the modeling approach on the equity measures of the optimal solutions of $\boldsymbol{P_4 - PO}$ and $\boldsymbol{P_7 - PO}$ is also checked. The optimal solutions of $\boldsymbol{P_4 - PO}$ and $\boldsymbol{P_7 - PO}$ found by complete enumeration and the best solutions of $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ found by using *KN++* are compared to the best solution for $\boldsymbol{P_R}$ (where we minimize mean response time as in Chapter 3) in terms the mean response time, the variance of region-wise mean response time and the Gini coefficient.

In order to quantify the difference of the models, the mean absolute percent deviation of measures for the optimal solution of the models, $MAPD$, from $\boldsymbol{P_R}$ is calculated for each measure over all instances. We report the average percent positive deviations, $avgpd(\%)$, of the mean response time of the best solution in the models from $\boldsymbol{P_R}$. The percent of instances with positive deviation, $ppd(\%)$, in mean response time is reported in order to show the fraction of instances with positive deviations. For $VarR_j$ and G, the average percent negative deviation, $avgnd(\%)$, and the percent of instances with negative deviation, $pnd(\%)$,are reported instead since equity gets better as $VarR_j$ and G decrease.

In Table 5.25, these statistics are reported to show the changes in R, $VarR_j$ and G.

Table 5.25: Comparison of Models P_4 and P_7 with P_S in performance measures

Measure	Statistics	$P_4 - PO$	P_4	$P_7 - PO$	P_7
R	$MAPD(\%)$	22.00	14.74	4.58	11.51
	$ppd(\%)$	100	100	94.44	100
	$avgpd(\%)$	22.00	14.74	4.82	11.51
$VarRj$	$MAPD(\%)$	52.53	49.97	20.24	51.24
	$pnd(\%)$	97.22	100	63.89	97.22
	$avgnd(\%)$	53.92	49.97	26.19	51.80
G	$MAPD(\%)$	43.29	38.59	13.45	36.67
	$pnd(\%)$	100	100	75	97.22
	$avgnd(\%)$	43.29	38.59	16.76	37.48

The results show that $\boldsymbol{P_7 - PO}$ does not improve equity measures as good as P_7 in Section 3 with respect to P_R as expected since our modeling approach does not perform well for this objective function.

The effect of $\boldsymbol{P_4 - PO}$ on the mean response time, the variance of mean region-wise response time and the Gini coefficient are more prominent than in P_4. So, the solutions obtained $\boldsymbol{P_4 - PO}$ improve equity more while sacrificing from mean response time more in comparison to P_4 in Section 3. This is due to the fact that $\boldsymbol{P_4 - PO}$ cannot estimate the objective function well enough in the first place.

Overall, it seems that the modeling approach is not useful in objective functions which are constructed based on region-wise measures and the performance gets even worse when there is a summation of region-wise measures in the objective function. With this last experiment, we finish the experimental study and conclude this chapter in the next section.

5.4 Conclusion

In this chapter, a modeling approach is proposed which can be used to approximate the performance measures of the exact queueing model in closed-form. Then, these closed form formulations are used to construct mathematical models that can be solved with commercial package solvers.

Via an extensive experimental study, it is showed that proposed models **MESD-PO** and **MECD-PO** are well performing with respect to the measures used in their objective function, expected satisfied demand and expected covered demand. Although **MMRT-PO** is specifically formulated to minimize mean response time, it is seen that optimal solutions for **MESD-PO** and **MECD-PO** are better than the optimal solutions of **MMRT-PO** in resulting mean response time values.

Models **MESD-PO** and **MMRT-PO** are compared to the decomposition method DM-S-CF in two performance measures. It is seen that the performance of **MESD-PO** in expected satisfied demand matches the performance of the more advanced method DM-S-CF, but both **MESD-PO** and **MMRT-PO** fall behind DM-S-CF in minimizing mean response time of the system.

MECD model variants are compared to s-MEXCLP which is used to find the solution that covers the maximum demand. s-MEXCLP are provided with the average busy probability for each feasible solution of an instance which is estimated from a simulation study. It is seen that **MECD-PO** is better than s-MEXCLP, marginally in expected conditionally satisfied demand and significantly in mean response time of the optimal solutions.

MESD-PO, **MECD-PO** and **MMRT-PO** models are tested under package solvers. It is seen that the optimality gap for the solution are relatively low for **MESD-PO**, **MECD-PO** and the performance of the models are not significantly affected. **MMRT-PO** has a higher average optimality gap under BARON solver. Hence, it is suggested to increase computation time under BARON or use meta-heuristic algorithms to find better performing solutions for **MMRT-PO**.

MESD-PO, **MECD-PO** and **MMRT-PO** models are also tested on real-life data set, Edmonton data. It is seen that **MESD-PO**, **MECD-PO** performs relatively worse on Edmonton data than on toy data while **MMRT-PO** has lower deviations on Edmonton data. With this, the applicability of models on larger data sets are shown on real-life data set.

