DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

A SIMULATION OPTIMIZATION STUDY FOR AN EMERGENCY DEPARTMENT OF A HOSPITAL IN İZMİR

by
Ceren GEGE

July, 2018
İZMİR

A SIMULATION OPTIMIZATION STUDY FOR AN EMERGENCY DEPARTMENT OF A HOSPITAL IN İZMİR

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science in Industrial Engineering, Industrial Engineering Program

by

Ceren GEGE

July, 2018

İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled and **"A SIMULATION OPTIMIZATION STUDY FOR AN EMERGENCY DEPARTMENT OF A HOSPITAL IN İZMİR"** completed by **CEREN GEGE** under supervision of **ASSIST. PROF. DR. ÖZGÜR YALÇINKAYA** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

....
Assist. Prof. Dr. Özgür YALÇINKAYA

Supervisor

....	....
Doç. Dr. Şener AKPINAR	ADALET ÖNER
	DR. ÖĞRETİM ÜYESİ
(Jury Member)	(Jury Member)

Prof. Dr. Latif SALUM
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Assist. Prof. Dr. Özgür YALÇINKAYA for the continuous support of my M.Sc. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor for my M.Sc. study.

I would like to thank the emergency department staff of the hospital for their help and guidance in making me have the opportunity to implement this study in the hospital.

I am grateful to my mother Türkan, my father Hüdayim and my brother Arif Cenker who have provided me moral and emotional support in my life. Their love and guidance are with me in whatever I pursue. I am also grateful to my other family members, especially my aunt Şafak, for being a great mentor and role model to me during my life. I would like to thank my fiancé and my friends who have supported me along the way.

Ceren GEGE

A SIMULATION OPTIMIZATION STUDY FOR AN EMERGENCY DEPARTMENT OF A HOSPITAL IN İZMİR

ABSTRACT

The long waiting times of patients and congestion are the most important problems of emergency departments (*EDs*) in the hospitals. In this thesis, for an *ED* of a hospital in İzmir, the aim is to find the optimal allocation of resources while minimizing the average overall flow time of patients (length of stay) and keeping the utilization ratios of doctors and nurses at the target levels. First of all, the simulation model of the *ED* is developed to evaluate the current system performance. Secondly, the data obtained from the simulation model is used in Nested Box-Behnken Design for building metamodels. Lastly, Derringer-Suich multi response optimization procedure is applied to minimize average overall flow time of patients and to keep utilization ratios of doctors and nurses at target levels. This study will provide help for healthcare decision makers to find the optimal configuration of resources in *ED* while considering to satisfy multi objectives simultaneously.

Keywords: Emergency department, healthcare systems, simulation, simulation optimization, response surface methodology, metamodel, desirability functions, multi response optimization, design of experiments

İZMİR’DE BİR HASTANENİN ACİL SERVİS DEPARTMANI İÇİN BİR SİMÜLASYON OPTİMİZASYONU ÇALIŞMASI

ÖZ

Hastaların uzun bekleme süreleri ve yoğunluk, hastanelerdeki acil servis departmanlarının en önemli problemleridir. Bu tezde, İzmir’deki bir hastanenin acil servis departmanı için, amaç, hastaların ortalama genel akış süresini (kalış süresi) minimize ederken ve doktor ve hemşirelerin faydalı kullanım oranlarını hedef seviyelerde tutarken, kaynakların eniyi tahsisini bulmaktır. İlk olarak, mevcut sistem performansını değerlendirmek için acil servis departmanının simülasyon modeli geliştirilmiştir. İkinci olarak, simülasyon modelinden elde edilen veriler Nested-Box-Behnken tasarımı içinde metamodeller oluşturmak için kullanıldı. Son olarak, Derringer-Suich çoklu yanıt optimizasyon prosedürü hastaların ortalama genel akış süresini en aza indirmek ve doktorların ve hemşirelerin faydalı kullanım oranlarını hedeflenen seviyelerde tutmak için uygulandı. Bu çalışma sağlık sektörü karar vericilerine çoklu hedefleri de eş zamanlı dikkate alarak, acil servislerde eniyi kaynak konfigürasyonunu bulmalarına yardım sağlayacaktır.

Anahtar Kelimeler: Acil servis departmanı, sağlık sistemleri, simülasyon, simülasyon optimizasyonu, yanıt yüzey metodolojisi, metamodel, istek fonksiyonları, çoklu yanıt optimizasyonu, deneyler tasarımı

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

In this chapter, the background, motivation and objectives of this study are stated, and also the organization of this thesis is outlined.

1.1 Background and Motivation

Nowadays healthcare system becomes more important and the medical levels of hospitals have greatly improved. However, waiting times and delays in hospitals and especially in emergency departments (*EDs*) still remain as a problem to be solved. When a waiting and delay occurs, this situation affects not only patients but also doctors and personnel. Because patient satisfaction is one of the most important aspects in healthcare systems and it tightly depends on waiting and delay times, it's really vital to analyze the key process to eliminate the bottlenecks, waiting times and optimize the process of healthcare system.

A hospital environment is a complex system that requires appropriate allocation of human and material resources in order to improve its effectiveness and efficiency. Therefore, finding the optimal allocation of resources in an *ED* is a problem studied mostly in the literature among the other problem types encountered in an *ED*.

1.2. Research Objectives

In this thesis, simulation optimization in an *ED* is studied with three main objectives;

- To review the relevant literature
- To develop a simulation model of a real *ED* system
- To do an optimization study considering multiple responses simultaneously

In order to meet the objectives;

- The studies on simulation optimization in *EDs* and also response surface methodology (*RSM*) studies in healthcare systems have been published in the last decade are reviewed. They are classified according to the problem type, aim, key performance index, modeling approach, software used and data gathering method and data sources.
- A simulation model is developed for evaluating a real *ED*'s current performance measures.
- Experiments are designed to obtained metamodels. Some special designs are used to obtain verified and validated metamodels. Derringer-Suich multi response optimization procedure is applied to minimize the average overall flow time of patients with keeping utilization ratios of the doctors and the nurses at target levels.

The contributions can be summarized as follows;

- During the literature review, we have found a few surveys investigating *ED* overcrowding, and recent *ED* simulation applications under normal and disaster conditions. Our literature review includes some studies only focused on simulation optimization in *EDs*. These studies in *EDs* mostly involve metaheuristics as optimization approaches. Additionally, we have reviewed studies including *RSM* in healthcare systems and this review shows us that there is a limited number of such studies in the literature.
- The *RSM* included studies focus on optimizing only single response. In this thesis multi-responses (three responses) are optimized simultaneously. To the best of our knowledge, this study is the first one that uses multi-response surface methodology to optimize a problem in an *ED* and the second one (the first one is Liang, Guo & Fung (2015) dealt with surgery scheduling in operation theaters) in all healthcare systems.
- In the design of experiments part, this study has a novel approach because Nested Box-Behnken Design (*NBBD*) is developed and applied to obtain second order metamodels again to the best of our knowledge it is the first time.

1.3 Organization of the Thesis

The organization of this thesis is as follows.

In chapter two, a literature review of the studies focused on simulation optimization studies in *EDs* and *RSM* in healthcare systems is given. Chapter three gives information about *RSM*. In chapter four, firstly the studied problem is introduced. Secondly the built simulation model of a real *ED* is explained then the application of optimization is given in details. Conclusions of the study and the optimal solution are given in the last chapter and the improvements are indicated elaborately.

CHAPTER TWO
RESPONSE SURFACE METHODOLOGY

2.1 Introduction to Response Surface Methodology

Response surface methodology (*RSM*) is a collection of mathematical and statistical techniques, which are useful for developing, improving, and optimizing processes. It also has important applications in the design, development, and formulation of new products, as well as in the improvement of existing product designs, and it is an effective tool for constructing optimization models (Myers & Montgomery, 2016, p.1).

RSM consists of the experimental strategy for exploring the space of the process or input factors, empirical statistical modelling to develop an appropriate approximating relationship between the yield and the process variables, and optimization methods for finding the levels or values of the process variables that produce desirable values of the response outputs (Myers & Montgomery, 1995, p.2).

"*RSM* was proposed by Box and Wilson in 1951 for finding the input combination that minimizes the output of a real, non-simulated system. Then, it was generally used for random simulation models" (Angün, Kleijnen, Hertog & Gürkan, 2002, p.377).

Box and Wilson (1951) laid the foundations for *RSM*. That paper was important not only because it described what became an entire field of research for the next fifty years but also because it changed dramatically the way that engineers, scientists, and statisticians approached industrial experimentation. They outlined a sequential philosophy of experimentation that encompasses experiments for screening, region seeking (such as steepest ascent), process/product characterization, and process/product optimization. Clearly, "*RSM* includes much more than second order model fitting and analysis. Indeed, *RSM*, broadly understood, has become the core of industrial experimentation" (Myers et al., 2004, p.53).

RSM is generally used for optimizing the performance (or a model) of an unknown system, which is subject to controllable, uncontrollable and unknown variables. *RSM* can be applied to any system that has key elements like a criterion of effectiveness known as the response of the system, which is measurable on continuous scale, and quantifiable independent variables that affect the system's performance. At these conditions, *RSM* is a group of techniques for finding the optimum response of the system in an optimum fashion. Although from a historical viewpoint *RSM* has been applied primarily to Operations Management, its greatest value to decision scientists is its potential application to simulation studies (Brightman, 1978, p.481).

There are two distinct phases in *RSM*. If the values of the decision variables are thought to produce a level of the system performance far below the maximum (minimum), the method of steepest ascent (descent) is employed to reach the optimum region quickly. The second phase employs canonical analysis to determine the exact values for decision variables that ensure optimum system performance (Brightman, 1978, p.482).

Graphical view of a response surface and its contour plot for a maxima problem with two independent variables is shown in Figure 2.1. In addition, current operating condition and maximum response point are denoted on the contour plot.

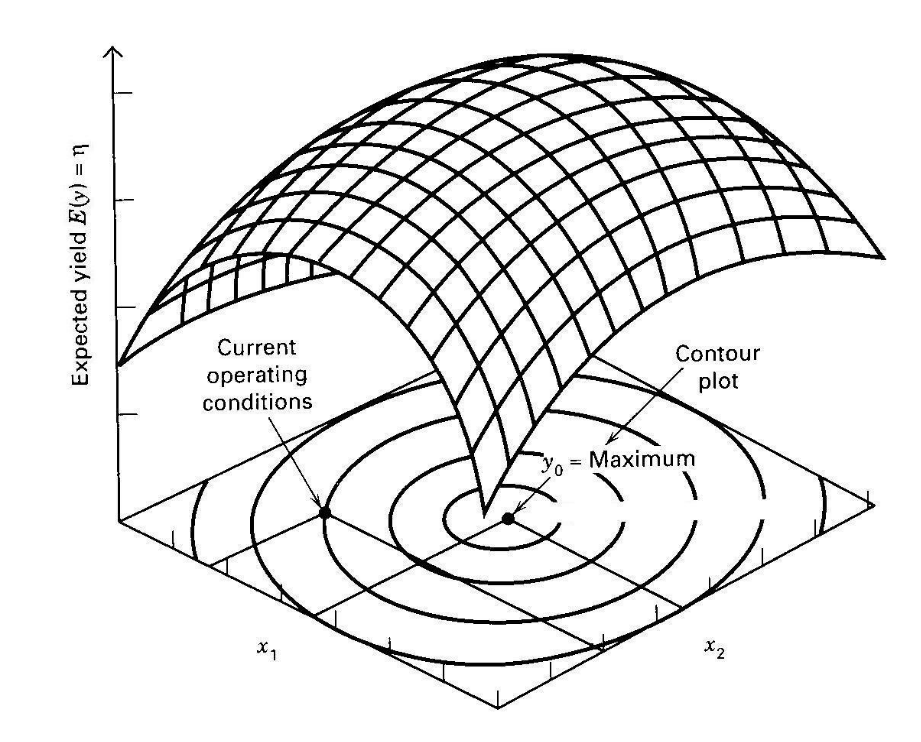


Figure 2.1 A three dimensional response surface

2.2 Approximating Response Functions

For example, if the objective is to find the levels of independent variables ($\phi_1, \phi_2, \ldots, \phi_k$) that maximize the response (y) of a specific process, the process should be identified as a function of independent variables, which is;

$$y = f(\phi_1, \phi_2, \ldots, \phi_k) + \varepsilon \tag{2.1}$$

where ε represents the noise or error that is observed in the response (y). Since the error term should be normally, identically and independently distributed with zero mean and constant variance (σ^2), the expected response is;

$$E(y) = E[f(\phi_1, \phi_2, \ldots, \phi_k)] + E(\varepsilon) = \eta \tag{2.2}$$

and then the surface represented by;

$$\eta = f(\phi_1, \phi_2, \ldots, \phi_k) \tag{2.3}$$

is called a response surface. The variables $\phi_1,\phi_2,\ldots,\phi_k$ in equation (2.3) are called natural variables, because they are expressed in the natural units of measurement. In much *RSM* works it is convenient to transform the natural variables to the coded variables $x_1,x_2,\ldots,x_k$ where these coded variables are usually defined to be dimensionless with mean zero and the same spread or standard deviation, and also fall between –1 and +1. The formulation, which is used for transforming natural variables to coded variables, is;

$$x_k = \frac{\phi_k - [\max(\phi_k) + \min(\phi_k)]/2}{[\max(\phi_k) - \min(\phi_k)]/2} \tag{2.4}$$

and the response surface in terms of coded variables is;

$$\eta = f(x_1,x_2,\ldots,x_k) \tag{2.5}$$

(Myers & Montgomery, 1995, p.3).

In most *RSM* problems, the form of the relationship between the independent variables and the response is unknown, it is approximated. Thus, the first step in *RSM* is to find an appropriate approximation for the true functional relationship between response and the set of independent variables. Usually, a low-order polynomial in some region of the independent variables is employed. If the response is well modelled by a linear function of the independent variables, then the approximating function is the first order model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \tag{2.6}$$

If there is curvature in the system, then a polynomial of higher degree must be used, such as the second order model;

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^{\,2} + \sum_{i}^{k-1} \sum_{j}^{k} \beta_{ij} x_i x_j + \varepsilon \tag{2.7}$$

where $i = 1,2,\ldots,k\text{-}1$ and $j = 1,2,\ldots,k$ also $i<j$.

Almost all the *RSM* problems utilize one or both of the approximating polynomials mentioned above. However, it is not the case that a polynomial model will be a reasonable approximation of the true functional relationship over the entire space of the independent variables. These models usually work quite well only for a relatively small region. The method of least squares is used to estimate the parameters in the approximating polynomials.

The response surface analysis is then performed using the fitted surface. If the fitted surface is an adequate approximation of the true response function, then the analysis of fitted surface will be approximately equivalent to analysis of the actual system. This type of model is called as metamodel. The model parameters can be estimated most effectively if proper experimental designs are used to collect the data. Designs for fitting response surfaces are called response surface designs. The response surface design used for fitting response surface directly affects the quality of the estimate (Montgomery, 2012, p.479).

The second order models are widely used in *RSM* because; it has a flexible structure that means it can take on a wide variety of functional forms, so it will often work well as an approximation to the true response surface, and estimating the regression parameters in the second order model is easy, also there is considerable practical experience indicating that second order models work well in solving real response surface problems (Myers & Montgomery, 2016, p.6).

2.2.1 The Metamodel Concept

Batmaz and Tunali (2003, p.455) say that “metamodelling is a process of developing a mathematical relationship between a response measure of interest and a set of input variables” (Yalçınkaya, 2004).

“Metamodel first described by Blanning (1975) as a mathematical relationship between one or more sets of sensitivity measures of interest and the sets of input to the metamodel” (Aytuğ, Doğan & Bezmez, 1996, p.24).

Let δ_j denote a factor j influencing the outputs of the real-word system (j = 1,2,…,s), and let W_c denote the system response vector (c = 1,2,…,w). Without loss of generality, the discussion can be simplified by considering a system with a single response W, since a multiple response system can be considered as a set of single response systems. The relationship between the response variable W and the inputs δ_j of the system is represented by;

$$W = f_1(\delta_1,\delta_2,\ldots,\delta_s) \tag{2.8}$$

A simulation model is then an abstraction of the real system, in which it is considered only a selected subset of the input variables $\{\delta_j \mid j = 1,2,\ldots,r\}$ where r is significantly smaller than the unknown s. The response of the simulation W' is then defined as a function f_2 of this subset and a vector of random numbers v representing the effect of the excluded inputs;

$$W' = f_2(\delta_1,\delta_2,\ldots,\delta_{r\,,v}) \tag{2.9}$$

A metamodel is a further abstraction, in which a subset of the simulation input variables are selected $\{\delta_j \mid j = 1,2,\ldots,m,\ m\leq r\}$ and the system described as;

$$W'' = f_3(\delta_1,\delta_2,\ldots,\delta_m) + \varepsilon \tag{2.10}$$

where ε denotes a fitting error, which has an expected value of zero. These levels of abstraction are shown in Figure 2.2 (Yu & Popplewell, 1994, p.788).

The simulation model is leaner than the real-word system, it contains fewer variables and these are all under the control of the experimenter. The simulation model, although simpler than the real-word system, is still a very complex way of relating input to output. A simpler analytical model may be used as an auxiliary to the simulation model in order to better understand the more complex model and to provide a framework for testing hypotheses about it. This auxiliary model is referred to as a metamodel (Friedman, 1995, p.15-20).