The modeling approach used to construct closed form formulations and mathematical models are also employed to estimate other performance measures. Models $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ from Chapter 3 are reconstructed in closed form as $\boldsymbol{P_4 - PO}$ and $\boldsymbol{P_4 - PO}$. The optimal solutions for those models are found via complete enumeration. A comparison with the best solutions obtained with *KN*++ for $\boldsymbol{P_4}$ and $\boldsymbol{P_7}$ shows that $\boldsymbol{P_7 - PO}$ has unacceptable deviations from the best solutions while $\boldsymbol{P_4 - PO}$delivers solutions with an average deviation of 6.90% over instances for which the differences are statistically significant.

To conclude, the proposed models compete with more sophisticated algorithmic approaches in finding an optimal solution for the problem instances. The models incorporate important measures of EMS systems and are easy to apply since performance measures of the EMS systems are expressed as a function of the decision variables in closed form. Package solvers could be easily used to find good solutions for the EMS vehicle location problems defined in this chapter that handle the uncertainty in demand, travel times or incident handling times based on queueing models.

CHAPTER 6

CONCLUSION

In this thesis, the emergency medical service (EMS) vehicle location problem is studied. The problem is defined based on stochastic processes addressing the uncertainty in demand calls, travel times and incident handling times. Incorporating uncertainties in the form of stochastic processes into the problem environment enables one to construct queuing models to assess the performance measures resulting from a candidate location solution for the EMS system. However, an exact queueing model requires a state definition where the size of the state space increases exponentially in the number of vehicles and demand regions. Hence, the exact queueing model is not used in the evaluation of the performance measure of the system in this thesis. A discrete event simulation model is constructed and used to evaluate the performance measures in the rest of the study.

In Chapter 3, we work on the equity aspect in the EMS vehicle location problem which is not extensively studied in the literature. Various location models are used to locate emergency vehicles at candidate locations and the measures related to equity such as the Gini coefficient and the variance of region-wise measures are checked. In an extensive computational study based on simulation optimization is conducted. The effect of network features and model choices on equity measures, and the trade-off between overall performance and equity are revealed.

It is shown that equity is overlooked when overall performance measures such as mean response time and expected satisfied demand are used in the objective function. Deterministic coverage constraints based on mean travel time decrease the disparities among regions. Region-wise performance measures improves equity at the most in the expense of overall measures. It is seen that network features affect equity differently under different models depending on the network. Therefore, it might be beneficial to incorporate the network specifications in the determination of the mathematical model to be used.

Simulation of real-life systems is computationally expensive and exhaustive as the gap between the simulation model and reality decreases. Another issue with using simulation is the issue of selecting the best solution from alternative solutions where simulation output analysis requires handling of randomness in the output variables. As the number of alternative solutions increases, an algorithmic approach is needed to select the best solution. In line with this requirement, we use a selection and ranking algorithm, $KN{+}{+}$, to select the best solution for a problem instance in Chapter 3. This algorithm still requires running the simulation models for every feasible solution for an initial number of demand calls to eliminate inferior alternatives. Although the algorithm continues eliminating inferior solutions until the termination, it still uses simulation models, and the computational burden persists.

Due to the reasons mentioned, a decomposition method with problem specific variants is proposed in Chapter 4. The evaluation of the performance measures of an EMS system relies on the queueing models constructed in the decomposition methods. EMS system is decomposed into separate interdependent queueing models, and performance measures of the EMS system are estimated based on the steady-state distributions of those interdependent models. Differently from the literature, the performance of the decomposition methods are checked on optimization problems. A meta-heuristic algorithm is proposed to obtain well performing solutions for the optimization problems by using the decomposition methods to assess the objective function values.

The proposed decomposition methods performs very well in optimization problems for several objective functions, regardless of the network features. The proposed meta-heuristic algorithm is easy-to-apply, gives well performing solutions, and the required computation time is significantly less than the time for simulation optimization. Hence, we deliver a good approximation algorithm for the performance measures of an EMS system based on queueing theory which can be used within meta-heuristic algorithms to find solution to optimization problems.

The decomposition methods proposed in Chapter 4 requires algorithmic solution approaches in approximating the steady-state distribution which is the solution to a nonlinear set of equations. Therefore, the decomposition methods are more computationally burdensome than closed-form mathematical models for which commercial solvers are easily exploited to solve. So, the interdependence among queueing models in Chapter 4 is ignored in Chapter 5 in order to develop approximate closed form formulations for the performance measures of the EMS system. Due to the simplifications in the problem environment, closed-form formulations for the performance measures are constructed based decision variables which are used to propose mixed integer nonlinear programs for the EMS vehicle location problems. Hence, the decision makers are provided with a mathematical model that could be solved with commercial solvers and incorporates uncertainties in the demand, travel time and incident handling time based on queueing theory.

With an extensive computational study, it is shown that MINLP models performs well under the objective functions, maximum expected satisfied demand and maximum expected covered demand, on both toy and real life data. The performance of the MINLP models varies under different objective functions, models having regional measures in the objective functions are particularly not performing well while the performance might be acceptable under minimum mean response time objective regarding the simplicity of the modeling approach. Overall, the proposed closed-form approximations can be used to maximize expected satisfied or covered demand in EMS vehicle location problems without any need for the estimation of problem parameters in advance as in MEXCLP.

With this thesis, we study EMS vehicle location problem in terms of equity in depth, deliver decomposition methods that can be used to estimate the performance measures of the EMS systems, propose MINLP models for the EMS vehicle location problems that can be solved with commercial solvers.