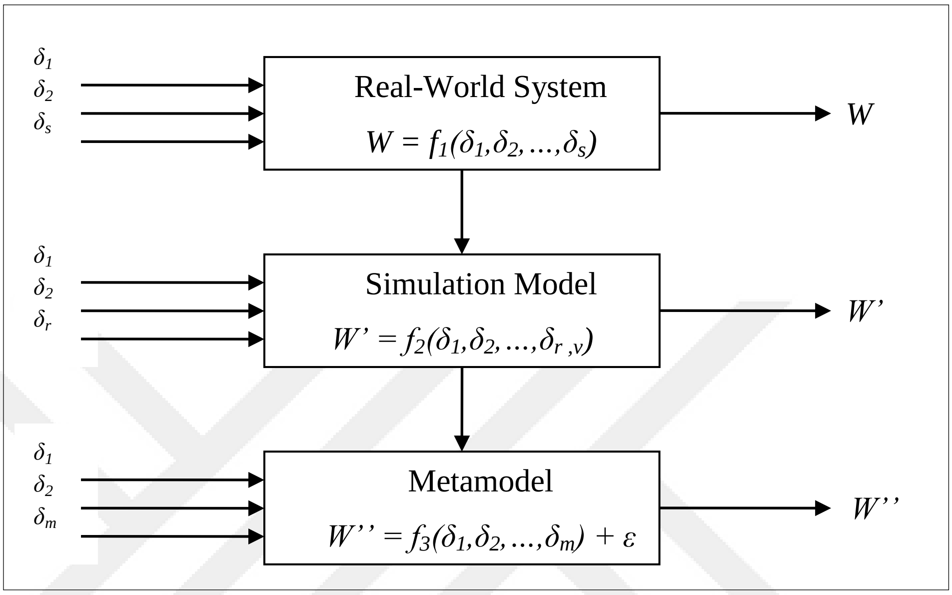


Figure 2.2 Metamodel concept

Metamodelling provides one approach to statistical summarization of simulation results, allowing some extrapolation from the simulated range of system conditions and therefore potentially offering some assistance in optimization. While simulation model is an abstraction of the real system, metamodel is an abstraction of simulation model; so metamodel is a further abstraction of the real system (Yu & Popplewell, 1994, p.778).

The objective of building a metamodel is to determine a (relatively simple) functional relationship between the system response and selected decision variables. Thus, it becomes much easier (cheaper) not only analyzing the simulation output, but also predicting how the real system will react to specify combinations of the set of controllable input variables. It is also straightforward to perform sensitivity analysis of the simulation model parameters and "what-if" questions without having to perform additional simulation runs (Santos & Nova, 1999, p.502).

The purpose of a metamodel is to estimate or approximate the response surface. A metamodel is often specified to be a regression model where the independent variables for the regression are the simulation input parameters and the dependent variable is the response of interest. Then the metamodel, instead of the actual simulation model, can be used economically to learn about how the response surface would behave over various regions of input-factor space and thus to estimate how the response would change at a particular point of the input factors are changed slightly or perhaps to find approximately optimal settings of the input factors (Sridharan & Babu, 1998, p.592).

Some of the benefits of the metamodel are;

- Simulation models are flexible and solve real problems without making too many restricting assumptions as in most analytical models. By using metamodels in post simulation analysis a good approximation to reality is provided.
- Through the use of experimental design the number of simulation runs required to generate a metamodel is drastically reduced. This also reduces the amount of time and effort that is spent in conducting simulation. Thus simulation becomes an attractive technique to use.
- The use of metamodels allows some generalizations to simulation output. However, these generalizations have to be within the defined boundaries of the problem. Sensitivity analysis on model parameters can also be easily carried out without re-running costly simulation programs.
- The model is valid and yields satisfactory solutions that are comparable to simulation results.
- Regression metamodels are usually simpler and easier to use than most analytical models (Madu, 1990, p.388).

2.3 The Followed Steps by a *RSM* Study

A typical *RSM* study consists of the following steps;

A) Estimation Process

a) The Studies Before Estimating the Metamodel

- Determining the objectives of developing a metamodel for the simulation model.
- Identifying all of the decision variables and their characteristics such as being discrete or continuous, and the performance measure(s).
- Identifying the operability regions of all the decision variables. The developed and validated simulation model that represents the behaviour of real system is used for this purpose.
- Identifying the accuracy of the metamodel with respect to the selected performance measure(s).
- Determining the validity measure(s) of the metamodel.
- Applying the factor screening. In this phase, the important decision variables (i.e., the decision variables which have statistically more effects on the performance measures) are determined.

b) The Studies for Estimating the Metamodel

- According to the factor screening results, build a 2^k full factorial or 2^{k-p} fractional factorial with centre points design for estimating the metamodel.
- Determining the low and the high levels for each decision variable. The operability region of the related decision variable should cover the low and high level of the related decision variable. The low and the high levels for the decision variables are coded as –1, +1 respectively.
- Making tactical decisions on executing the simulation model such as specifying the number of replications at each design point, the length of the replication, and the variance reduction technique which will be used.

- After executing the model at each design point, applying the least square estimation procedure for fitting the response to a first order model.
- Analyzing significance of the model parameters, main and interaction effects, lack-of-fit, and curvature effects. If the first order model is proper according to the results, the optimization process starts. If it is not proper, the first order model with interaction or second order model is fitted.

B) Optimization Process

- Determining the gradient vector of the performance measure, if the first order model is proper. This gradient vector is used in one of the process improvement techniques such as Steepest Ascent/Descent. In these methods, the centre of the design is changed to increase the performance of the system. If there is no improvement, the method stops for estimating a new first order model in new ranges, then, same procedure which was explained will be followed. This process continues until the new design has the interaction or curvature effects. If it is so, the interaction or second order model is fitted.
- After fitting second order model, applying the canonical analysis for calculating the optimal levels of the decision variables (Yıldız, 2003, p.41-43).

2.4 The Sequential Nature of *RSM*

Most applications of *RSM* are sequential in nature. That means, at first, some ideas are generated concerning which factors or variables are likely to be important in the response surface study. This usually leads to an experiment designed to investigate these factors with a view toward eliminating the unimportant ones. This type of experiment is usually called a screening experiment. The objective of factor screening is to reduce the list of candidate variables to a relatively few, so that subsequent experiments will be more efficient and require fewer runs or tests. A screening experiment is referred as *phase zero* of a response surface study.

After identifying important independent variables *phase one* of the response surface study begins. In phase one, the objective is to determine if the current levels or settings of the independent variables result in a value of the response that is near the optimum, or if the process is operating in some other region that is (possibly) remote from the optimum. If the current settings or levels of the independent variables are not consistent with optimum performance, then the experimenter must determine a set of adjustments to the process variables that will move the process toward the optimum. Phase one of the *RSM* makes considerable use of the first order model and an optimization technique called the method of steepest ascent/descent.

Phase two of a response surface study begins when the process is near the optimum. At this point the experimenter usually wants a model that will accurately approximate the true response function within a relatively small region around the optimum. Because the true response surface usually exhibits curvature near the optimum, a second order model (or perhaps some higher order polynomial) will be used. Once an appropriate approximated model has been obtained, this model may be analyzed to determine the optimum conditions for the process (Myers & Montgomery, 2016, p.7-8).

This sequential experimental process is usually performed within some region of the independent variable space called the operability region. This sequential nature is denoted in Figure 2.3.

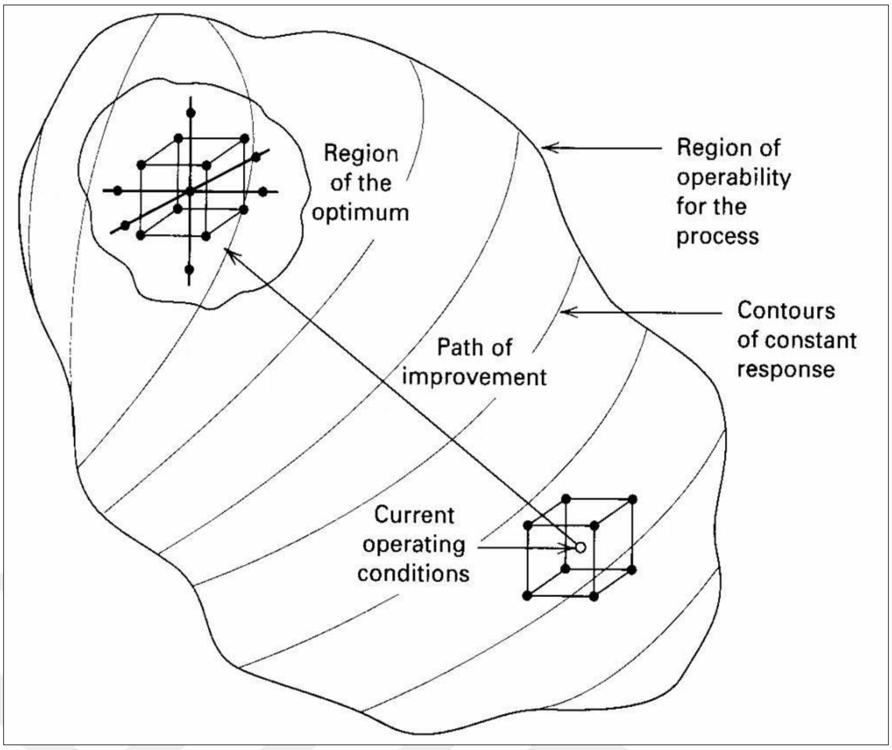


Figure 2.3 The sequential nature of RSM

2.5 The Method of Steepest Ascent/Descent

RSM is hill climbing (for a maxima objective), and the objective is locating the mountain summit. As in most mountain-climbing expeditions, the base camp is far below the summit. The objective of the method of steepest ascent is to move swiftly up the mountain without stopping for exploring the lower peaks (Brightman, 1978, p.482).

Frequently, the initial estimate of the optimum operating conditions for the system will be far from the actual optimum. At this situation, the objective is to move rapidly to the general vicinity of the optimum. The method of steepest ascent is a procedure for moving sequentially along the path of steepest ascent, that is, in the direction of the maximum increase in the response. If minimization is desired, then this technique is called as the method of steepest descent method.

As an example, the region of fitted first order response surface and the path of steepest ascent for a response with two independent variables (x_1 and x_2) is shown in Figure 2.4. The fitted first order model is as;

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i \tag{2.11}$$

and the contours of $\hat{y}$ is a series of parallel lines as denoted in Figure 2.4.

The direction of steepest ascent is the direction in which response increases most rapidly. This direction is parallel to the normal to the fitted first order response surface. The coordinates along the path of steepest ascent depend on the nature of the regression coefficients in the fitted first order model (Montgomery, 2012, p.481).

The method of steepest ascent contains the following steps;

- Fitting a first order model by using an orthogonal design. Two-level designs will be quite appropriate, although centre runs are recommended.

Consider a situation in which N experimental runs are conducted on k design variables x1, x2,…, xk and a single response y. A model is postulated of the types;

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i \text{ , } (i = 1,2,...,N) \tag{2.12}$$

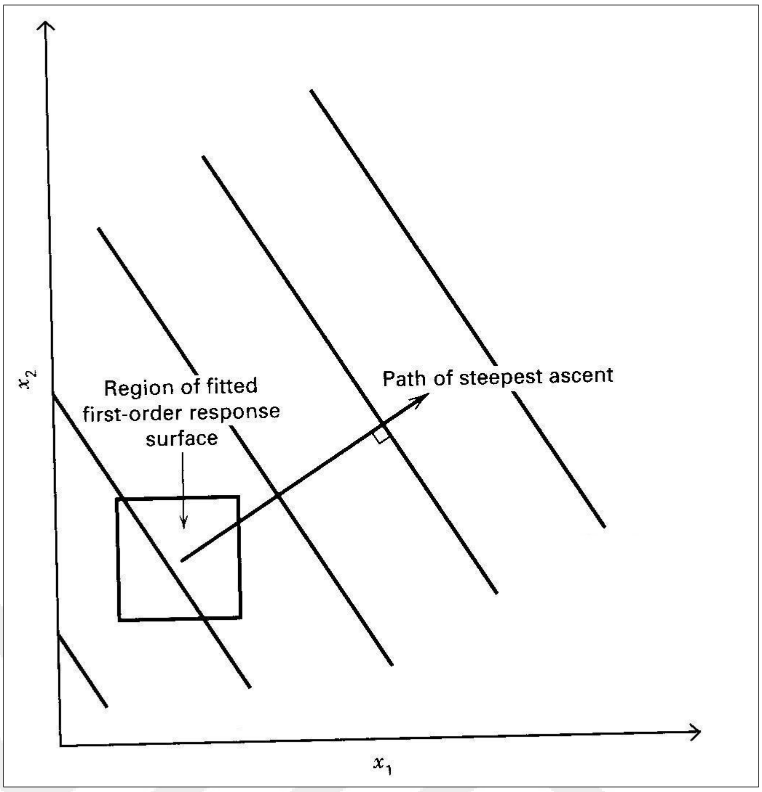


Figure 2.4 Path of steepest ascent

and the fitted model is given by;

$$y = b_0 + b_1 x_{i1} + b_2 x_{i2} + ... + b_k x_{ik} \tag{2.13}$$

where b_i are found by the method of least squares. For the first order model and a fixed sample N, if $x_j \in [-1,+1]$ for j = 1,2,…,k, then $Var(b_i / \sigma^2)$ for i = 1,2,…,k is minimized if the design is orthogonal and all x_i levels in the design are ± 1 for i = 1,2,…,k (Myers & Montgomery, 2016, p.376).

- Computing a path of steepest ascent (descent) if maximizing (minimizing) response is required.
- Conducting experimental runs along the path. That is, doing single or replicated runs and observing the response value. The results will normally show improving values of response. At some region along the path, the improvement will decline and eventually disappear. This stems from the deterioration of the

simple first-order model once one strays too far from the initial experimental region.

- Choosing a base at some location for a second experiment, where an approximation of the maximum (or minimum) response is located on the path. Again, the design should be a first order design. It is quite likely that centre runs for testing curvature, and degrees of freedom for interaction-type lack-of-fit are important at this point.
- Conducting a second experiment and fitting another first order model. Making a test for lack-of-fit. Computing a second path based on the new model, if lack-of-fit is not statistically significant. Conducting single or replicated experiments along this second path. It is quite likely that the improvement will not be as strong as that in the first path. After improvement is diminished, one has a base for conducting a more elaborate experiment and a more sophisticated process optimization (Myers & Montgomery, 2016, p.234).

A general algorithm for determining the coordinates of a point on the path of steepest ascent (with the assumption that the points x_i are the base or origin point, namely $x_1 = x_2 = , \ldots, = x_k = 0$) is as follows;

- Choosing a step size in one of the variables, say Δx_j. Usually, the variable that is known most about or has the largest absolute regression coefficient $|\hat{\beta}_j|$ is selected.
- The step sizes in the other variables are;

$$\Delta x_i = \frac{\hat{\beta}_i}{\hat{\beta}_j / \Delta x_j} \tag{2.14}$$

- Converting the Δx_i from coded variables to the natural variables (Montgomery, 2012, p.485-486).

As the experimental region moves near the region of optimum conditions, it is certainly expected that curvature would be more prevalent…. If a test for curvature finds significant quadratic terms, it would be suspected that the steepest ascent (descent) methodology would become effective (Myers & Montgomery, 2016, p.242). “At this point, the investigators will surely be interested in finding optimum conditions through the use of a fitted second order model” (Myers & Montgomery, 2016, p.244).

2.6 Analysis of Second Order Response Surface

2.6.1 Central Composite Designs (CCDs)

When the experimenter is relatively close to the optimum, a model which includes curvature is usually required to approximate the response. In most cases, the second order model is adequate.

Second order models cannot be fitted with two-level designs plus centre points. The minimum conditions to fit a second order model are, (a) at least $1+2k+k(k-1)/2$ distinct design points, where k is the number of design variables, and (b) at least three level of each design variable. In the case of first order designs the dominant property is orthogonality. In the case of second order designs, orthogonality ceases to be such an important issue; and estimation of individual coefficients, while still important, becomes secondary to the scaled prediction variance ($N\ Var\,\hat{y}(x)/\sigma^2$) (Myers & Montgomery, 2016, p.393).

For fitting a second order model there is a class of designs, central composite designs (*CCD*s) are common. *CCD*s are two-level full or fractional factorial designs that have been augmented with a small number of carefully chosen treatments to permit estimation of the second order response surface models (Neter, Kutner, Nachtsheim & Wasserman, 1996, p.1281).

Hood & Welch (1993, p.117) indicated *CCDs* are obtained from resolution V, full or fractional factorial designs by the adding of star points and perhaps more centre points. Star points are points where one of the x_i takes on the values $\pm\,\alpha$ while the remaining x_i are all zero (Yalçınkaya, 2004).

“Resolution V designs are designs in which no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with three-factor interactions” (Myers & Montgomery, 2016, p.165).

The total number of experimental trials planned, that is indicated by n_T is;

$$n_T = 2^{k-f} n_c + 2^k n_s + n_o \tag{2.15}$$

and the α that generates a rotatable design is given by;

$$\alpha = \left[\frac{2^{k-f}(n_c)}{n_s} \right]^{1/4} \tag{2.16}$$

where k is the number of factors, f is the level of fractional in the two-level factorial design selected, n_c is the number of replications at each design point, n_s is the number of replications at each star points, n_o is the number of replications at the centre point, and α is the axial distance.

“A rotatable design is one for which $N\,Var\,\hat{y}(x)/\sigma^2$ has the same value at any two locations that are the same distance from the design centre. In other words, $N\,Var\,\hat{y}(x)/\sigma^2$ is constant on spheres” (Myers & Montgomery, 2016, p.401).

While rotatability is a desirable property of a *CCD*, it should not be the sole basis for making the choice of α, in some situations it may be physically impossible or difficult to extent the star points beyond the experimental region defined by the upper and lower limits of each factor, where α must not exceed 1 (α = 1 is often called a face-centered design (*FCD*)). In this circumstance the resulting lack of rotatability may not be considered a serious disadvantage (Neter, Kutner, Nachtsheim & Wasserman, 1996, p.1283-1287).