Bibliography

Akdogan, M.A., Bayındır, Z.P., Iyigun, C., 2018. Locating emergency vehicles with an approximate queuing model and a meta-heuristic solution approach. Transportation Research Part C: Emerging Technologies 90, 134–155. doi:10.1016/j.trc.2018.01.014.

Aly, A.A., White, J.A., 1978. Probabilistic formulation of the emergency service location problem. The Journal of the Operational Research Society 29, 1167–1179. doi:10.1057/jors.1978.261.

Aringhieri, R., Bruni, M.E., Khodaparasti, S., van Essen, J.T., 2017. Emergency medical services and beyond: Addressing new challenges through a wide literature review. Computers & Operations Research 78, 349–368. doi:10.1016/j.cor.2016.09.016.

Atkinson, Kovalenko, Kuznetsov, Mykhalevych, 2008. A hypercube queueing loss model with customer-dependent service rate. European Journal of Operational Research 191, 223–239.

Ball, M.O., Lin, F.L., 1993. A reliability model applied to emergency service vehicle location. Operations Research 41, 18–36. doi:10.1287/opre.41.1.18.

Bélanger, V., Ruiz, A., Soriano, P., 2019. Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. European Journal of Operational Research 272, 1–23. doi:10.1016/j.ejor.2018.02.055.

Berman, O., Drezner, Z., Krass, D., 2009. Cooperative cover location problems: the planar case. IIE Transactions 42, 232–246. doi:10.1080/07408170903394355.

Berman, O., Krass, D., 2002. The generalized maximal covering location problem. Computers & Operations Research 29, 563–581. doi:10.1016/S0305-0548(01)00079-X.

Bertsimas, D., Farias, V.F., Trichakis, N., 2011. The price of fairness. Operations Research 59, 17–31. doi:10.1287/opre.1100.0865.

Box, G.E.P., 1949. A general distribution theory for a class of likelihood criteria. Biometrika 36, 317–346.

Boyacı, B., Geroliminis, N., 2015. Approximation methods for large-scale spatial queueing systems. Transportation Research Part B: Methodological 74, 151–181. doi:10.1016/j.trb.2014.12.011.

Brandeau, M.L., Larson, R.C., 1986. Extending and applying the hypercube queueing model to deploy ambulances in boston. TIMS Studies in the Management Sciences 22, 121–153.

Budge, S., Ingolfsson, A., Erkut, E., 2009. Approximating vehicle dispatch probabilities for emergency service systems with location-specific service times and multiple units per location. Operations Research 57, 251–255. doi:10.1287/opre.1080.0591.

Chanta, S., Mayorga, M.E., Kurz, M.E., McLay, L.A., 2011. The minimum p-envy location problem: a new model for equitable distribution of emergency resources. IIE Transactions on Healthcare Systems Engineering 1, 101–115. doi:10.1080/19488300.2011.609522.

Chanta, S., Mayorga, M.E., McLay, L.A., 2014. Improving emergency service in rural areas: a bi-objective covering location model for ems systems. Annals of Operations Research 221, 133–159. doi:10.1007/s10479-011-0972-6.

Church, R., ReVelle, C.S., 1974. The maximal covering location problem. Papers of the Regional Science Association 32, 101–118. doi:10.1007/BF01942293.

Czyzyk, J., Mesnier, M.P., Moré, J.J., 1998. The neos server. IEEE Journal on Computational Science and Engineering 5, 68.

Daskin, M.S., 1983. A maximum expected location model: formulation, properties and heuristic solution. Transportation Science 7, 48–70. doi:10.1287/trsc.17.1.48.

Daskin, M.S., Stern, E.H., 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. Transportation Science 15, 137–152. doi:10.1287/trsc.15.2.137.

Dolan, E.D., 2001. The NEOS Server 4.0 Administrative Guide. Technical Memorandum ANL/MCS-TM-250. Mathematics and Computer Science Division, Argonne National Laboratory.

Drezner, T., Drezner, Z., Guyse, J., 2009. Equitable service by a facility: minimizing the gini coefficient. Computers & Operations Research 36, 3240–3246. doi:10.1016/j.cor.2009.02.019.

Drezner, Z., Wesolowsky, G.O., Drezner, T., 2004. The gradual covering problem. Naval Research Logistics 51, 841–855. doi:10.1002/nav.20030.

El Itani, B., Ben Abdelaziz, F., Masri, H., 2019. A bi-objective covering location problem. Management Decision 57, 432–444. doi:10.1108/MD-04-2018-0474.

Erkut, E., Ingolfsson, A., Erdoğan, G., 2008. Ambulance location for maximum survival. Naval Research Logistics 55, 42–58. doi:10.1002/nav.20267.

Galvão, R.D., Chiyoshi, F.Y., Morabito, R., 2005. Towards unified formulations and extensions of two classical probabilistic location models. Computers & Operations Research 32, 15–33. doi:10.1016/S0305-0548(03)00200-4.

Gendreau, M., Laporte, G., Semet, F., 1997. Solving an ambulance location model by tabu search. Location Science 5, 75–88. doi:110.1016/S0966-8349(97)00015-6.

Geroliminis, N., Karlaftis, M.G., Skabardonis, A., 2009. A spatial queueing model for the emergency vehicle districting and location problem. Transportation Research Part B 43, 798–811. doi:10.1016/j.trb.2009.01.006.

Geroliminis, N., Kepaptsoglou, K., Karlaftis, M.G., 2011. A hybrid hypercube – genetic algorithm approach for deploying many emergency response mobile units in an urban network. European Journal of Operational Research 210, 287–300. doi:10.1016/j.ejor.2010.08.031.

Gini, C., 1912. Variabilità e mutabilità. Studi economico-giuridici pubblicati per cura della Facoltà di Giurisprudenza della Regia Università di Cagliari Anno III, parte 2.

Goldsman, D., Kim, S., Marshall, W.S., Nelson, B.L., 2002. Ranking and selection for steady-state simulation: procedures and perspectives. INFORMS Journal on Computing 14, 2–19. doi:10.1287/ijoc.14.1.2.7710.

Gropp, W., Moré, J.J., 1997. Optimization environments and the neos server, in: Buhman, M.D., Iserles, A. (Eds.), Approximation Theory and Optimization. Cambridge University Press, p. 167.

Hakimi, S.L., 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. Operations Research 13, 462–475. doi:10.1287/opre.12.3.450.

Halpern, J., 1977. The accuracy of estimates of the performance criteria in certain emergency service queueing systems. Transportation Science 11, 223–241. doi:10.1287/trsc.11.3.223.

Hogan, K., ReVelle, C.S., 1986. Concepts and application of backup coverage. Management Science 34, 1434–1444. doi:10.1287/mnsc.32.11.1434.

Iannoni, A.P., Morabito, R., 2007. A multiple dispatch and partial backup hypercube queuing model to analyze emergency medical systems on higways. Transportation Research Part E: logistics and transportation review 43, 755–771. doi:10.1016/j.tre.2006.05.005.

Iannoni, A.P., Morabito, R., Saydam, C., 2008. A hypercube queueing model embedded into a genetic algorithm for ambulance deployment on highways. Annals of Operations Research 157, 207–224. doi:10.1007/s10479-007-0195-z.

Iannoni, A.P., Morabito, R., Saydam, C., 2009. An optimization approach for ambulance location and the districting of the response segments on highways. European Journal of Operational Research 195, 528–542. doi:10.1016/j.ejor.2008.02.003.

Iannoni, A.P., Morabito, R., Saydam, C., 2011. Optimizing large-scale emergency medical system operations on highways using the hypercube queueing model. Socio-Economic Planning Sciences 45, 105–117. doi:10.1016/j.seps.2010.11.001.

Ingolfsson, Erkut, Budge, 2003. Simulation of single start station for edmonton ems. Journal of the Operational Research Society 54, 736–746.

Jarvis, J.P., 1975. Optimization of Stochastic Service Systems with Distinguishable Server. Ph.D. thesis. Operations Research Center, IRP-TR-19-75, Massachusetts Institute of Technology. Cambridge.

Jarvis, J.P., 1985. Approximating the equilibrium behavior of multi-server loss systems. Management Science 31, 235–239. doi:110.1287/mnsc.31.2.235.

Karasakal, O., Karasakal, E.K., 2004. A maximal covering location model in the presence of partial coverage. Computers & Operations Research 31, 1515–1526. doi:10.1016/S0305-0548(03)00105-9.

Khodaparasti, S., Maleki, H.R., Bruni, M.E., Jahedi, S., Beraldi, P., Conforti, D., 2016. Balancing efficiency and equity in location-allocation models with an application to strategic ems design. Optimization Letters 10, 1053–1070. doi:10.1007/s11590-015-0929-9.

Kim, S., Nelson, B.L., 2006. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. Operations Research 54, 475–488. doi:10.1287/opre.1060.0281.

Korkmaz, S., Goksuluk, D., Zararsiz, G., 2014. Mvn: An r package for assessing multivariate normality. The R Journal 6, 151–162. doi:10.32614/RJ-2014-031.

Kulpa, 1997. The poincaré-miranda theorem. The American Mathematical Monthly 104, 545–550.

Larson, 1975. Approximating the performance of urban emergency service systems. Operations Research 23, 845–868.

Larson, R.C., 1974. A hypercube queueing model for facility location and redistricting in urban facility service. Computers & Operations Research 1, 67–95. doi:10.1016/0305-0548(74)90076-8.

Li, X., Zhao, Z., Zhu, X., Wyatt, T., 2011. Covering models and optimization techniques for emergency response facility location and planning: a review. Mathematical Methods of Operations Research 74, 281–310. doi:10.1007/s00186-011-0363-4.

Liu, Y., Li, Z., Liu, J., Patel, H., 2016. A double standard model for allocating limited emergency medical service vehicle resources ensuring service reliability. Transportation Research Part C: Emerging Technologies 69, 120–133. doi:10.1016/j.trc.2016.05.023.

Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530. doi:10.1093/biomet/57.3.519.

Marianov, V., ReVelle, C., 1994. The queuing probabilistic location set covering problem and some extensions. Socio-Economic Planning Sciences 28, 167–178. doi:10.1016/0038-0121(94)90003-5.