The three components of *CCD* play important roles in building second order models;

- The resolution V fraction contributes in a major way in estimation of linear terms and two factor interactions. It is variance-optimal for these terms. The factorial points are the only points that contribute to the estimation of the interaction terms.
- The axial points contribute in a large way to estimation of quadratic terms, and do not contribute to the estimation of interaction terms.
- The centre runs provide an internal estimate of error (pure error) and contribute toward the estimation of quadratic terms (Myers & Montgomery, 2016, p.394).

The natural competitor for the *CCD* is three-level (3^k) factorial design. Actually, 3^2 design is a face centre cube, thus a *CCD*. But when k becomes large, the 3^k factorial design includes an excessive number of design points. For $k>3$ the number of design points for the 3^k design is usually considered impracticable for most applications, due to that *CCD* is the most popular class of second order designs (Myers & Montgomery, 2016, p.413).

Table 2.1 shows a three-variable face-centered design when the k=3 and the last rows represents the center point.

Table 2.1 A three-variable face-centered design

Runs	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1
9	-1	0	0
10	1	0	0
11	0	-1	0
12	0	1	0
13	0	0	-1
14	0	0	1
15	0	0	0

2.6.2 Nested Faced Center Design (NFCD)

As given in Section *2.6.1*, if $\alpha = 1$ the *CCD* is called as face-centered design (*FCD*).

Landman, Simpson, Mariani, Ortiz & Britcher (2007) investigated a wind-tunnel testing of high performance aircraft. In an exploratory study, they developed a hybrid design called a "nested face-centered design" (*NFCD*) when classic *CCDs* were found to have inadequate prediction qualities over a cuboidal design space with factors at five levels to allow for potential joining of individual model subspaces. Inspired by a "nested" type of approach of rotatable designs in three or four factors (Draper, 1960, 1962), they come up with the idea of nested *FCD*. To leverage some of the advantages of the fractional factorial and the *FCD* with a new approach allowing five levels, equal increments in the control surface levels are adopted. The full *NFCD* in two factors is shown in Table 2.4 and geographically shown in Figure 2.5. With more factors, the nested FCD geometry becomes hypercubes, in which case the inner cube and outer cube could be fractionated and therefore be alternate fractions (Yang, 2008). A two-factor *NFCD* in the study of Yang (2008) is shown in Table 2.2.

Table 2.2 A two-factor NFCD

Runs	x_1	x_2
1	-1	-1
2	1	-1
3	-1	1
4	1	1
5	0	-1
6	0	1
7	-1	0
8	1	0
9	-0.5	-0.5
10	0.5	-0.5
11	-0.5	0.5
12	0.5	0.5
13	0	-0.5
14	0	0.5
15	-0.5	0
16	0.5	0
17	0	0

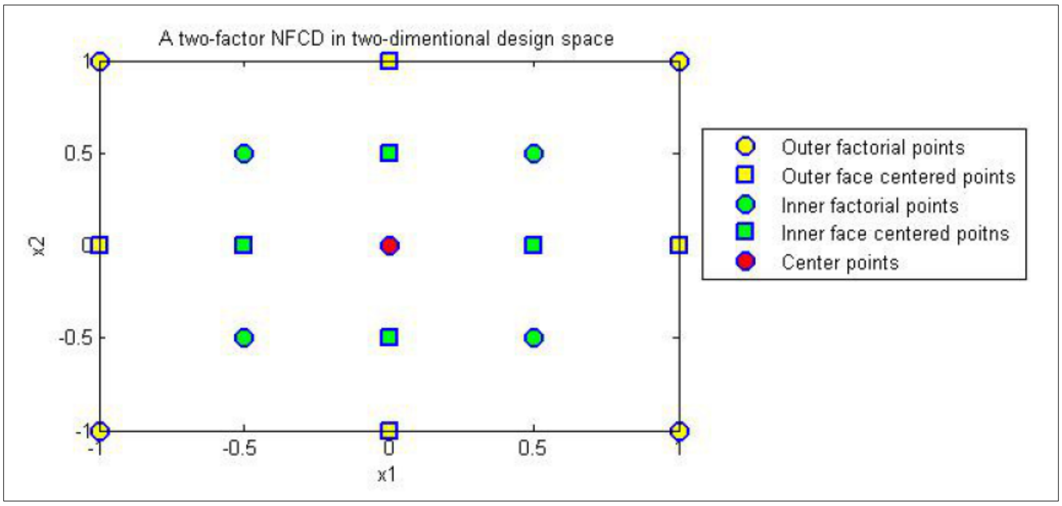


Figure 2.5 A two-factor NFCD in 2-D design space

2.6.3 Box-Behnken Design (BBD)

Box and Behnken (1960) developed a family of efficient three-level designs for fitting second order response surfaces. The methodology for design construction is interesting and quite creative. The class of designs is based on the construction of balanced incomplete block designs. For example, a balanced incomplete block design with three treatments and three blocks is given by Figure 2.6 (Myers & Montgomery, 2016, p.413).

	Treatment		
	1	2	3
Block 1	X	X	
Block 2	X		X
Block 3		X	X

Figure 2.6 A balanced incomplete block design

The pairing together of treatments 1 and 2 symbolically implies, in the response surface setting, that design variables x_1 and x_2 are paired together in a 2^2 factorial (scaling ±1) while x_3 remains fixed at the center (x_3=0). The same applies for blocks 2 and 3, with a 2^2 factorial being represented by each pair of treatments while the third factor remains fixed at 0 (Myers & Montgomery, 2016, p.413).

Table 2.3 shows a three-variable Box–Behnken Design (*BBD*) when the k=3 (Myers & Montgomery, 2016, p.413). The design is also shown geometrically in Figure 2.7 (Montgomery, 2012, p.504).

Table 2.3 A three-variable Box-Behnken Design

Runs	x_1	x_2	x_3
1	-1	-1	0
2	-1	1	0
3	1	-1	0
4	1	1	0
5	-1	0	-1
6	-1	0	1
7	1	0	-1
8	1	0	1
9	0	-1	-1
10	0	-1	1
11	0	1	-1
12	0	1	1
13	0	0	0

Each pair of factors is linked in a 2^2 factorial.

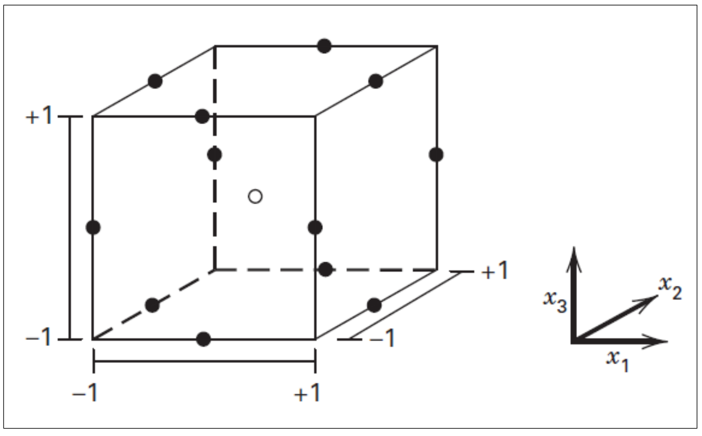


Figure 2.7 A Box-Behnken design for three factors

The *BBD* is a spherical design, with all points lying on a sphere of radius. Also, the *BBD* does not contain any points at the vertices of the cubic region created by the upper and lower limits for each variable. This could be advantageous when the points on the corners of the cube represent factor-level combinations that are prohibitively expensive or impossible to test because of physical process constraints (Montgomery, 2012, p.503-504).

The spherical nature of the *BBD*, combined with the fact that the designs are rotatable or near-rotatable, suggests that ample center runs should be used. In fact, for k=4 and 7, center runs are necessary to avoid singularity. The use of three to five center runs is recommended for the *BBD* (Myers & Montgomery, 2016, p.416).

"In many scientific studies that require *RSM*, researchers are inclined to require three evenly spaced levels. Thus, the Box–Behnken design is an efficient option and indeed an important alternative to the central composite design" (Myers & Montgomery, 2016, p.415).

2.6.4 Nested Box-Behnken Design (NBBD)

Inspired from the *NFCD*, a *NBBD* is developed as given in Table 2.4.

Table 2.4 A four-factor NBBD

Runs	x_1	x_2	x_3	x_4	Runs	x_1	x_2	x_3	x_4
1	-1	-1	0	0	26	0.5	-0.5	0	0
2	1	-1	0	0	27	-0.5	0.5	0	0
3	-1	1	0	0	28	0.5	0.5	0	0
4	1	1	0	0	29	0	0	-0.5	-0.5
5	0	0	-1	-1	30	0	0	0.5	-0.5
6	0	0	1	-1	31	0	0	-0.5	0.5
7	0	0	-1	1	32	0	0	0.5	0.5
8	0	0	1	1	33	-0.5	0	0	-0.5
9	-1	0	0	-1	34	0.5	0	0	-0.5
10	1	0	0	-1	35	-0.5	0	0	0.5
11	-1	0	0	1	36	0.5	0	0	0.5
12	1	0	0	1	37	0	-0.5	-0.5	0
13	0	-1	-1	0	38	0	0.5	-0.5	0
14	0	1	-1	0	39	0	-0.5	0.5	0
15	0	-1	1	0	40	0	0.5	0.5	0
16	0	1	1	0	41	-0.5	0	-0.5	0
17	-1	0	-1	0	42	0.5	0	-0.5	0
18	1	0	-1	0	43	-0.5	0	0.5	0
19	-1	0	1	0	44	0.5	0	0.5	0
20	1	0	1	0	45	0	-0.5	0	-0.5
21	0	-1	0	-1	46	0	0.5	0	-0.5
22	0	1	0	-1	47	0	-0.5	0	0.5
23	0	-1	0	1	48	0	0.5	0	0.5
24	0	1	0	1	49	0	0	0	0
25	-0.5	-0.5	0	0					

2.6.5 Location of the Stationary Point

The aim is to find the levels of $x_1, x_2, \ldots, x_k$ that optimize the predicted response. This point, if it exists will be the set of $x_1, x_2, \ldots, x_k$ for which partial derivates are zero, namely;

$$\partial\hat{y}/\partial x_1 = \partial\hat{y}/\partial x_2 = \ldots = \partial\hat{y}/\partial x_k = 0 \qquad (2.17)$$

This point, say $x_{1,s}$, $x_{2,s}$, ..., $x_{k,s}$ is called the stationary point (Montgomery, 2012, p.486).

The fitted second order model in matrix notation is;

$$\hat{y} = b_0 + x'b + x'\hat{B}x \tag{2.18}$$

where b_0 contains estimate of the intercept, b contains estimate of the linear, and $\hat{B}$ contains estimate of the second order coefficients. Actually;

$$x' = [x_1, x_2, \ldots, x_k] \tag{2.19}$$

$$b' = [b_1, b_2, \ldots, b_k] \tag{2.20}$$

$$\hat{B} = \begin{bmatrix} b_{11} & b_{12}/2 & \ldots & b_{1k}/2 \\ & b_{22} & \ldots & b_{2k}/2 \\ & & \ldots & \ldots \\ sym. & & & b_{kk} \end{bmatrix} \tag{2.21}$$

If $\hat{y}$ is differentiated respect to x,

$$\partial\hat{y}/\partial x = b + 2\hat{B}x \tag{2.22}$$

and this derivative is set to 0, the stationary point will be found as;

$$X_s = -B^{-1}b/2 \tag{2.23}$$

X_s is the stationary point of the system. The predicted response at the stationary point can be found by substituting equation (2.23) in to equation (2.18), it is;

$$\hat{y}_s = b_0 + x_s'b/2 \tag{2.24}$$

(Myers & Montgomery, 2016, p.278).

After finding stationary point, we determine whether the stationary point is a point of maximum response (Figure 2.8) or minimum response (Figure 2.9) or a saddle point (Figure 2.10). The most straightforward way to do this is to examine a contour plot of the fitted model. If there are only two or three process variables, the construction and interpretation of this contour plot is relatively easy.

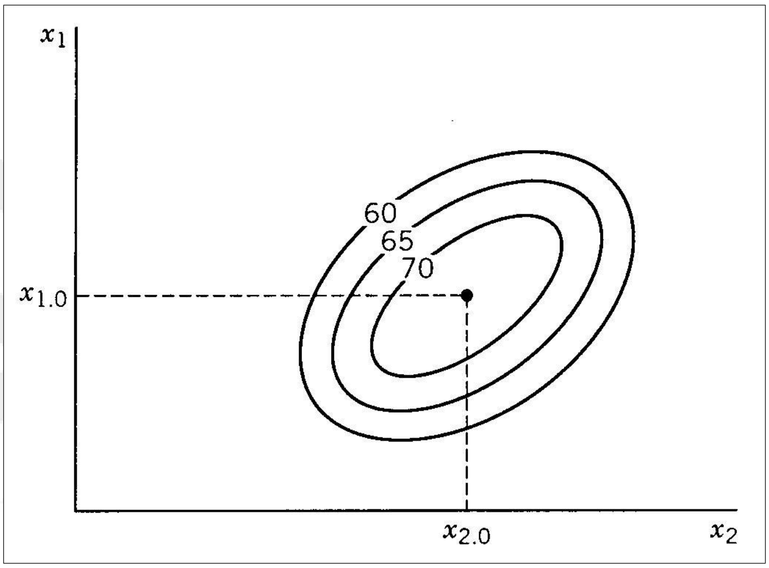


Figure 2.8 Stationary point is a point of maximum response

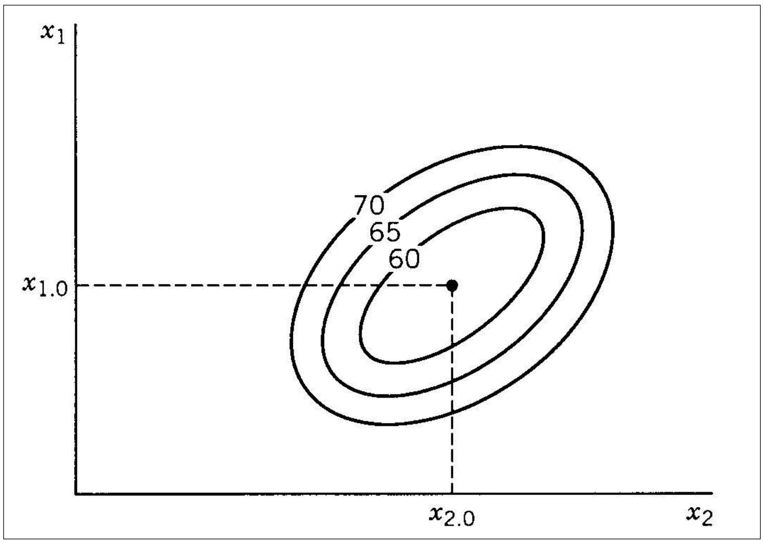


Figure 2.9 Stationary point is a point of minimum response

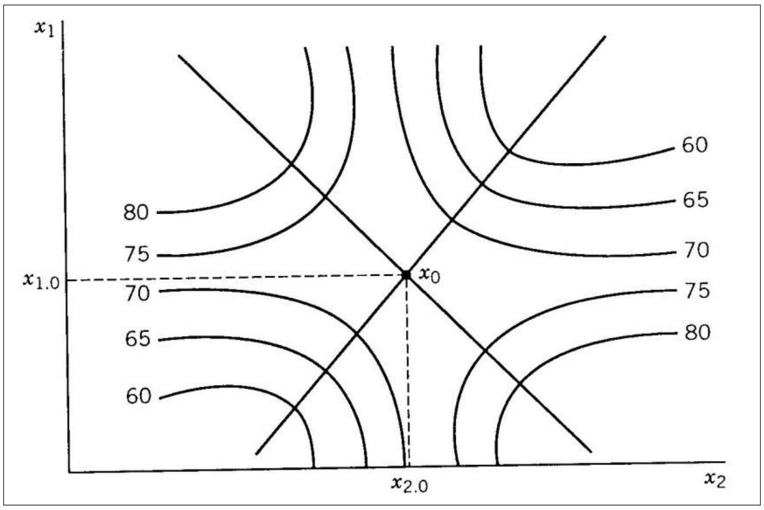


Figure 2.10 Stationary point is a saddle point

However, even when there are relatively few variables, a more formal analysis, called the canonical analysis, can be useful. It is helpful first to transform the model into a new coordinate system with the origin at the stationary point X_s and then to rotate the axes of this system until they are parallel to the principal axes of the fitted response surface. Results in the fitted model that is called canonical form of the model is;

$$\hat{y} = \hat{y}_s + \sum_{i=1}^{k} \lambda_i w_i^2 \tag{2.25}$$

where $\hat{y}_s$ is the estimated response at the stationary point, λ_i are the eigenvalues of $\hat{B}$ and w_i are canonical variables (transformed independent variables). Canonical form of the second-order model with two independent variables is denoted in Figure 2.11 (Montgomery, 2012, p.489).

Equation (2.25) nicely describes the nature of the stationary point and the nature of the system around stationary point. The nature of X_s (stationary point) is determined by the sings of the λ's;

- If λ_i are all negative, the stationary point is a point of maximum response.
- If λ_i are all positive, the stationary point is a point of minimum response.

- If λ_i are mixed in sign, the stationary point is a saddle point (Myers & Montgomery, 2016, p.279).