Marianov, V., ReVelle, C., 1996. The queueing maximal availability location problem: A model for the siting of emergency vehicles. European Journal of Operational Research 93, 110–120. doi:10.1016/0377-2217(95)00182-4.

Mclay, L.A., Mayorga, M.E., 2010. Evaluating emergency medical service performance measures. Health Care Management Science 13, 124–36. doi:10.1007/s10729-009-9115-x.

Mendonca, F.C., Morabito, R., 2001. Analysing emergency medical service ambulance deployment on a brazilian highway using the hypercube model. Journal of the Operational Research Society 52, 261–268. doi:10.1057/palgrave.jors.2601097.

NHS-England, 2015. Handbook to the nhs constitution for england. URL: https://www.gov.uk/government/publications/supplements-to-the-nhs-constitution-for-england/the-handbook-to-the-nhs-constitution-for-england.

Rawls, J., 1971. A Theory of Justice. Harvard University Press, Cambridge, MA. doi:10.2307/j.ctvkjb25m.

ReVelle, C.S., Hogan, K., 1988. A reliability constrained siting model with local estimates of busy fractions. Environment and Planning B: Urban Analytics and City Science 15, 143–152. doi:10.1068/b150143.

ReVelle, C.S., Hogan, K., 1989. The maximum availability location problem. Transportation Science 23, 192–200. doi:10.1287/trsc.23.3.192.

Sacks, Grief, 1994. Orlando police department uses or/ms methodology new software to design patrol district. OR/MS Today , 30–42.

Saydam, C., Aytuğ, H., 2003. Accurate estimation of expected coverage: revisited. Socio-Economic Planning Sciences 37, 69–80. doi:10.1016/S0038-0121(02)00004-6.

Sorensen, P., Church, R., 2010. Integrating expected coverage and local reliability for emergency medical services location problems. Socio-Economic Planning Sciences 44, 8–18. doi:10.1016/j.seps.2009.04.002.

Tabachnick, B.G., Fidell, L.S., 2013. Using multivariate statistics. 6 ed., Pearson, Boston, MA.

Takeda, R.A., Widmer, J.A., Morabito, R., 2007. Analysis of ambulance decentralization in an urban emergency medical service using hypercube queueing model. Computers & Operations Research 34, 727–741. doi:10.1016/j.cor.2005.03.022.

Toregas, C., Swain, R.W., ReVelle, C.S., Bergman, L., 1971. The location of emergency service facilities. Operations Research, 19, 1363–1373. doi:10.1287/opre.19.6.1363.

Toro-Díaz, H., Mayorga, M., Mclay, L., Rajagopalan, H.K., Saydam, C., 2015. Reducing disparities in large-scale emergency medical service systems. Journal of the Operational Research Society 66, 1169–1181. doi:10.1057/jors.2014.83.

Wang, W., Wu, S., Wang, S., Zhen, L., Qu, Z., 2021. Emergency facility location problems in logistics: Status and perspectives. Transportation Research Part E: Logistics and Transportation Review 154, 102465. doi:10.1016/j.tre.2021.102465.

Yitzhaki, S., Schechtman, E., 2013. The Gini Methodology: A Primer on a Statistical Methodology. Springer New York, New York, NY. chapter 8. pp. 11–31.

APPENDICES

Appendix A

DETAILS OF MANOVA ANALYSIS

Multivariate analysis of variance (MANOVA) is used to test the significance of effects of factors on multiple response variables. MANOVA requires the assumptions of analysis of variance (ANOVA) to be satisfied and some additional assumptions. These are the absence of multivariate outliers, absence of multicollinearity, multivariate normality, homogeneity of covariance matrices, and linear relationship between response variables in each treatment group.

To satisfy the absence of multicollinearity, response variables with a correlation higher than .90 are suggested to be excluded from the test by Tabachnick and Fidell (2013). Mardia's test is used to assess multivariate normality, which is proposed by Mardia (1970). Heterogeneity in covariance matrices is said to affect the test's significance minimally when equal sample sizes for each group are satisfied, which is also the case in our tests. Still, we use Box M test proposed by Box (1949) to test the homogeneity of covariances. The existence of a linear relationship between response variables in each group is checked visually with scatter plots.

The statistical test are conducted on R environment. We use `manova` function from `stats` package to apply MANOVA, `boxM` function for Box M test from `biotools` package, `mvn` function for Mardia's test from `MVN` package by Korkmaz et al. (2014) to assess multivariate normality , Shapiro-Wilk test from `MVN` package to assess univariate normality. As an example, the output of the MANOVA on R environment where *Model* factor has only two levels as $\boldsymbol{P_1}$ and $\boldsymbol{P_3}$ are given in Table A.1.

According to the results in Table A.1, it is seen that p-values ($Pr(> F)$) for all factors and interaction terms are smaller than 0.05. So, all the terms are significant. Therefore, we conclude that the difference between $\boldsymbol{P_1}$ and $\boldsymbol{P_3}$ is statistically

Table A.1: MANOVA results where *Model* has two levels as P_1 and P_3

Type II MANOVA Tests: Pillai test statistic	Df	test stat	approx F	num Df	den Df	Pr(>F)
Model	1	0.88966	1302	2	323	< 2.2e-16
Pattern	2	1.99632	87935	4	648	< 2.2e-16
Number of Vehicles	2	1.86597	2255	4	648	< 2.2e-16
Incident Handling Rate	1	0.99643	45118	2	323	< 2.2e-16
Model:Pattern	2	1.57626	603	4	648	< 2.2e-16
Model:Number of Vehicles	2	1.13732	214	4	648	< 2.2e-16
Pattern:Number of Vehicles	4	1.97014	5344	8	648	< 2.2e-16
Model:Incident Handling Rate	1	0.64860	298	2	323	< 2.2e-16
Pattern:Incident Handling Rate	2	1.92597	4215	4	648	< 2.2e-16
Number of Vehicles:Incident Handling Rate	2	1.01429	167	4	648	< 2.2e-16
Model:Pattern:Number of Vehicles	4	1.55479	283	8	648	< 2.2e-16
Model:Pattern:Incident Handling Rate	2	0.84775	119	4	648	< 2.2e-16
Model:Number of Vehicles:Incident Handling Rate	2	0.78150	104	4	648	< 2.2e-16
Pattern:Number of Vehicles:Incident Handling Rate	4	1.46324	221	8	648	< 2.2e-16
Model:Pattern:Number of Vehicles:Incident Handling Rate	4	0.57631	33	8	648	< 2.2e-16

significant, although they are found to be very similar by hierarchical agglomerative clustering.

Appendix B

DETAILED DOE RESULTS FOR THE GENETIC ALGORITHM

Table B.1: Percent of Time IQM-CF-E Optimum Found in a single GA run (%)

			Form			Nb. of Vehc.			Incident Hand. Rate			
S	u_c	u_m	Uni.	C-A.	O-A.	4	5	6	3	6	9	12
50	0.8	0.05	36	31	68	51	43	40	58	49	42	30
		0.1	50	43	79	68	56	48	69	66	41	53
	0.9	0.05	33	30	65	53	43	32	57	41	37	36
		0.1	52	44	73	61	62	47	62	61	50	52
100	0.8	0.05	61	57	93	78	75	57	82	70	60	68
		0.1	73	67	97	83	86	68	87	83	74	71
	0.9	0.05	66	55	91	79	69	63	77	74	64	67
		0.1	77	66	100	91	80	72	92	80	74	77

Table B.2: Percent of Time IQM-CF-E Optimum Found among 10 GA runs (%)

			Form			Nb. of Vehc.			Incident Hand. Rate			
S	u_c	u_m	Uni.	C-A.	O-A.	4	5	6	3	6	9	12
50	0.8	0.05	100	83	100	100	100	83	100	100	89	89
		0.1	100	100	100	100	100	100	100	100	100	100
	0.9	0.05	100	83	100	92	100	92	100	100	89	89
		0.1	100	92	100	100	100	92	100	100	89	100
100	0.8	0.05	100	100	100	100	100	100	100	100	100	100
		0.1	100	100	100	100	100	100	100	100	100	100
	0.9	0.05	100	100	100	100	100	100	100	100	100	100
		0.1	100	100	100	100	100	100	100	100	100	100

Table B.3: Percent of Time KN++ Optimum Found in a single GA run (%)

			Form			Nb. of Vehc.			Incident Hand. Rate			
S	u_c	u_m	Uni.	C-A.	O-A.	4	5	6	3	6	9	12
50	0.8	0.05	28	22	33	33	13	36	40	27	24	19
		0.1	35	28	39	37	23	42	40	33	26	37
	0.9	0.05	27	18	30	33	18	24	37	19	23	21
		0.1	39	28	34	37	23	42	40	34	27	34
100	0.8	0.05	46	28	40	43	21	49	42	33	36	40
		0.1	58	28	43	42	31	57	47	44	36	46
	0.9	0.05	51	29	41	43	25	53	52	37	34	38
		0.1	58	35	42	48	29	58	52	40	42	44

Table B.4: Percent of Time KN++ Optimum Found among 10 GA runs (%)

			Form			Nb. of Vehc.			Incident Hand. Rate			
S	u_c	u_m	Uni.	C-A.	O-A.	4	5	6	3	6	9	12
50	0.8	0.05	92	75	67	83	67	83	89	67	89	67
		0.1	92	75	83	75	75	100	78	89	78	89
	0.9	0.05	83	75	58	75	67	75	78	78	78	56
		0.1	92	75	67	67	67	100	89	67	78	78
100	0.8	0.05	83	67	50	67	50	83	67	67	67	67
		0.1	92	58	58	58	58	92	67	67	56	89
	0.9	0.05	92	75	58	67	67	92	89	67	67	78
		0.1	92	67	42	58	50	92	67	67	67	67

Appendix C

ON EXPECTED SATISFIED DEMAND AND MEAN RESPONSE TIME

Assume that there are M regions where a single vehicle is to be located with the objective of minimizing mean response time. Let λ_j be the demand rate of region $j = 1, .., M$ and, let the travel time between region i and j be exponentially distributed with mean ω_{ij}, and let the incident handling time for a region be exponentially distributed with known mean ϕ_j for region j.