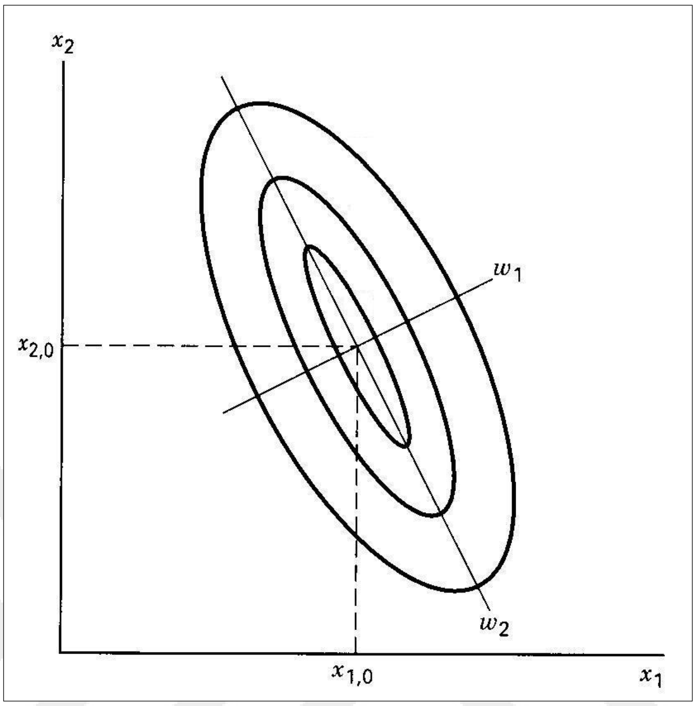


Figure 2.11 Canonical form of second order model

It is not unusual to encounter variations of the pure maximum, minimum, or saddle point response surfaces, that system is called ridge system. The two type of ridge system are the stationary ridge system and the rising ridge system.

If the stationary point is in the region of the experimental design and one (or more) λ_i is near zero, the system is a stationary ridge system. As an example in Figure 2.12, a contour plot of a stationary ridge system with two independent variables is shown, and maximization is the objective of response. For this example λ_1<0 and λ_2<0 but $\lambda_1 \cong 0$, the canonical model for this response surface is theoretically;

$$\hat{y} = \hat{y}_s + \lambda_2 w_2^{\;2} \tag{2.26}$$

Clearly the stationary point is a point of maximum response, but there is essentially a line maximum. The response variable is very insensitive to the variable w_1 multiplied by the small λ_1. Optimum may be taken anywhere along w_1 axis.

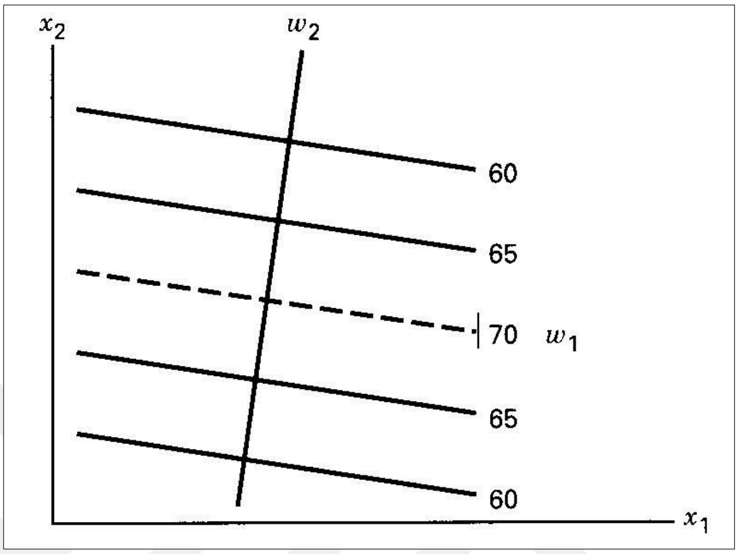


Figure 2.12 A contour plot of a stationary ridge system

If the stationary point is far outside the region of exploration for fitting the second order model and one (or more) λ_i is near zero, then the surface may be a rising ridge. In this type of ridge system, inferences about the true surface or the stationary point cannot be drawn because X_s is outside the region where the model fitted.

As an example in Figure 2.13, a contour plot of a rising ridge system with two independent variables is demonstrated, and again maximization is the objective of response.

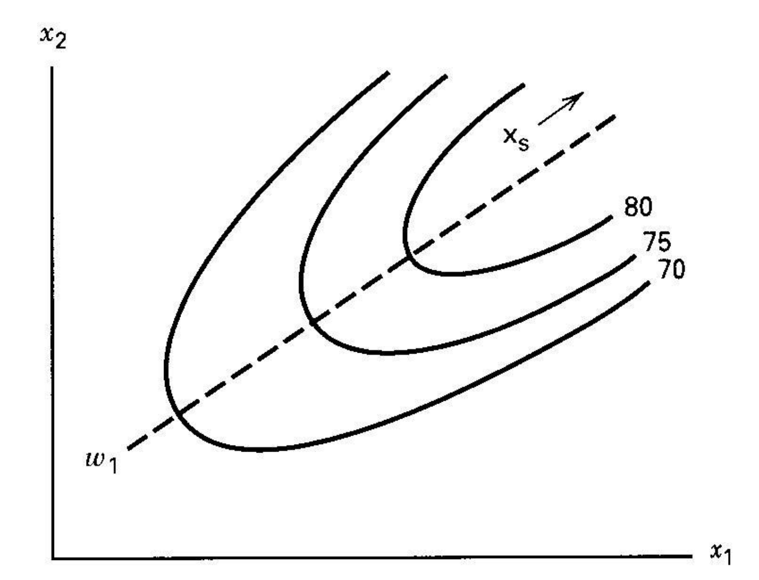


Figure 2.13 A contour plot of a rising ridge system

Again, λ_1<0 and λ_2<0 but $\lambda_1 \cong 0$, the canonical model for this response surface is the same as in equation (2.26). Also it is seen that, further exploration is warranted in the w_1 direction. Indeed the rising ridge often is a signal to the researcher that he/she has perhaps made a faulty or premature selection of the experimental design region (Myers & Montgomery, 2016, p.283) and (Montgomery, 2012, p. 495).

2.7 Response Surface Analysis with Multiple Responses

Up to now, it is focused on modelling a measured response or a function of design variables and letting the analysis indicate areas in the design region where the process is likely to give desirable results, the term "desirable" being a function of the predicted response. However, in many instances the term desirable is a function of more than one response (Myers & Montgomery, 2016, p.325).

Multiple response optimization is a method that allows for compromise among the various responses. Simultaneous consideration of multiple responses involves first building an appropriate response surface model for each response and then trying to find a set of operating conditions that in some sense optimizes all responses or at least keeps them in desired ranges (Myers & Montgomery, 2016, p.333).

A relatively straightforward approach to optimizing several responses that works well when there are only a few process variables is to overlay the contour plots for each response. The experimenter can visually examine the contour plot to determine appropriate operating conditions (Montgomery, 2012, p.496-497).

2.7.1 The Desirability Function

The overlaying of contour plots along with separate response surface analyses often give the user workable solutions for product improvement as long as the number of responses is not too great. However, when the problem involves four or more responses or design variables, then contour overlay methodology becomes unruly. Derringer & Suich (1980) developed an interesting procedure, which can be very useful when several responses are involved. The method makes use of a desirability function in which the researchers own priorities and desires on the response values are built into one optimization procedure (Myers & Montgomery, 1995, p.247-248).

Ribardo & Allen (2003, p.227) indicated that engineers design products or processes by selecting x_1, x_2, ..., x_p that will result in a desirable combination of properties or quality criteria, $Y_1, Y_2, \dots, Y_m$. Functions that transform a set of properties into a single objective are called “desirability” functions and are written as $D(Y_1, Y_2, \dots, Y_m)$ (Yalçınkaya, 2004).

The desirability function approach is one of the most widely used methods in industry for the optimization of multiple response processes. It is based on the idea that the “quality” of a product or process that has multiple quality characteristics, with one of them outside of some “desired” limits, is completely unacceptable. The method finds operating conditions that provide the “most desirable” response values. The desirability approach is a popular method that assigns a “score” to set of responses and chooses factor settings that maximize that score (NIST/SEMATECH, 2003).

Castillo, Montgomery & McCarvile (1996) say that “the basic idea of the desirability function approach is to transform a multiple response problem in to a single response problem by means of mathematical transformations” (Yalçınkaya, 2004).

First, Harrington (1965) introduced the concept of a “desirability function” for determining input parameter settings, which optimize the tradeoffs among multiple process measurements. Then, the procedure refined by Derringer & Suich (1980). In addition, a recent article by Castillo et al. (1996) suggested a slight modification to the procedure so that it can be easily implemented using the “Solver” function of Microsoft Excel (Fuller & Scherer, 1998, p.4016).

In Harrington (1965) exponential functional forms were selected to calculate the desirabilities associated with individual criteria, Y_i, and the use of the geometric mean for weighting these criteria together to calculate overall desirability. Derringer & Suich (1980) criticized the functional forms and weighting scheme in Harrington for being overly rigid. As an alternative, they suggested a family of functions that permitted the target value to be anywhere in the region between product specifications. Castillo et al. (1996) improved the individual criteria desirabilities of Derringer (1994) in order to achieve greater smoothness and differentiability. This smoothness is useful because it can improve the performance of gradient-based solvers in optimizing the derived desirability functions (Ribardo & Allen, 2003, p.228).

Özler (1999) reviewed the desirability function proposed by Derringer & Suich (1980) in his study. He also proposed a new d_i function which is not very sensitive to deviations below a target value of a quality characteristic, but is more sensitive to deviations at high magnitudes.

In Costa, Lourenço & Pereira (2011) a review on desirability-based methods under adverse variance conditions is presented. These methods are assessed according to bias, quality of predictions and robustness through optimization measures and the usefulness of those measures to select the compromise solution is evaluated. In another study, Costa & Lourenço (2016) aimed at helping who need to select an effective criterion for solving multiresponse optimization problems developed under the *RSM* framework, and find better solutions. Therefore, they evaluated the working abilities of several criteria of varying sophistication built on popular and novel approaches.

Candioti, De Zan, Cámara & Goicoechea (2014) presented a review about the application of *RSM* when several responses have to be optimized at the same time in the field of analytical methods development is presented. Also they discussed some critical issues like response transformation, multiple response optimization and modeling with least squares and artificial neural networks. They presented most of the recent analytical applications in some specific contexts.

In Akteke-Öztürk, Weber & Köksal (2015) the nondifferentiability issues of desirability funcitons of Derringer & Suich (1980) with a structural approach are elaborated. A new multi-objective optimization method is suggested which is motivated by more than one nondifferentiable point case of desirability functions. In another study, Akteke-Öztürk et al. (2017) focus on the uncertainty effect related with response models in desirability functions of Derringer & Suich (1980). The study addresses core areas of analytic optimization theory and contributes to the usage of disjunctive optimization and generalized semi-infinite optimization, in the course of robustifying optimization of generalized desirability functions.

2.7.1.1 The Desirability Function of Harrington

Harrington (1965) used two steps to calculate the desirability function. The first step concentrated on each criterion/response to assign an individual desirability. Criteria divided into two types: ‘two-sided’ criteria whose acceptable values were bounded by both an upper specification limit (*USL*) and a lower specification limit (*LSL*) and 'one-sided' criteria whose acceptable values were bound by a single specification limit (Ribardo & Allen, 2003). For two-sided criteria, the process of assigning desirability by calculating the scaled response value $Y'_j(x)$ was performed using;

$$Y'_j(x) = \frac{2Y_j(x) - (USL + LSL)}{USL - LSL} \tag{2.27}$$

Then, the user would need to input a single scaled criterion value for criterion j, $Y_{j,0}$, and its assumed desirability d_0 (somehow independently of the values of other criteria), e.g. when $Y'_{j,0} = -0.1$ then $d_0 = 0.63$. This would be appropriate if −0.1 (slightly lower than the midpoint between the specification limits) corresponded to a 'good' system. This pair of numbers ($Y'_{j,0}$, d_0) was used to calculate the parameter n using the following formula;

$$n = \frac{\ln[\ln(1/d_0)]}{\ln|Y'_{j,0}|} \tag{2.28}$$

Next, the desirability for the two-sided characteristic was obtained using;

$$d_j(Y_j(x)) = \exp[-|Y'_j(x)|^n] \tag{2.29}$$

For single-sided criteria, the individual desirability measures or 'desirabilities' were calculated as follows. The engineer had to input two pairs ($Y_{j,1}$, d_1) and ($Y_{j,2}$, d_2) with the criteria values, $Y_{j,1}$ and $Y_{j,2}$, given in actual response units and it was assumed, without loss of generality, that $Y_{j,1} > Y_{j,2}$. Each of the response values was then scaled using the formula;

$$Y'_{j,1} = -\ln[-\ln(d_1)] \quad \text{and} \quad Y'_{j,2} = -\ln[-\ln(d_2)] \tag{2.30}$$

Then, the scaled criteria value, $Y'_j(x)$, corresponding to the actual response $Y_j(x)$ was found through the following linear transformation;

$$Y'_j(x) = [(Y_j(x) - Y_{j,2})/(Y_{j,1} - Y_{j,2})](Y'_{j,1}(x) - Y'_{j,2}(x)) + Y'_{j,2}(x) \tag{2.31}$$

Next, the desirability for the one-sided characteristic was estimated using;

$$d_j\left(Y_j(x)\right) = \exp[-\exp(-Y'_j(x))] \tag{2.32}$$

In the second stage to estimate the system desirability, individual criteria desirabilities were combined using the following formula involving subjective weights of each of the criteria, w_i;

$$D(x) = \left[d_1\left(Y_1(x)\right)^{w_1} d_2\left(Y_2(x)\right)^{w_2} \ldots d_m\left(Y_m(x)\right)^{w_m}\right]^{1/S} \tag{2.33}$$

where $S = \Sigma_i \; w_i$.

In its original formulation, it was argued that w_i = 1 for all responses i was sufficient for most cases of interest. In subsequent research, several advantages for considering unequal weights were described in Derringer (1994). These included that weighting provides a more direct method to adjust the relative importance of alternative criteria than changing the other desirability parameters, e.g. Y_{j0} and d_0. Harrington's rating system for interpreting the desirability is denoted in Table 2.5.

Table 2.5 Harrington's rating system for interpreting the desirability, d

Rating	Description
1.00	The ultimate in satisfaction and quality (an improvement beyond this point would have no appreciable value)
1.00–0.80	Acceptable and excellent (represents unusual quality or performance well beyond anything commercially available)
0.80–0.63	Acceptable and good (represents an improvement over the best commercial quality)

Table 2.5 continues

0.63–0.40	Acceptable but poor (quality is acceptable to the specification limits but improvement is desired)
0.40–0.30	Borderline (if specification exists, then some of the product quality lies exactly on the specification maximum or minimum)
0.30–0.00	Unacceptable (materials of this quality would lead to failure)
0.00	Completely unacceptable

2.7.1.2 The Desirability Function of Derringer-Suich

One and two-sided desirability functions are used depending on whether the response is to be maximized or minimized or has an assigned target value. For a response with target value level, *A*, *B*, and *C* are assigned so that $A \leq B \leq C$. A product is considered unacceptable if $\hat{y} < A$ or $\hat{y} > C$. The value *B* is the "most desirable value" (target). The quantity d_i, the desirability, is defined as;

$$d_i = \begin{cases} \left(\dfrac{\hat{y} - A}{B - A}\right)^s, A \leq \hat{y} \leq B \\ \left(\dfrac{\hat{y} - C}{B - C}\right)^t, B \leq \hat{y} \leq C \end{cases} \tag{2.34}$$

with *d* being *0* if $\hat{y} > C$ or $\hat{y} < A$. One can then use Figure 2.14 to allocate power values *s* and *t* according to one's subjective impression about the role of this response in the total desirability of the product.

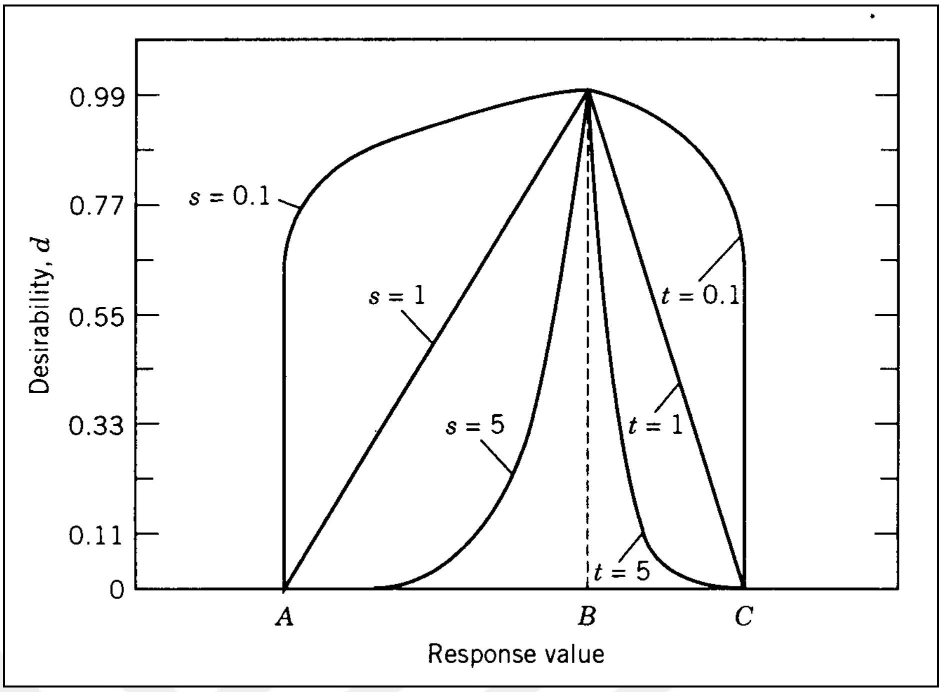


Figure 2.14 Desirability function for target value B

In the case where a response should be maximized, one chooses a value *B* such that $d = 1$ for any $\hat{y} > B$. However, we assume that the product is unacceptable ($d = 0$) if $\hat{y} < A$. Here, $B = C$, then the desirability function is given by;

$$d_i = \left\{ \left(\frac{\hat{y} - A}{B - A} \right)^s, A \le \hat{y} \le B \right. \tag{2.35}$$

In the case where a response is to be minimized, a value *B* is chosen such that $\hat{y} \le B$ is quite desirable and produces a $d = 1$. A value for $\hat{y} > C$ is considered unacceptable and therefore results in a $d = 0$. In this case, $A = B$. The desirability function is given by;

$$d_i = \left\{ \left(\frac{\hat{y} - C}{B - C} \right)^t, B \le \hat{y} \le C \right. \tag{2.36}$$

In most applications, values *A*, *B*, and *C* are chosen according to the researcher's priorities. Choices for *s* and *t* may be more difficult. Choice of *s* and *t* are determined by how important it is for $\hat{y}$ to be close to the target *B*. Using small values for *s* and *t* essentially does not require the response to be close to target, but a choice of *s* and *t* as large as implies that the desirability value is very low unless $\hat{y}$ gets very close to target. For this reason, it can be tried for various levels of *s* and *t* (Myers & Montgomery, 1995, p.250-251).

The general approach is to first convert each response y_i into an individual desirability function d_i that varies over the range $0 \le d_i \le 1$, where if the response y_i is at its goal or target, then $d_i = 1$, and if the response is outside an acceptable region, $d_i = 0$. Then the design variables are chosen to maximize the overall desirability;

$$D = (d_1 * d_2 * ... * d_m)^{1/m} \tag{2.37}$$

where there are *m* responses (Montgomery, 2012, p. 498).

"Although a given increase in $\hat{y}$ does not necessarily result in a uniform increase in the decision maker's preference, in the transformed desirability space, however, each increment in d_i represents the same marginal return to the decision maker" (Fuller & Scherer, 1998, p.4017).

The overall desirability, *D*, is the geometric mean of the individual desirability values. This provides an overall assessment of the combined response functions levels. Maximizing *D* enables the simultaneous optimization of the geometric mean of the transformed response functions. The value of *D* that close to 1 implies that, all responses are in a desirable range simultaneously (Osborne, Armacost & Edwards, 1997, p.3834).

The rationale behind using the geometric mean in formula (2.33) and (2.37) is that if any quality characteristics has an undesirable value (i.e., $d_i = 0$) at some operating conditions, then the overall result is usually a product which is wholly unacceptable, regardless of the values taken on by the other responses (Castillo, Montgomery & McCarvile 1996, p.337).

"After the "optimum" or desirable condition on *x* has been determined, the researcher should do confirmatory runs at that condition to be sure that all responses are in a satisfactory region" (Myers & Montgomery, 1995, p.252).

The appearance of the desirability function for the maximum $\hat{y}$ case, the impact of various *s* values are shown in Figure 2.15.

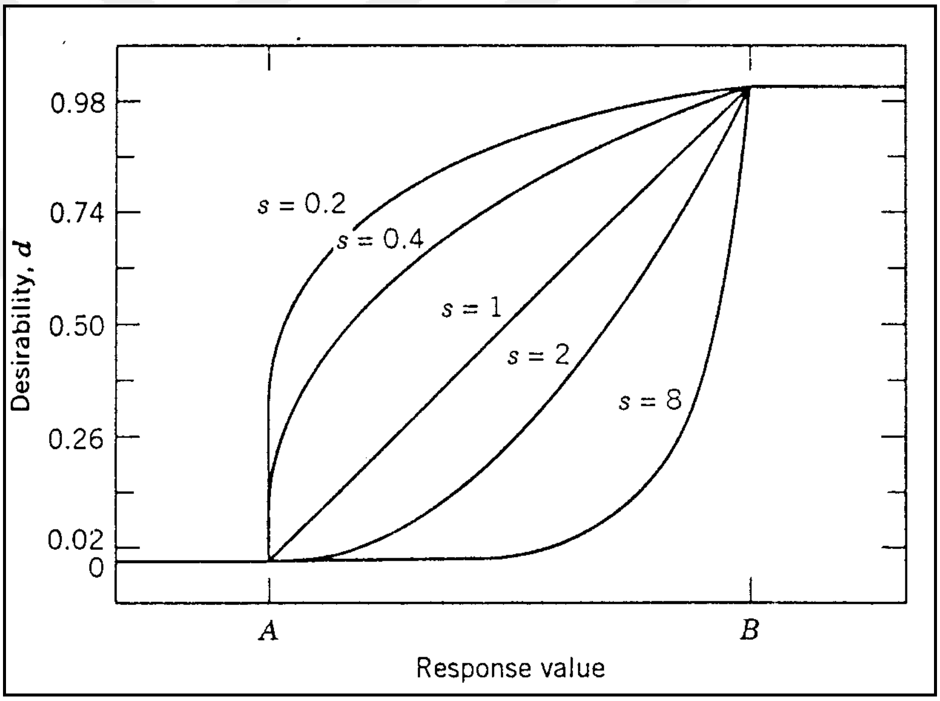


Figure 2.15 Desirability function for a response to be maximized

The decision maker may consider gains in the i^{th} response to be of constant, increasing or decreasing marginal worth. The rate of the marginal desirability of increases in the i^{th} response is controlled by the variable "*s*" in the transformation equation (2.35). If the marginal desirability is constant then there is a linear relationship between $\hat{y}$ and d_i and as such the value of *s* is set to 1. If additional units of $\hat{y}$ above the minimum acceptable value A are of decreasing marginal worth, then there is a concave relationship between $\hat{y}$ and d_i and as such the value of s is less than 1. Finally, if additional units of $\hat{y}$ above the minimum acceptable value A are of increasing marginal worth, then there is a convex relationship between $\hat{y}$ and d_i and as such the value of s is greater than 1. In the case that any value above the minimum is acceptable, a small value s would be selected (e.g. s = 0.2). In the case where approaching the process maximum, B, is considered highly desirable, then a large value of s would be selected (Fuller & Scherer, 1998, p.4017).

CHAPTER THREE
LITERATURE REVIEW

After searching the last decade of literature, it has been seen that there are 51 studies have been published are reviewed and listed in chronological order and discussed. Given the review in Paul, Reddy & DeFlitch (2010) can be seen for the studies had been published in previous years. This review can be classified into two main groups and examined under two headings; simulation optimization in emergency departments (*EDs*) and response surface methodology (*RSM*) in healthcare systems.

3.1 Literature on Simulation Optimization in *EDs*

The first part of the review includes 42 studies on simulation optimization in *EDs* listed in Table 3.1 in a chronological order. The studies are briefly given below in a chronological order.

In their studies, Ahmed & Alkhamis (2009) presented a simulation optimization study to design a decision support tool for the *ED* of a hospital in Kuwait. Their aim was to determine the optimal staffing level which maximizes patient throughput and reduces the waiting time of patients subject to budget constraints. The simulation optimization model presented in this study provided optimal staffing allocation that would allow a 28% increase in patient throughput and an average of 40% reduction in patients' waiting time with using the same resources.

Paul, Reddy & DeFlitch (2010) reviewed the simulation studies within the area of *ED* overcrowding. This review organized the analysis of simulation studies have been published from 1970 to 2006 under several headings. They stated that although simulation had been useful in identifying critical resources and process improvements that can relieve overcrowding somehow, these studies still had important limitations that needed to be addressed. Additionally, they made suggestions about the subjects which are needed to be focused in the future works.

Weng, Cheng, Kwong, Wang & Chang (2011) implemented a simulation optimization approach for the allocation of resources optimally in an *ED* in Taiwan. After demonstrating the waiting time and system time of patients in *ED* by simulation, then the optimization approach was applied to increase the performance in *ED* and management of treated patients and thus increase the degree of satisfactions of patients. The results of the analysis showed that the overall performance in *ED* can be increased by 8% by new human resources allocation studied.

In the first study of Cabrera, Taboada, Iglesias, Epelde & Luque (2011), they presented an agent-based simulation optimization to design an optimal staff configuration in an *ED*. They used an exhaustive search technique which was stated as implying a lot of search time. In their following study (Cabrera, Taboada, Iglesias, Epelde & Luque, 2012), the main objective was to optimize the number of staff in an *ED* to minimize length of stay under a budget constraint. They stated the pipeline scheme as a better optimization approach rather than exhaustive search technique. Also, in another paper of Cabrera, Luque, Taboada, Epelde & Iglesias (2012, December), an agent-based modeling and simulation to find the best *ED* staff configuration for designing a decision support system for the *ED* of a hospital in Spain was presented. Lastly, a simulation optimization study to find the optimal staff configuration in an *ED* was presented in a PhD dissertation (Cabrera, 2013).

In his study, Feili (2013) used a simulation optimization technique to determine the best allocation of resources in an *ED* of a hospital. The objective was reducing the wait times of patients. As a result of the study, it was seen that by implementing the outcome of the study and optimization of the hospital resources, waiting time for the patients could be reduced significantly, and also the related costs could be controlled properly.

Tan, Lau & Lee (2013) presented dynamic patient prioritization strategies and dynamic resource adjustment policies to concurrently manage demand and supply, respectively. The aim of the study was to improve patient’s length of stay in an *ED* considering demand and supply. They verified through a simulation that using an integrated framework improves the patients' length of stay in the *ED* without restricting the demand.

Chen (2014) presented a multi-objective simulation optimization algorithm designed by integrating a non-dominated sorting genetic algorithm II with multi-objective computing budget allocation. The aim of the study was optimizing medical resources allocation to minimize the patient average length of stay and medical resource wasted costs. The effectiveness and feasibility of this method were identified by conducting the experiment, and the experimental analysis showed that the proposed distributed simulation optimization framework can effectively reduce the simulation time.

Pinto, de Campos, Perpétuo & Ribeiro (2014) proposed a queuing model to estimate the required beds for a region in Brazil and also built a simulation optimization model. The objective of the simulation optimization was to minimize the total number of beds with a constraint which limits the refusal rate up to a specific point. They compared the number of beds estimated by both queuing model and simulation model.

Ghanes et al. (2014) constructed a simulation optimization model of staffing levels in a hospital *ED* in France. Indistinguishably, their aim was to minimize the average length of stay under a budget constraint and improve the *ED* operations performance. With this study, the *ED* managers with the general management of the hospital were allowed to take an important tactical decision on increasing the current human staffing budget by 10% in order to reduce the current average length of stay by 33%. In another study, Ghanes et al. (2015) proposed a simulation optimization study which optimizes the human resource staffing levels in an *ED* of a hospital in France. Their simulation model was based on a comprehensive understanding of the *ED* functioning and through optimization they studied the effect of the staffing budget on length of stay, and showed that it has a diminishing marginal effect.

Lal, Roh & Huschka (2015) explained the methodology and applications of simulation based optimization, emphasized advantages, challenges and opportunities of using this method in healthcare through an application. In the paper, they presented the concept of simulation based optimization, its potential areas of application in healthcare and detailed examples of problems where simulation based optimization was successfully applied, to provide the basic education needed to encourage its usage in healthcare systems.

Keshtkar, Salimifard & Faghih (2015) presented a simulation optimization study in order to find the optimal configuration in *ED* resources that affects a patient's length of stay, subject to budget constraints. Simulation was used to analyze the system and to estimate target function. Then an optimization model was solved under different budget constraints to see the improvement in patient length of stay in *ED*.

In their studies, Zeinali, Mahootchi & Sepehri (2015) developed a simulation based optimization model to reduce the total waiting time of patients due to budget and capacity constraints considering the configuration of *ED* resources. After constructing the simulation model, Artificial Neural Networks was selected as the best metamodel within three candidate metamodels which are *RSM*, Radial Basis Functions, and Artificial Neural Networks. The total average waiting time of patients was reduced by 48% by implementing the optimum allocation obtained from proposed optimization.

In one of the recent applications, an algorithmic approach that seeks to enable *ED* decision makers especially in the triage to optimally schedule evaluations for patients who are waiting for treatment in the *ED* was proposed. An algorithm with simulation was used for reducing patients' length of stay, queues for beds in departments and the *ED*, in addition cumulative treatment time in the *ED* (Elalouf & Wachtel, 2015).

El-Rifai, Garaix, Augusto & Xie (2015) presented a stochastic optimization model to solve the shift scheduling problem in an *ED* of a hospital in France. The objective of the study was to optimize shift distribution among employees and to minimize patient's waiting time. The proposed optimization approach in the study allowed finding the promising cyclic schedules. The framework proposed was easily extensible and allowed the integration of several strategic and operational constraints.

Uriarte, Zúñiga, Moris & Ng (2015) presented discrete event simulation and multi-objective optimization approach for a system improvement analysis in a Swedish *ED*. They chose the non-dominated sorting genetic algorithm II as the optimization algorithm. Multi-objective optimization was proved to be a suitable technique to support design and improvement processes of healthcare systems.

Ibrahim, Liong, Bakar, Ahmad & Najmuddin (2015) aimed to develop a simulation optimization model to determine the best allocation of resources in *ED* to reduce patients' length of stay in the review paper of work in progress of a study was conducted in a government hospital in Malaysia. In their following study, Ibrahim, Liong, Bakar, Ahmad & Najmuddin (2017) applied a simulation optimization approach to find the best configuration of resources that can minimize patient waiting time in an *ED* of a hospital in Malaysia. Firstly, by using simulation the actual *ED*'s operations of a public hospital in Malaysia were modeled. Then an optimization approach was used to find the possible combinations of a number of resources that can minimize patients waiting time while increasing the number of patients served. The improvement model significantly improved *ED*'s efficiency with an average of 32% reduction in average patients waiting times and 25% increase in the total number of patients served.

Guo & Tang (2015) proposed a simulation optimization approach to create a decision support system for the *ED* managers to design the optimal configuration of resources with the objective minimize patients length of stay and minimize the overall costs separately. A methodology was presented, which used simulation combined with optimization to determine the optimal resource staffing and scheduling plans, called configurations, to minimize total costs from the standpoint of hospital and to minimize the patients' length of stay in the *ED* system in the perspective of patients.

One of the most comprehensive reviews of *ED* simulation applications for normal conditions was conducted by Gul & Guneri (2015). This review included all papers related to the *ED* simulation from Winter Simulation Conference paper archive and journal articles from various databases between the years 1974–2014. This paper also included *ED* simulation applications for disaster conditions and showed the gaps in this area. They aimed to reveal the importance of simulation for disaster preparedness of *ED*s and to show the innovative aspects of recent *ED* simulation applications unlike the available literature surveys. Another paper (Almagooshi, 2015) described simulation models in healthcare which was developed in the past two decades. Many of the simulation systems were reviewed which were ranging from simulation of patient flow in emergency rooms to simulation of populations with specific chronic diseases. Discrete event simulation and agent based simulation were included as simulation types. The paper has explored the trends within the researches in simulation platforms used to design these models. A trend in data collection methodology was also identified. Agent characteristics in agent based simulations were also examined in addition to the all process of discrete and agent based simulations.

Chen & Wang (2016) proposed a simulation and multi-objective optimization approach to determine the optimal allocation of *ED* resources to minimize the patient's average length of stay and medical resource wasted cost subject to budget constraints. A multi-objective stochastic optimization model was proposed to solve medical resource allocation problems in the *ED* system. A multi-objective simulation optimization algorithm integrating two different algorithms was developed to handle the stochastic and the multi-objective nature of the mathematical model. After that, a discrete event simulation model of the *ED* flow generated from a hospital case in Taiwan was also constructed to estimate the expected performance values of each medical allocation solution obtained.

He, Cai & Wang (2016) considered changing the procedure to improve resource allocation and the comprehensive benefits of West China Hospital of Sichuan University *ED* system. After designing a new registration procedure based on the deficiencies of the previous model and compute-assisting method were used to find the optimal allocation of resources. As a result, the resource utilization rates and the patients' waiting times of the new procedure were more optimal than the previous model.

Another study for estimating the impact of changes in *ED* was done by Kang & Lobo (2016). To that end, a discrete event simulation model and a first-order linear regression model were built to quantify the impact of factors on the performance measures such as; patients' length of stay in express care area, overall length of stay, the time between patient arrival and the time patient is seen by a physician (door-to-door) in express care area, and overall door-to-door. The results showed that the current *ED* express care area functions were so close to the optimal conditions. However, the used discrete event simulation and metamodel indicated that they were useful tools to support decision making in *ED*.

In the paper of Azadeh, Ahvazi, Haghighii & Keramati (2016), a simulation optimization of an *ED* in a general hospital in Iran by modeling human errors based on skill, rule, and knowledge based behavior were conducted. This study combined simulation technique with stochastic data envelopment analysis and data envelopment analysis in order to improve the quality of care in *ED* by modeling three different human errors. The results showed that addition of physician and nurse in *ED* would reduce the human errors, patient duration and queue length. This paper differed from the others as being the first paper that examined human errors in a general hospital by simulation, data envelopment analysis and stochastic data envelopment analysis approaches.

K & Harris (2016) presented a meta-model based simulation optimization to allocate resources to reduce the total average waiting time of patients in an *ED*. In this paper Monte Carlo Sampling approach was proposed for efficient selection of training data. Then experimentation for decision support using metamodels was for resource allocation and thereby to address accuracy and robustness of the corresponding metamodels. This concept was used in the proposed work to reduce the average waiting time of patients in need of emergency treatment.

Baesler (2016) used simulation and *RSM* to find the optimum combination of additional resources required to meet the increase in patient demand and to minimize patient flow time in an emergency room of a private hospital in Chile. This study used *RSM* as an optimization approach but it considered each time only one response which was the minimization of patients' flow time in the system.