The EMS system having a single vehicle located at region $i$could be represented as a queueing model as follows. $\{B_t, t \geq T\}$ is a continuous time Markov chain with state space L. A state B_t could be one of the following:

- j_k when the vehicle is responding to a call from region j and it is in the mode k of the service at time t where k is set to $1, 2$ or 3 if the vehicle is en-route to demand region j, handling the incident or en-route to the vehicle location, respectively,
- 0_i when the vehicle being available at time t,
- i_2 when the vehicle is serving a demand call from the region it is located at time t.

Then, the state space for the queueing model located at region i is $L = \{0_i, i_2, j_k\}$, $j = 1, ..., M, j \neq i$, $k = 1, 2, 3$.

An example transition diagram is shown in Figure C.1 where t, u, v are the regions and the vehicle is located at region t.

Let π_b denote the steady state probability of state $b \in L$, balance equations for the

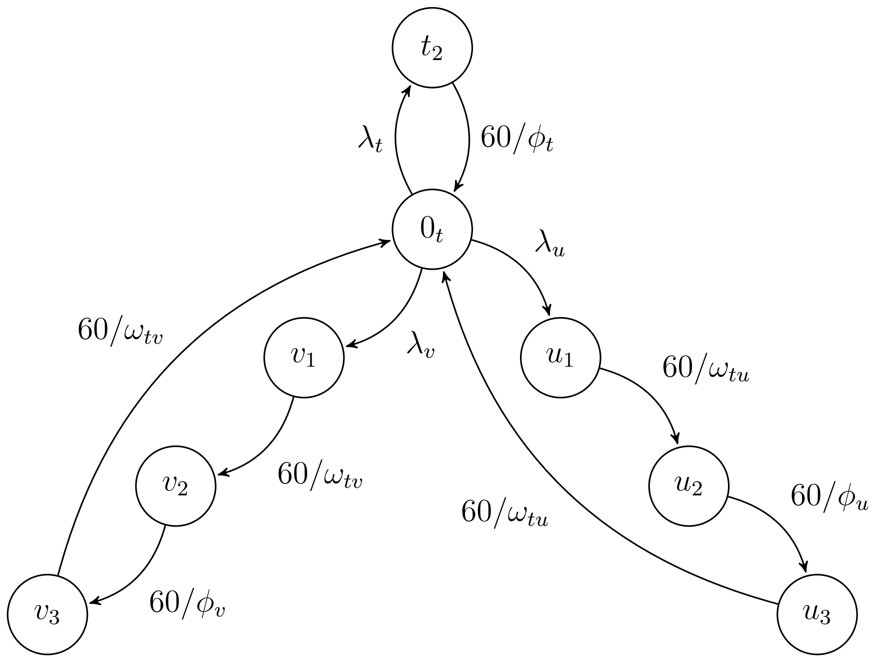


Figure C.1: Rate diagram for the queueing model in the example

general queueing model are written as follows:

$$\pi_{0_i}\sum_{j=1}^{M}\lambda_j = \pi_{i_2}\frac{60}{\phi_i} + \sum_{j=1, j\neq i}^{N}\pi_{j_3}\frac{60}{\omega_{ij}} \tag{C.1}$$

$$\pi_{0_i}\lambda_i = \pi_{i_2}\frac{60}{\phi_i} \tag{C.2}$$

$$\pi_{0_i}\lambda_j = \pi_{j_1}\frac{60}{\omega_{ij}} = \pi_{j_2}\frac{60}{\phi_i} = \pi_{j_3}\frac{60}{\omega_{ij}}, \quad j = 1, .., M, j \neq i \tag{C.3}$$

To find the steady state probabilities, (C.2), (C.3) and (C.4) are solved simultaneously.

$$\pi_{0_i} + \pi_{i_2} + \sum_{j=1, j\neq i}^{M}\sum_{k=1}^{3}\pi_{j_k} = 1. \tag{C.4}$$

Based on the definition of steady state probabilities in (C.2) and (C.3), we can rewrite (C.4) as follows:

$$\pi_{0_i} + \pi_{0_i}\frac{\lambda_i\phi_i}{60} + \sum_{j=1, j\neq i}^{M}\pi_{0_i}\frac{\lambda_j\omega_{ij}}{60} + \sum_{j=1, j\neq i}^{M}\pi_{0_i}\frac{\lambda_j\phi_j}{60} + \sum_{j=1, j\neq i}^{M}\pi_{0_i}\frac{\lambda_j\omega_{ji}}{60} = 1. \tag{C.5}$$

$$\pi_{0_i}\left(1 + \sum_{j=1}^{M}\frac{\lambda_j\phi_j}{60} + 2\sum_{j=1, j\neq i}^{M}\frac{\lambda_j\omega_{ij}}{60}\right) = 1. \tag{C.6}$$

Note that the second term inside the parenthesis in (C.6,) $\sum_{j=1}^{M} \frac{\lambda_j \phi_j}{60}$, is a constant and it is independent of the choice of vehicle location i. Let this term be equal to ρ. Then, π_{0_i} for the queueing model where the vehicle is located at region i is

$$\pi_{0_i} = \frac{1}{1 + \rho + 2\sum_{j=1, j \neq i}^{M} \frac{\lambda_j \omega_{ij}}{60}}. \tag{C.7}$$

The mean response time of the system where vehicle is located at region i is found as follows:

$$R(i) = \frac{\pi_{0_i} \lambda_i 0 + \pi_{0_i} \sum_{j=1, j \neq i}^{M} \lambda_j \omega_{ij}}{\pi_{0_i} \sum_{j=1}^{M} \lambda_j}. \tag{C.8}$$

where the numerator is the mean travel time weighted with the number assignments in unit time to regions and the denominator is the total number of assignments in unit time.