El-Zoghby, Farouk & El-Kilany (2016) presented a framework with combining genetic algorithm, discrete event simulation and system dynamics to optimize the *ED* and other areas of the hospital linked to the *ED*. This paper firstly included a review of literature clearly showed that multi-objective simulation modeling of the *ED*'s patient flow was one of the most frequently studied topics in healthcare management. In the framework used, the genetic algorithm was used for the optimization of resource levels throughout the *ED*, which directly affects patient flow. Two simulation tools were utilized in the framework, discrete event simulation, which replicates the performance of that *ED* with optimized resources, and system dynamics, which models the entire healthcare system, studying the interrelationships that exist between the *ED* and the rest of the hospital. Results of simulation confirmed the need for an integrated framework to analyze the performance of a healthcare system.

Oh, Novotny, Carter, Ready, Campbell & Leckie, (2016) presented a simulation based decision support tool for the redesign of an *ED*. A throughput time goal was keeping the time from arrival to departure under three hours for 80% of *ED* patients. Target areas for improvement were identified including optimized process flow, resource allocation and operational policies by using discrete event simulation modeling. After the changes in *ED* processes, 81% of patients had a length of stay in the *ED* of less than three hours and also a 30% improvement in average patient length of stay was achieved.

A simulation and optimization method is proposed to characterize the durations of service times when the process duration data are not available in an *ED* in Guo, Goldsman, Tsui, Zhou & Wong (2016). To characterize the key service durations, firstly lognormal distributions were used. Thereafter, a new meta-heuristic approach was proposed, which combines an improved adaptive genetic algorithm and simulated annealing to search for the optimal set of service time distribution parameters. Then this combined approach was used along with optimal computing budget allocation technology. Optimal computing budget allocation technology minimized the total simulation cost while achieving the best parameter sets. The experimental results showed that the proposed method can find accurate estimates of service time distribution parameters in a short time.

In one of the recent papers, defining a medical resource allocation problem in an *ED* in Taiwan, a multi-objective stochastic optimization model which simultaneously minimizes average patient length of stay and total medical wasted costs were constructed (Feng, Wu & Chen, 2017). By integrating a non-dominated sorting genetic algorithm II with multi-objective computing budget allocation, better and more suitable allocation solutions for *ED* decision-makers were generated.

In the paper of Liu (2017), an agent-based model and simulation techniques were used to understand, predict and optimize the complex behavior of *ED*. This simulation optimization method created a model to see the interaction of *ED* components (patient, nurse, doctor, equipment etc.) of a hospital in Spain. The two contributions of the study were to help quantitatively predicting and analyzing the complex behavior of *ED*s by an agent-based model for and to develop a systematic method to automatically calibrate a general *ED* model with incomplete data by using simulation and optimization based methodologies.

Another study was done to create a systematic method to calibrate a general *ED* model with incomplete data (Liu, Rexachs, Epelde & Luque, 2017a). For this purpose, a simulation-based optimization was used and by a real case study, the proposed method was proven to be suitable way of calibrating and validating the simulation model with incomplete data. In other paper of Liu et al. (2017b), an agent-based simulation optimization model was presented to understand the complexity of *ED* units, evaluate policies and study problems such as prolonged waiting times, inefficient use of *ED* resources, and unbalanced staff scheduling.

In Guo, Gao, Tsui & Niu (2017) which was applied in an *ED* of a public hospital in Hong Kong presented a simulation optimization for the staff configuration problem in order to minimize the total labor cost considering service level requirements (patients' waiting time for treatment). The results indicated significantly higher practicability and efficiency of the proposed method with different patient arrival rates and service constraints.

Uriarte, Zúñiga, Moris & Ng (2017) presented a novel approach while combining discrete event simulation, simulation-based multi-objective optimization and data mining techniques to provide support to decision makers within a Swedish *ED*. As a result of this study, the decision makers were provided with a range of nearly optimal solutions and design rules which reduce the length of stay and waiting times for *ED* patients. These solutions included the optimal number of resources and the required level of improvement in key processes.

Another study is made to present a simulation-based optimization tool for resource planning in an *ED* of a hospital (Yousefi, Yousefi, Ferreira, Kim, & Fogliatto, 2017). The proposed genetic algorithm based optimization approach was tested in a public *ED*, and it was shown to decrease the average length of stay in this *ED* case study by 14%. Furthermore, the proposed metamodel showed a 26.6% improvement compared to the average results of other algorithms of mean absolute percentage error.

Rabbani, Farshbaf-Geranmayeh & Yazdanparast (2018) used simulation and multi-response optimization approach and considered clinical pathway method in order to consider multi-departments. This study used artificial neural network for the multi-response optimization part. After the application of the approaches in this paper, the new resource level decreased patients' waiting time for bed by 32%, reduced mean waiting time in triage queue by about 16%, and decreased mean waiting time in the drugstore by 64%.

Yazdanparast, Hamid, Azadeh & Keramati (2018) presented a study which includes an integrated algorithm for optimizing resource allocation with respect to human error in an *ED* in Tehran, Iran. The algorithm was composed of simulation, artificial neural network, design of experiment and fuzzy data envelopment analysis. It was a multi-response optimization approach to optimize human error, cost, wait time, and patient safety, and productivity.

Table 3.1 Literature on simulation optimizations in EDs

	Study	**Country**	**Problem type**	**Aim(s) of the study**	**KPIs**	**Software**	**Data gathering method/source**
1	Ahmet & Alkhamis (2009)	Kuwait	Staff allocation/ decision support system (DSS)	To determine the optimum number of staff levels in order to maximize patient throughput and reduce waiting time under budget constraints	Patient throughput, patient average waiting time in system	Excel, Simscript	A comprehensive survey, observations, interviewed doctors, nurses and hospital personnel
2	Paul, Reddy & DeFlitch (2010)	USA	Reviewed the simulation studies investigating emergency department (ED) overcrowding				Healthcare systems engineering, operations research and computer science publication venues
3	Weng, Cheng, Kwong, Wang & Chang (2011)	Taiwan	Resource allocation	To allocate the resources optimally in ED	National Emergency Department Overcrowding Scale, score used to measure the degree of congestion of ED	Simul8	Data of an ED in Taiwan

Table 3.1 continues

4	Cabrera, Taboada, Iglesias, Epelde & Luque (2011)	Spain	Staff configuration/ DSS	To find the optimal staff configuration in an ED	Patient waiting time, CT (cost*time)	NetLogo	Interviews with ED staff
5	Cabrera, Taboada, Iglesias, Epelde & Luque (2012)	Spain	Staff configuration/ DSS	To optimize the number of staff members of EDs	Patient’s length of stay (LOS)	NetLogo	ED data
6	Cabrera, Luque, Taboada, Epelde & Iglesias (2012)	Spain	Staff configuration	To find the best ED staff configuration	LOS	NetLogo	Taken from the studied hospital
7	Cabrera (2013)	Spain	Staff configuration	To find the optimal staff configuration in an ED	LOS, throughput, CLoS (Cost-LOS)	NetLogo	ED data
8	Feili (2013)	Iran	Resource allocation	To determine the best allocation of resources	Waiting times	ExtendSim	Observations

Table 3.1 continues

9	Tan, Lau & Lee (2013)	Singapore	Queue management	To improve patient's LOS in an ED by dynamic queue management considering demand and supply	Average LOS	Java, Sas	Historical data, database of the studied hospital
10	Chen (2014)	Taiwan	Resource allocation	To optimize medical resource allocation to minimize the patient average LOS and medical resource wasted costs (MWCs)	Patient average LOS, MWCs	eM-Plant	Interviews, analysis
11	Pinto, de Campos, Perpétuo & Ribeiro (2014)	Brazil	Bed capacity planning	To estimate the required bed capacity	Number of beds, occupancy rate, refusal rate	Arena	Databases of the Ministry of Health and the Central Admissions of the city
12	Ghanes, Jouini, Jemai, Wargon, Hellmann, Thomas & Koole (2014)	France	Staff configuration	To minimize the average LOS under a budget constraint and improve the ED operations performance	Average LOS	Arena	ED database, other databases, interviews with experts, on-site observations

Table 3.1 continues

13	Ghanes, Wargon, Jouini, Jemia, Diakogiannis, Hellman, Thomas & Koole (2015)	France	Staff configuration	To optimize the human resource staffing levels in ED	LOS, door-to-door (DTD) time	Arena	Records from databases, interviews with experts and decision makers, on-site observations
14	Lal, Roh & Huschka (2015)	USA	Shift scheduling	To discuss the concept of simulation optimization and highlighting its use in healthcare through an application	The absolute error between the offered load and the scheduled capacity for the physician led teams	Arena, Excel	Existing data sources, time studies
15	Keshtkar, Salimifard & Faghih (2015)	Iran	Resource allocation	To minimize patients' total LOS in an ED subject to budget constraint	Resource utilization, average patient waiting time, patients' total average LOS in an ED	Arena	Database, observations, interviews with ED staff
16	Zeinali, Mahootchi & Sepehri (2015)	Iran	Resource allocation	To minimize the total average waiting time of patients subject to both budget and capacity constraints	Total average waiting time of patients	Arena, Design-Expert, Matlab	Archived files, hospital's managers
17	Elalouf & Wachtel, (2015)	Israel	Patient scheduling	To propose an algorithmic approach that enables ED decision makers to optimally schedule evaluations for ED patients	Waiting times, ED queues	Arena	Empirical data obtained from a hospital

Table 3.1 continues

18	El-Rifai, Garaix, Augusto & Xie (2015)	France	Shift scheduling	To solve the shift scheduling problem in order to optimize shift distribution among employees and minimize patient's waiting time	Patient's total expected waiting time, maximum waiting time in queues	IBM Ilog Cplex, AnyLogic	Data of ED in the hospital, data in the literature, experts' opinion
19	Uriarte, Zúñiga, Moris & Ng (2015)	Sweden	DSS	To provide a DSS for system improvement in an ED	LOS , time to meet the physician	FlexSim	Provided data from stakeholders, time studies, ED personnel
20	Ibrahim, Liong, Bakar, Ahmad & Najmuddin (2015)	Malaysia	Resource allocation	To develop a simulation optimization model to find the best allocation of resources to reduce patient's LOS	Average long waiting time, average total LOS, resource's utilization rates	Arena	Interviews with ED administrators, review of documents, observations
21	Guo & Tang (2015)	China	Resource configuration/ DSS	To help ED decision makers to design optimal resource configuration within different constraints	LOS, total costs of all the ED resources	Matlab	Questionnaires, discussion with ED managers, communication with patients waiting in the ED system
22	Gul & Guneri (2015)	Turkey	To reveal the importance of simulation for disaster preparedness of EDs and the innovative aspects of recent ED simulation applications with available literature survey				Winter Simulation Conference archive, journals, databases

Table 3.1 continues

23	Almagooshi (2015)	Saudi Arabia	To present a literature review of simulation models in healthcare				ACM digital library, IEEE Explore, Winter Simulation Conference archive, Journal of Simulation archive
24	Chen & Wang (2016)	Taiwan	Resource allocation	To determine the optimal allocation of ED resources to minimize the patient's LOS and MWC subject to budget constraints	Patient average LOS, MWC	ExpertFit	Real world ED environment, ED administrators, hospital information system
25	He, Cai & Wang (2016)	China	Resource allocation	To improve resource allocation for optimal use of medical workers for revised registration procedure in ED	Resources' utilization, average waiting time	Arena	Observations, real operation data, questionnaire, information from human resource department
26	Kang & Lobo (2016)	USA	Design parameters identification	To estimate the impact of changes in an ED using discrete event simulation and metamodeling	Express care area LOS, overall LOS, express DTD, overall DTD	No clear information	Electronic medical records of patients

Table 3.1 continues

27	Azadeh, Ahvazi, Haghighii & Keramati (2016)	Iran	Human error modeling	To improve the quality of care in ED by modeling three different human errors	Average patient duration, average queue length, number of three human errors	Awesim	Observations of the processes of ED, interviews with technicians, supervisor, nurses and physicians, recorded data
28	K & Harris (2016)	India	Resource allocation	To allocate resources appropriately to reduce the total average waiting time of patients	Total average waiting time of patients	Matlab	Available training data sets
29	Baesler (2016)	Chile	Resource combination	To find the optimum combination of additional resources required for the increase in demand and for minimizing patient flow time	Patient flow time	Flexsim Hc, classical mathematical programming tools	Hospital data
30	El-Zoughby, Farouk & El-Kilany (2016)	Egypt	Resource optimization	To optimize the resource levels in ED	Average LOS	ExtendSim	Data taken from a case study in a literature from a university hospital in Dublin

Table 3.1 continues

31	Oh, Novotny, Carter, Ready, Campbell & Leckie (2016)	USA	DSS	To improve ED throughput	Patient throughput (throughput time, wasting time, left without being seen rate, number of patients in ED/work in process), Resource utilization (staff, beds, equipment)	Arena	Hospital information system, floor observation, staff interviews
32	Guo, Goldsman, Tsui, Zhou & Wong (2016)	Hong Kong	Incomplete data	To characterize the durations of service times when the process duration data are not available in an ED	Time differences between registration and triage, time differences between triage and consultation and time differences between consultation and departure	Arena	ED data
33	Feng, Wu & Chen (2017)	Taiwan	Resource allocation	To define medical resource allocation problem in an ED and minimize LOS and MWC simultaneously	LOS, MWC	eM-Plant, Visual Studio, ExpertFit	Real world ED environment, ED administrators, hospital information system

Table 3.1 continues

34	Liu (2017)	Spain	DSS	To develop a simulation optimization tool to better understand, predict and evaluate the complex behavior of an ED	Patients’ LOS	NetLogo, Java	Real operation data, expertise ED staff, information system database of the hospital
35	Liu, Rexachs, Epelde & Luque (2017a)	Spain	Model calibration	To create a method to calibrate a general ED model with incomplete data	LOS, leave without-being-seen	NetLogo	Information system database, actual operation data
36	Liu, Rexachs, Epelde & Luque (2017b)	Spain	DSS	To understand the complexity of ED units, evaluate policies and study ED operations	Patients’ LOS	NetLogo	Real operation data, interviews with experienced ED staff, historical data
37	Guo, Gao, Tsui & Niu (2017)	Hong Kong	Staff configuration	To find the best staff configuration in order to minimize total labor cost considering service level requirements	Waiting time of categorized patients, LOS	Arena	Comprehensive survey
38	Uriarte, Zúñiga, Moris & Ng (2017)	Sweden	DSS	To provide a decision support tool for an ED	LOS, waiting times	FlexSim, ModeFrontier	Interviews, visits to ED

Table 3.1 continues

39	Yousefi, Yousefi, Ferreira, Kim & Fogliatto (2017)	Brazil	Resource allocation	To present a simulation-based optimization tool for resource planning in an ED	Average LOS	NetLogo, Matlab	Collected data from the hospital
40	Ibrahim, Liong, Bakar, Ahmad & Najmuddin (2017)	Malaysia	Resource allocation	To develop a simulation optimization model to find the best allocation of resources to reduce patient's waiting time	Average patient's waiting time, total number of patient served in the system	Arena	Interviews with ED administrators, review of documents, field observation, time recordings, log books, database system
41	Rabbani, Farshbaf-Geranmayeh & Yazdanparast (2018)	Iran	Resource allocation	To present a simulation optimization approach for optimizing resource allocation in an ED	Mean waiting times of patients in stages 1-4, patients mean waiting time for bed, mean waiting time in drug store queue, mean waiting time in triage queue	Arena, Minitab	Collected data from the ED of the hospital, historical data, expert judgment method from department experts

Table 3.1 continues

42	Yazdanparast, Hamid, Azadeh & Keramati (2018)	Iran	Resource allocation	To present an integrated algorithm for optimizing the resource allocation with respect to human error in an ED	Mean weighted waiting time of patients, mean waiting time for patients in triage, frequency of skill-based error, mean waiting time for bed, redundancy score, cost	Arena	ED data

3.2 Literature on *RSM* in Healthcare Systems

The second part of the review includes 12 studies on *RSM* in healthcare systems listed in Table 3.2 in a chronological order, and also discussed below in a chronological order.

In the master thesis of Grummels (2012) a simulation optimization study was presented for resource allocation in hospitals by using a studied case from the literature. The objective of the thesis was to investigate the applicability of the *RSM* for model based optimal resource allocation in hospitals. Only one response, which included the patient flow time and the resource costs, was aimed to be minimized in this study.

Akcan & Kokangul (2013) built a new approximation for inventory control system and examined this by a healthcare case study. The optimal levels of r (reorder point) and Q (order quantity) were determined in order to minimize the expected annual total inventory cost. A simulation metamodel was constructed to obtain equations for the average on-hand inventory and average number of orders per year. Thereafter, these equations were used to determine the optimal levels of r and Q while minimizing the total cost in an integer non-linear model.

In their study, Kasaie & Kelton (2013a) studied the resource allocation problem in the control of epidemics and proposed a simulation optimization framework to the control of influenza pandemic with several interacting healthcare interventions. It was stated that analyzing the behavior of resource allocation outcomes with regard to different investment strategies and seek optimal allocations were enabled by using the simulation model with optimization techniques. In a companion study, Kasaie & Kelton (2013b) provided a detailed discussion on calibration, analysis and optimization of an agent-based simulation of the problem they discussed in Kasaie & Kelton (2013a).

Sun & Li (2013) proposed a simulation-based *RSM* to optimize surgery start times and to find a way of distributing the buffer time into different surgeries for minimizing total idling and overtime costs. This study addressed the problem of determining start times for pre-sequenced surgery jobs for a single operating room, where stochastic surgery durations were represented by an infinite set of stochastic scenarios. The numerical studies indicated that the levels of the per time unit idling cost, per time unit overtime costs, and variability of random duration have effects on such allocations of the buffer time.

The research of Azadeh, Farahani, Torabzadeh & Baghersad (2014) focused on patient scheduling problem in *ED* laboratories of a hospital by the aim of minimizing the total average waiting time of prioritized patients. The scheduling of patients was modeled as a flexible open shop scheduling problem and a mixed integer linear programming model was developed for the objective. The aim of the proposed approach was to grant higher priority to the most urgent patients while achieving the most overall outcome of the resources by paying enough attention to other patients. A genetic algorithm was proposed to solve the problem and the *RSM* was used only for tuning the parameters and enhancing the performance of the genetic algorithm.

Liang, Guo & Fung (2015) proposed a simulation based optimization approach for surgery scheduling in operating theaters in a hospital. They used *RSM* with two optimization objectives which are maximizing the patient throughput and minimizing the patient waiting time. A combined scheduling policy consists of three simple scheduling rules was proposed to optimize the performance of scheduling operation theatre. The results indicated that the combined scheduling policy was more effective and efficient.

In their studies, Zeinali, Mahootchi & Sepehri (2015) developed a simulation based optimization model to reduce the total waiting time of patients due to budget and capacity constraints considering the configuration of *ED* resources. After constructing the simulation model, Artificial Neural Networks was selected as the best metamodel within three candidate metamodels which are *RSM*, Radial Basis

Functions, and Artificial Neural Networks. The total average waiting time of patients was reduced by 48% by implementing the optimum allocation obtained from proposed optimization.

Baesler (2016) used simulation and *RSM* to find the optimum combination of additional resources required to meet the increase in patient demand and to minimize patient flow time in an emergency room of a private hospital in Chile. This study used *RSM* as an optimization approach but it considered each time only one response which was the minimization of patients' flow time in the system.

K & Harris (2016) presented a meta-model based simulation optimization to allocate resources to reduce the total average waiting time of patients in an *ED*. In this paper Monte Carlo Sampling approach was proposed for efficient selection of training data. Then experimentation for decision support using metamodels was for resource allocation and thereby to address accuracy and robustness of the corresponding metamodels. This concept was used in the proposed work to reduce the average waiting time of patients in need of emergency treatment.

Kokangul, Akcan & Narli (2017) presented a simulation and metamodeling approach to optimize nurse levels subject to multiple performance measures at a neonatal unit of a hospital. A metamodel was built firstly to generate simulation results under various numbers of nurses. Then, those experimented data were used to obtain the mathematical relationship between inputs (number of nurses at each level) and performance measures (admission number, occupation rate, and satisfaction rate) using statistical regression analysis. Finally, several integer nonlinear mathematical models were proposed to find optimal nurse capacity subject to the targeted levels on multiple performance measures.

In another study (Niessner, Rauner & Gutjahr, 2018), a better decision support system for the staff re-allocation decisions of a discrete event simulation-based mass casualty incident management system was provided. Three simulation-optimization techniques were investigated: the Kiefer-Wolfowitz Algorithm, the metaheuristic

OptQuest Approach, and the *RSM*. The results of the study indicated that the optimized automated policies can improve the performance of the small field hospitals called advanced medical posts (*AMP*) compared to the management by simple heuristics or by human decision makers. The policy implications for improving strategic decision making and process management for incident commanders at the Austrian *AMP* based on the results of the dynamic simulation optimization techniques were discussed.

Table 3.2 Literature on RSM in healthcare systems

	Study	Country	Problem type	Aim(s) of the study	KPIs	Modeling approach (software)	Data gathering method/source
1	Grummels (2012)	The Netherlands	Resource allocation	To investigate the applicability of simulation optimization for resource allocation in hospitals	Patients' flow time, costs	Simulation, analytical queuing model (Matlab)	A studied case in the literature
2	Akcan & Kokangul (2013)	Turkey	Inventory control	To determine the optimal levels of reorder point and order quantity for minimizing the expected annual total inventory cost	Cost	Simulation, mathematical model (OptQuest, Design-Expert, Arena)	Data obtained/ collected from Neonatal Intensive Care Unit of a university hospital
3	Kasaie & Kelton (2013a)	USA	Resource allocation	To present an approach for resource allocation problem in the control of epidemics	Attack rate (AR), disability-adjusted life years (DALY)	Simulation, mathematical model (Design-Expert, Microsoft Excel VBA)	Available public health data
4	Kasaie & Kelton (2013b)	USA	Resource allocation	To provide a detailed discussion on calibration, analysis and optimization of an agent-based simulation of a resource allocation problem in the control of influenza epidemics	AR, DALY	Simulation, mathematical model (Design-Expert, Microsoft Excel VBA)	Available public health data

Table 3.2 continues

5	Sun & Li (2013)	USA	Operating room scheduling	To optimize surgery start times for a single operating room	Total cost, buffer time	Simulation, mathematical model (Design-Expert, Matlab, Arena)	Assumptions of a case study
6	Azadeh, Farahani, Torabzadeh & Baghersad (2014)	Iran	Patient scheduling	To improve the ED efficiency by minimizing the total average waiting time of prioritized patients in ED laboratories	Total weighted completion time of patients	Simulation, mathematical model (Lingo, Matlab)	ER data
7	Liang, Guo & Fung (2015)	China	Surgery scheduling	To optimize surgery scheduling in operating theaters	Patient throughput, patient waiting time	Simulation (Arena, Minitab)	Hospital Information System
8	Zeinali, Mahootchi & Sepehri (2015)	Iran	Staff configuration	To minimize the total average waiting time of patients subject to both budget and capacity constraints	Total average waiting time of patients	Simulation, mathematical model (Arena, Design-Expert, Matlab)	Archived files, hospital’s managers
9	Baesler (2016)	Chile	Resource combination	To find the optimum combination of additional resources required for the increase in demand and for minimizing patient flow time	Patient flow time	Simulation, mathematical model (Flexsim HC, classical mathematical programming tools)	Hospital data

Table 3.2 continues

10	K & Harris (2016)	India	Resource allocation	To allocate the resources optimally in ED	Total average time of patients	Simulation, mathematical model (Matlab)	Available training data sets
11	Kokangul, Akcan & Narli (2017)	Turkey	Optimal level of nurse	To optimize nurse capacity subject to multiple performance measures	Admission number, occupancy level, satisfied demand	Simulation, mathematical model (Matlab, Design-Expert)	Collected from the hospital
12	Niessner, Rauner & Gutjahr (2018)	Austria	Decision support system (DSS)	To provide a better DSS for the staff re-allocation decisions of a discrete event simulation-based disaster management system	Total rescue time, total number of patients deceased	Simulation (AnyLogic, JAVA, OptQuest)	Previous study from the literature

CHAPTER FOUR
APPLICATION IN A HOSPITAL

4.1 Information about the System

In this section, general description of the hospital, information about the *ED* of the hospital and the problem encountered in the *ED* are given.

4.1.1 General Description of the Hospital

The hospital is located in İzmir/Turkey and it is a member of a group that has many hospitals all over the country, serves in 50000 m^2 with a capacity of 303 beds. It has nine surgery rooms, 35 newborn incubators, seven coronal care, seven cardiovascular surgery care, 39 general care beds. The hospital is accredited by Joint Commission International (JCI) according to international medical services accreditation standards. It is active in many branches such as liver transplantation, bone marrow transplantation, cardiovascular surgery, and In Vitro Fertilization (IVF) center. The hospital provides intensive services to İzmir and the surrounding population by specially negotiated corporation and government contracts. The acceptance of foreign patients is carried out by international agreements. It has a high rate of foreign patients in the region and the country.

4.1.2 Emergency Department of the Hospital

Emergency status is specified in Social Security Institution Health Practice Declaration as situations like sudden illness, accident, injuries, and similar cases which require urgent responses in the first following 24 hours that the incident occurred. Therefore, it is accepted that loss of life and health integrity occurs when an urgent response to patient or transfer of patience to other emergency departments (*EDs*) are not made.

ED of the hospital accepts patients which have situations obedient to the definition above regardless of their health insurance and ability to pay their treatment expenses. The *ED* accepts all emergency admissions without discrimination. The urgent medical assessment, treatment and stabilization for each patient are provided. Emergency medical service is given from the moment when patient needs service until the exact treatment process finishes without interruption.

4.1.3 Problem Definition

ED is one of the complex systems which have the most intensity of patient. Therefore, this situation causes some problems in the processes of the system. The problem, in this case, is the bottlenecks which cause delays, long length of stays, patient and personnel dissatisfaction in the hospital in İzmir. This problem causes more intensity in the *ED* and decrease in the quality of service. Therefore, determining the optimum personnel and resource numbers plays a vital role for improving the current system. In this study, for an *ED* of a hospital in Izmir, the aim is to find the optimal allocation of resources while minimizing the overall flow time (length of stay) of patients and keeping the utilization ratios of doctors and nurses at the required levels.

4.2 Simulation Modeling

The features of the *ED* system, patient acceptance, assessment and treatment processes, assumptions and the inputs of the simulation model are given in this section. After giving detailed information about developing the simulation model, the verification and validation of the simulation are depicted.

4.2.1 Features of the System

The *ED* system in this case is a non-terminating system which operates 24 hours in a day. Triage is commonly used in crowded emergency rooms to determine which patients should be seen and treated immediately. In the *ED* of the hospital, there are three triage colors defined to categorize patients.

Red code (immediate) are used to label those who cannot survive without immediate treatment but who have a chance of survival. Red code patients are indicated as *Category 3*.

Yellow code is used for labeling situations which have the possibility of threatening life, limb loss and the risk of significant morbidity. Yellow code patients are indicated as *Category 2*.

Green code is used to label patients whose conditions are stable and can heal with a simple treatment. Green code patients are indicated as *Category 1*.

In this study, all these three categories have six different subcategories. These subcategories are obtained and decided from the *ED* information system as they are the most common patients who come to the *ED* to have treatment.

4.2.2 Patient Acceptance, Assessment and Treatment Processes

Patients who come to the *ED* by ambulance or by their own means (walk-in) are welcomed by the *ED* nurse. In the emergency room entrance, patients who need a porter are taken to emergency rooms by a stretcher or a wheelchair. The registration procedures are not made by the patients. They have immediate treatment without waiting registration procedures. The relatives of patients are routed to make registration procedures of the patients.

The triages of patients who come by ambulance and also by walk-in are done after arrival. Next, they have first treatment directly. After the first treatment process, if there is a need for medical examination, vascular access and tests, firstly these examinations in laboratory and radiology departments are made, then the observation room treatment can start. If there is not a need for medical examination, the patients have observation room treatment immediately.

After the assessment of the patient, the *ED* physician writes needed treatments and practices and informs the *ED* nurse. During this time, the patient follow-up is made by the related nurses. After controls, the doctor decides whether the patient can discharge or not. If the doctor decides that the patient can discharge, then the emergency service room treatment of the patient is done. Otherwise, the patient follow-up is made by the related nurses and the doctor.

Emergency patients who need surgical operations are prepared for surgical procedures after assessment. If the patient is discharged, the patient is informed by the *ED* physician. The patient can also be sent to other health institutions when there is an inability to continue the needed healthcare service. If the patient dies in the *ED*, the department acts on accordance with the determined procedure.

The *Category 1* and *Category 2* patient flow (conceptual models) in the *ED* of the hospital are shown in Figure 4.1. The *Category 3* patient flow (conceptual model) is depicted in Figure 4.2.

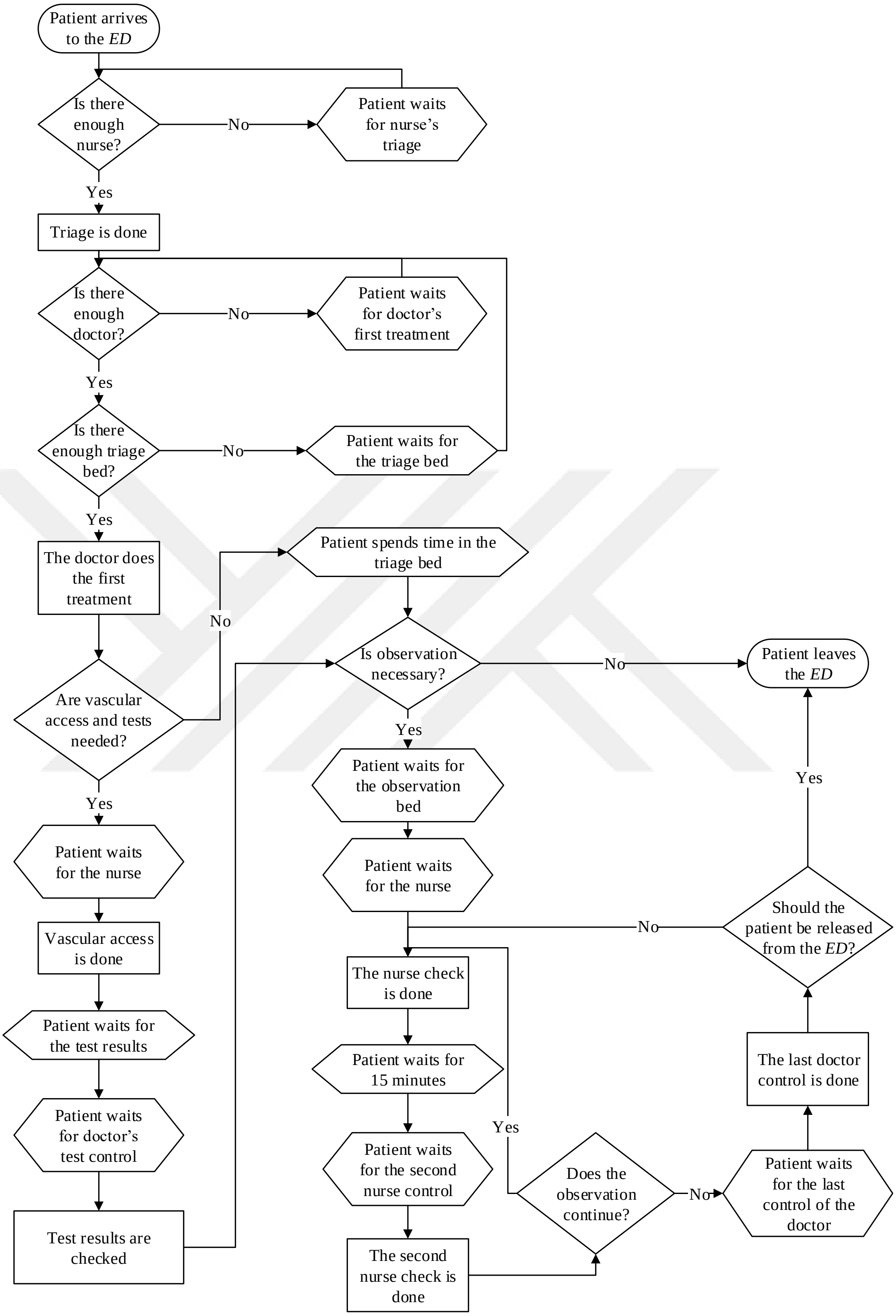


Figure 4.1 Category 1 and 2 patient flow conceptual model

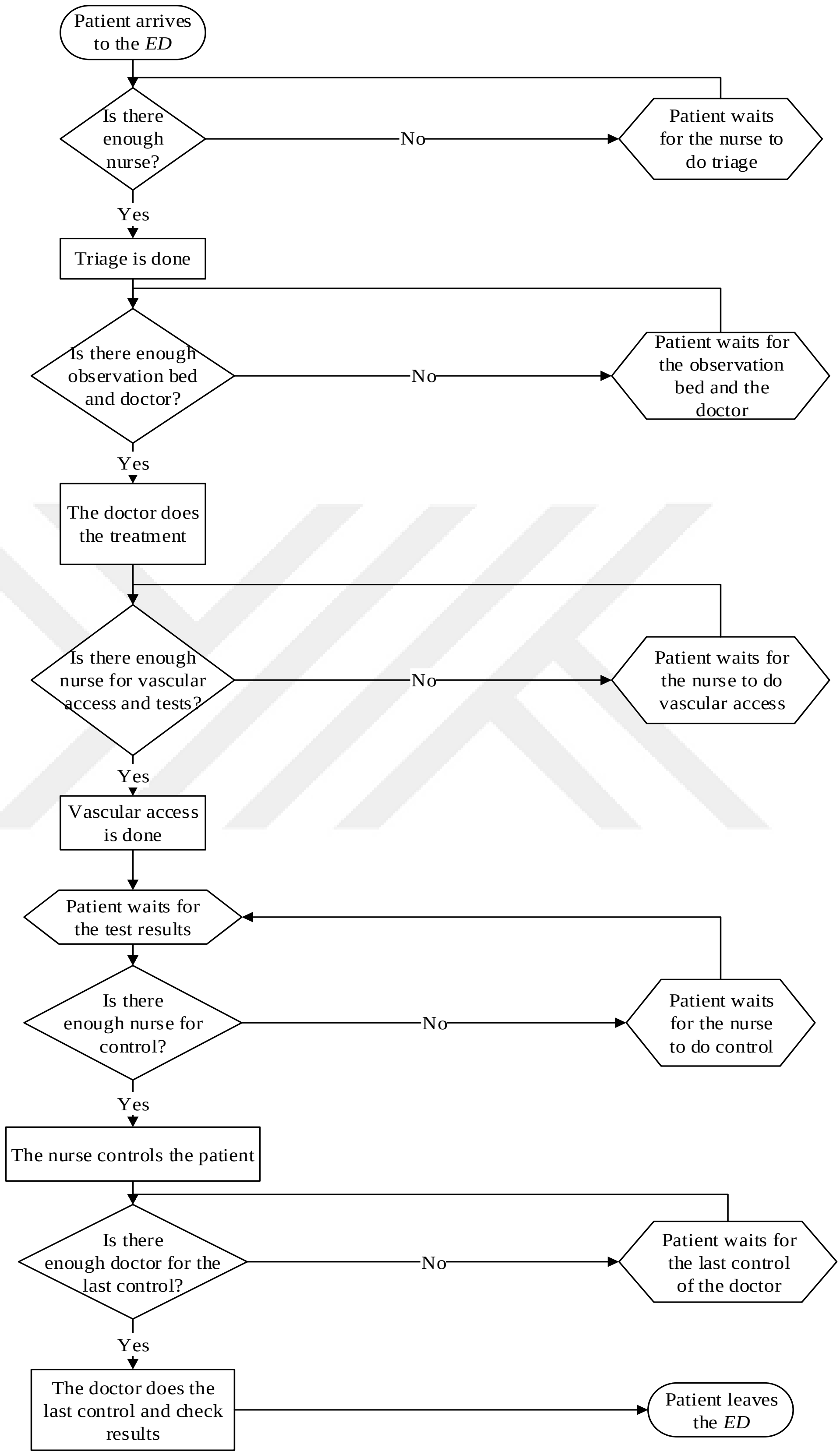


Figure 4.2 Category 3 patient flow conceptual model

4.2.3 Assumptions

- No patient dies in the queues.
- The unit used for time is minutes.
- Each patient can hold only one doctor.
- Each patient can hold only one nurse.
- Each patient can hold only one bed.
- The service times of the patients, the probabilities of the patient categories are collected by observations and expert opinions. The distributions of arrivals are determined by using Input Analyzer tool of Arena simulation modeling software.
- It is assumed that *Category 3* patients will leave the *ED* system when the doctor decides the transfer of these patients to the other units in the hospital by the head doctor's approval.