Simplifying C.8, the mean response time of the system is calculated as follows:

$$R(i) = \frac{\sum_{j=1, j \neq i}^{M} \lambda_j \omega_{ij}}{\sum_{j=1}^{M} \lambda_j}. \tag{C.9}$$

where the denominator, $\sum_{j=1}^{M} \lambda_j$, is a constant since it is independent form the choice of vehicle location i. Then, the optimal vehicle location which minimizes $R(i)$ is $\operatorname{argmin}_{i=1,\ldots,M} \left(\sum_{j=1, j \neq i}^{M} \lambda_j \omega_{ij} \right)$.

Note that π_{0_i} in (C.7) is maximized when $\sum_{j=1, j \neq i}^{M} \lambda_j \omega_{ij}$ is minimized since ρ is constant. Hence, minimization of mean response time maximizes π_{0_i} in the special case where one vehicle is located.

The expected satisfied demand in unit time is

$$ED(i) = \pi_{0_i} \sum_{j=1}^{M} \lambda_j \tag{C.10}$$

where the sum, $\sum_{j=1}^{M} \lambda_j$, is again a constant. Then, the solution that minimizes mean response time maximizes expected satisfied demand by maximizing π_{0_i}.

From the other end, one needs to maximize π_{0_i} to maximize expected satisfied demand. π_{0_i} is maximized when i is equal to $\operatorname{argmin}_{i=1,\ldots,M} \left(\sum_{j=1, j \neq i}^{M} \lambda_j \omega_{ij} \right)$. Then, the optimal solution i that maximizes expected satisfied demand also minimizes mean response time by minimizing $\sum_{j=1,j\neq i}^{M}\lambda_j\omega_{ij}$ (see (C.9)). This completes the proof for the special case where one vehicle is to be located.

For the case where two vehicles, the balance equations does not allow for an analytical proof where the state definition would require three entries as it is explained in the problem definition. In an effort to show a numerical example for the case of two vehicles, the modes of the service for a call is combined to a single mode where the time for service completion is assumed exponentially distributed with known mean $(\omega_{ij}+\phi_j+\omega_{ij})$. For this queueing model the state definition should store the state of the vehicle, being free or busy serving a region. Therefore, $B_t = (k,l)$ where $k,l = 0,1,...,M$ is the state where $(0,2)$ states that the first vehicle is available and second vehicle is busy serving demand Region 2 at time t, resulting in a total of $(M+1)^2$ states for the queueing model. Based on this definition, the balance equations of this queueing system is coded in MATLAB where two vehicles are located.

The problem instances from Section 4.4 are used for this experiment. The number of regions is set to 15, each considered a vehicle location. The demand rate is assumed to be the same for every region as 0.5 units per hour. The incident handling rate is set to 3, 6, 9 and 12. In total, 12 instances are generated based on a full factorial design using two factors: *Form* with three levels and *Incident Handling Rate* with four levels (3, 6, 9 and 12).

For each alternative solution of the problem instance, balance equations are solved with the help of MATLAB and mean response time (R) and expected satisfied demand in unit time (ED) is estimated based on the steady state probabilities. By complete enumeration, the optimal solution for each problem instance is found. It is seen that the optimal solution minimizing R and the optimal solution maximizing ED are the same solutions for all twelve instances. Although it is not possible to show analytically that two objectives are equivalent for greater number of vehicles than one in this problem environment, they are shown to be empirically equivalent for the two vehicle case as well by assessing the performance measures based on queueing theory.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Akdoğan, Muharrem Altan
Nationality: Turkish (TC)
Date and Place of Birth: 20.07.1989, Mersin
Marital Status: Married
Phone: +90 506 616 8360
E-mail: akdoganaltan@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
M.S.	METU	2015
B.S.	METU	2012
High School	Yusuf Kalkavan Anadolu Lisesi	2007

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2015-2022	METU	Research Asisstant
2013-2015	Yandex	Assessor, Search Quality Department
2012-2013	EnerjiSA	Functional Consultant, IS

PUBLICATIONS

Akdogan, M.A., Bayındır, Z.P., Iyigun, C. 2017. Location Analysis of Emergency Vehicles Using an Approximate Queueing Model. Transportation Research Procedia, 22, 430-439.

Akdogan M.A., Bayındır, Z.P., Iyigun, C. 2018. Locating Emergency Vehicles with an Approximate Queueing Model and a Meta-heuristic Solution Approach. Transportation Research Part C: Emerging Technologies, 90, 134–155.

Ertem, M.A., Akdogan, M.A., Kahya, M. 2022. Intermodal transportation in humanitarian logistics with an application to a Turkish network using retrospective analysis. International Journal of Disaster Risk Reduction, 72, 102828.