4.2.4 Inputs of the Simulation Model

The patient arrivals, the probabilities and distributions of tests, observation and service times, and the number of resources are given in detail in this section.

4.2.4.1 Patient Arrivals

Arrivals of the patients to the *ED* change hourly in a day. Therefore, in the model 24 different distributions of patient arrivals are used according to 24 timeslots in the *ED* (Table 4.1). Probabilities of patient arrivals according to categories and subcategories are shown in Table 4.2 and Table 4.3, respectively. The patient arrival times and probabilities of the patient categories are obtained from Hospital Information System and they are provided by the management of the hospital. Data on arrivals are given by the hospital is related to the first week of January 2016. Input Analyzer Tool of Arena simulation modeling software is used to find the distributions of the patient arrivals to the *ED* for each timeslot. The significance level of 5% (α=0.05) is used for chi-squared goodness of fit tests for the arrivals.

Table 4.1 Inter-arrival times distributions of patients

From to	Timeslot	Arrivals Distributions	Corresponding p-value
00:00 – 01:00	1	EXPO(5.15)	0.3400
01:00 – 02:00	2	EXPO(11.7)	> 0.1500
02:00 – 03:00	3	WEIB(8.31, 0.835)	0.1480
03:00 – 04:00	4	EXPO(8.18)	> 0.1500
04:00 – 05:00	5	EXPO(15)	> 0.1500
05:00 – 06:00	6	31 + WEIB(17.9, 0.668)	> 0.1500
06:00 – 07:00	7	2 + EXPO(36.8)	> 0.1500
07:00 – 08:00	8	1 + EXPO(36.2)	> 0.1500
08:00 – 09:00	9	EXPO(15)	> 0.1500
09:00 – 10:00	10	EXPO(11.9)	0.4290
10:00 – 11:00	11	EXPO(6.75)	0.0817
11:00 – 12:00	12	WEIB(5.32, 0.894)	> 0.1500
12:00 – 13:00	13	-0.001 + LOGN(12.8, 40.1)	> 0.1500
13:00 – 14:00	14	LOGN(4.66, 9.97)	0.0916
14:00 – 15:00	15	EXPO(4.25)	0.0767
15:00 – 16:00	16	EXPO(5.56)	0.4930
16:00 – 17:00	17	EXPO(5.38)	0.4690
17:00 – 18:00	18	-0.001 + EXPO(4.48)	0.1820
18:00 – 19:00	19	LOGN(4.38, 6.74)	> 0.1500
19:00 – 20:00	20	EXPO(3.81)	0.0987
20:00 – 21:00	21	EXPO(3.29)	0.2180
21:00 – 22:00	22	EXPO(3.54)	0.1910
22:00 – 23:00	23	EXPO(3.51)	0.6020
23:00 – 24:00	24	EXPO(4.38)	0.1320

The EXPO(mean), WEIB(scale, shape) and LOGN(mean, standart deviation) are exponential, weibull and lognormal distributions respectively and the values in front of some distributions are smoothing factors. “The exponential distribution is often used to model random arrival and breakdown processes” (Pegden, Shannon & Sadowski, 1990, p.561). “The Weibull distribution is widely used in reliability models to represent the lifetime of a device. …. This distribution is also used to represent non-negative task times that are skewed to the left” (Pegden, Shannon & Sadowski, 1990, p.568). “The lognormal distribution is used in situations in which the quantity is the product of a large number of random quantities. It is also frequently used to represent task times that have a non-symmetric distribution" (Pegden, Shannon & Sadowski, 1990, p.563).

Table 4.2 Arrival probabilities of patient categories

Category	Triage code	Percentage in all patients (%)
1	All greens	43.64
2	All yellows	38.69
3	All reds	17.67

Table 4.3 Arrival probabilities of patient subcategories

Category	Sub-category	Diagnosis	Percentage in all patients (%)	Percentage in its own category (%)
1	11	Upper respiratory infections	19.93	45.66
1	12	Urinary system dysfunction	1.41	3.22
1	13	Anxiety disorder	0.92	2.10
1	14	Acute pharyngitis	5.75	13.17
1	15	Acute tonsillitis	1.89	4.34
1	16	Other greens	13.75	31.51
2	21	Gastrenteritis and colitis, noninfected	5.62	14.53
2	22	Soft tissue dysfunction, fractures	6.85	17.69
2	23	Pneumonia	5.50	14.22
2	24	Essential (primary) hypertension	2.81	7.27
2	25	Acute bronchitis	6.72	17.38
2	26	Other yellows	11.19	28.91
3	31	Acute myocardial infarction	11.19	63.32
3	32	Intracranial haemorrhage (nontraumatic)	1.28	7.27
3	33	Atherosclerotic cardiovascular disease	0.24	1.38
3	34	Subarachnoid haemorrhage	0.61	3.46
3	35	Cerebrovascular disease	0.24	1.38
3	36	Other reds	4.10	23.19

4.2.4.2 Test and Observation Distributions and Probabilities

In the *ED* of the hospital, the probabilities of doing tests and observations change according to the patient subcategory. The tests probabilities, the observation probabilities and the test times are obtained by interviews with the *ED* doctors and nurses (Table 4.4). Input Analyzer Tool of Arena simulation modeling software is used to find the distributions of the test times. The significance level of 5% (α=0.05) is used for chi-squared goodness of fit tests.

Table 4.4 Probabilities and distributions of tests and observation times

Category	**Sub-category**	**Percentage of tests (%)**	**Percentage of non-tests (%)**	**Tests times**	**Percentage of observation (%)**	**Percentage of non-observation (%)**
1	11	5	95	NORM(45,5)	4	96
1	12	40	60	NORM(45,5)	4	96
1	13	0	100	0	4	96
1	14	5	95	NORM(45,5)	4	96
1	15	5	95	NORM(45,5)	4	96
1	16	20	80	NORM(45,5)	4	96
2	21	10	90	NORM(37.5,2.5)	95	5
2	22	80	20	NORM(60,10)	0	100
2	23	90	10	NORM(75,15)	95	5
2	24	50	50	NORM(60,10)	95	5
2	25	80	20	NORM(37.5,2.5)	0	100
2	26	70	30	NORM(45,5)	95	5
3	31	100	0	105	100	0
3	32	100	0	105	100	0
3	33	100	0	120	100	0
3	34	100	0	105	100	0
3	35	100	0	105	100	0
3	36	100	0	105	100	0

4.2.4.3 Activity Times Distributions and Probabilities

All other required activity times that every patient get during their process in the *ED* are collected by using stopwatch, observations and interviews with the *ED* personnel. Input Analyzer Tool of Arena simulation modeling software is used to find the distributions of activity times. The significance level of 5% (α=0.05) is used for chi-squared goodness of fit tests. These times are shown in Table 4.5.

Table 4.5 Distributions of activity times

Activity	Distribution
Triage by nurse	NORM(6,0.33)
First treatment by doctor	NORM(1.5,0.17)
Keeping doctor busy in case of no chance to get a triage bed	NORM(0.375,0.042)
Vascular access by nurse	NORM(5,0.33)
Test control by doctor	NORM(4,0.33)
Compulsory occupation of the triage bed in case of no tests	NORM(1.5,0.17)
Observation control (1) by nurse	NORM(6.5,0.17)
Compulsory waiting between observation checks	15
Observation control (2) by nurse	NORM(3.5,0.17)
Last observation control by doctor	NORM(3.5,0.17)
Check (category 3) by doctor	NORM(4,0.33)
Test control (category 3) by nurse	NORM(3.5,0.17)
Last control (category 3) by doctor	NORM(6,0.33)

4.2.4.4 Resource Numbers

In *ED*s, there are human resources who are qualified personnel such as doctors and nurses. And *ED*s have also material resources which helps the *ED* service such as beds used in triage and observation processes. In the studied current system, there are three doctors, four nurses, four triage beds and nine observation beds.

4.2.5 Developing the Simulation Model

In this section, all elements and blocks used in the simulation model are given in detail.

4.2.5.1 Elements

Elements are the program fragments for which model-specific definitions are made.

ATTRIBUTES

TimeIn is defined to record the entering times of the patients and used by TALLIES element to calculate the length of stay.

Category is defined to determine the categories of the patients.

Subcategory is defined to determine the sub-categories of the patients.

VARIABLES

For recording the data about the entities, *Timeslot* variable is defined with zero initial value. It is used for keeping the data of hourly changes during a day.

RESOURCES

Four types of resources are defined in the simulation model. These are:

- *Doctor*
- *Nurse*
- *BedTriage*
- *BedObservation*

EXPRESSIONS

This element is used for keeping the data of patient arrivals of 24 timeslots in *InterarrivalTimes* array.

QUEUES

Patients join the queue of doctors, nurses, beds and also for some tests in the process of emergency services.

DSTATS

It is defined to record the utilization of resources and also the average number of patients in the queues.

TALLIES

It is defined to record the average overall flow times of patients (length of stay) and times spent by each of the patient categories.

REPLICATE

Replication number is determined as 10. The model is run through 10 weeks (100800 minutes). Because the system is a non-terminating system, a warm-up period should be determined. The warm-up period is determined as 2000 minutes. The graphic in Figure 4.3 is used for determining the warm-up period. The replication length is 100800+2000=102800 minutes.

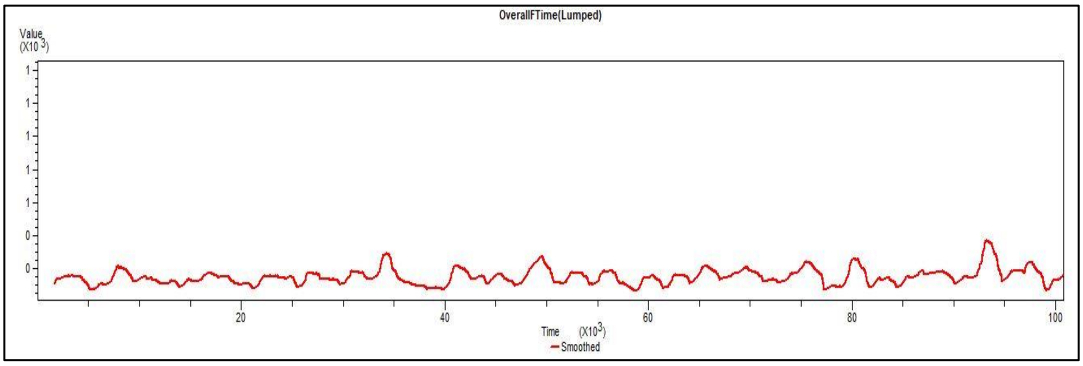


Figure 4.3 Moving average plot for warm-up

OUTPUTS

Outputs element is included in the model to record statistics and create an output file. The data will be used to compare the average overall flow time (length of stay), average doctor utilization and average nurse utilization of the current system and the optimized system.

All elements are shown in Figure 4.4.

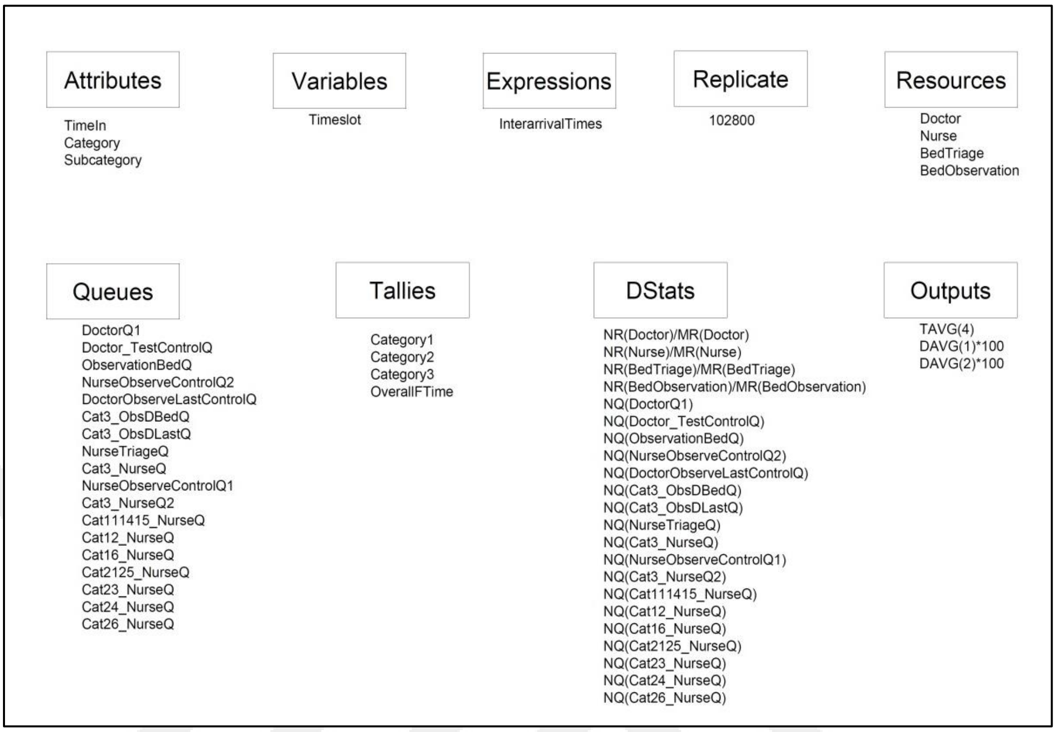


Figure 4.4 Elements of the simulation model

4.2.5.2 Blocks

Blocks are the program particles that make up the logical and mathematical relationships of the model.

As shown in Figure 4.5, this part of the simulation model is created for providing the hourly changes of patient arrivals to the *ED* of the hospital. First, *timeslot* is assigned to *timeslot+1*. After spending one hour (60 minutes), the value of *timeslot* is controlled with BRANCH block. If its value is equal to 24 then, the value of *timeslot* is assigned to zero. Next, the process continues in a loop.

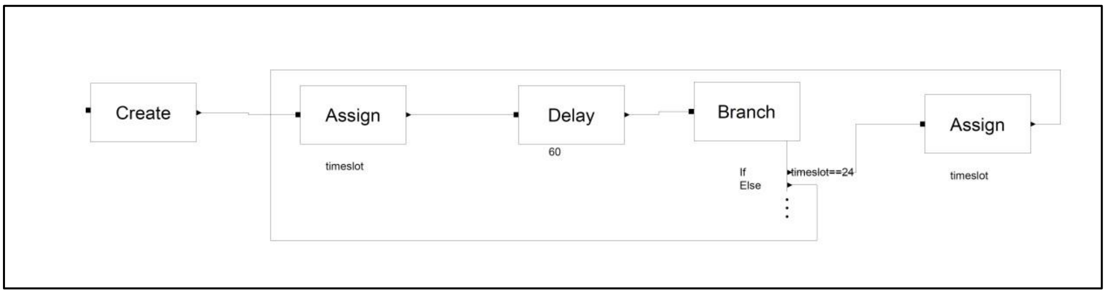


Figure 4.5 Blocks of the simulation model-1

As depicted in Figure 4.6, the patient arrivals to the *ED* are defined in the CREATE block of the model with distributions kept in *InterarrivalTimes* and they are called from this expression according to the *timeslot* which have hourly changes. First creation time is 0.0000001 because the other CREATE block is used to define hourly changes and its first creation is at time zero and must generate on entity at the beginning of simulation time. After the patient arrives to the *ED*, she/he gets a triage by the nurse. For the triage, they join the *NurseTriageQ* and they seize a nurse. They spend time in the DELAY block with a distribution of NORM(6,0.33). At the end of the triage, the category of patients is determined in ASSIGN block and the patient releases the nurse. Then, the subcategories of the patients are assigned to them within three ASSIGN blocks.

As given in Figure 4.7, the upper two ASSIGN blocks are used for assigning subcategories to *Category 1* and *Category 2* patients, which are green and yellow coded patients, respectively. The lower ASSIGN block is used for assigning subcategories to *Category 3* patients who are red coded patients and their flow in the *ED* is represented in detail later below and denoted in Figure 4.14, 4.15 and 4.16. *Category 1* and *Category 2* patients wait in QUEUE block for seizing a doctor. After seizing the doctor, the availability of triage bed is checked in the BRANCH block. If there is an empty triage bed for the patient, then the patient can have the triage bed and spends time in the DELAY block derived with NORM(1.5,0.17). If there is no available triage bed, the patients spends some time with the doctor subject to distribution of NORM(0.375,0.042) and then leaves the doctor. Lastly, in BRANCH block, *Category 1* and *Category 2* patients are checked according to their subcategories and they start to follow their own treatments.

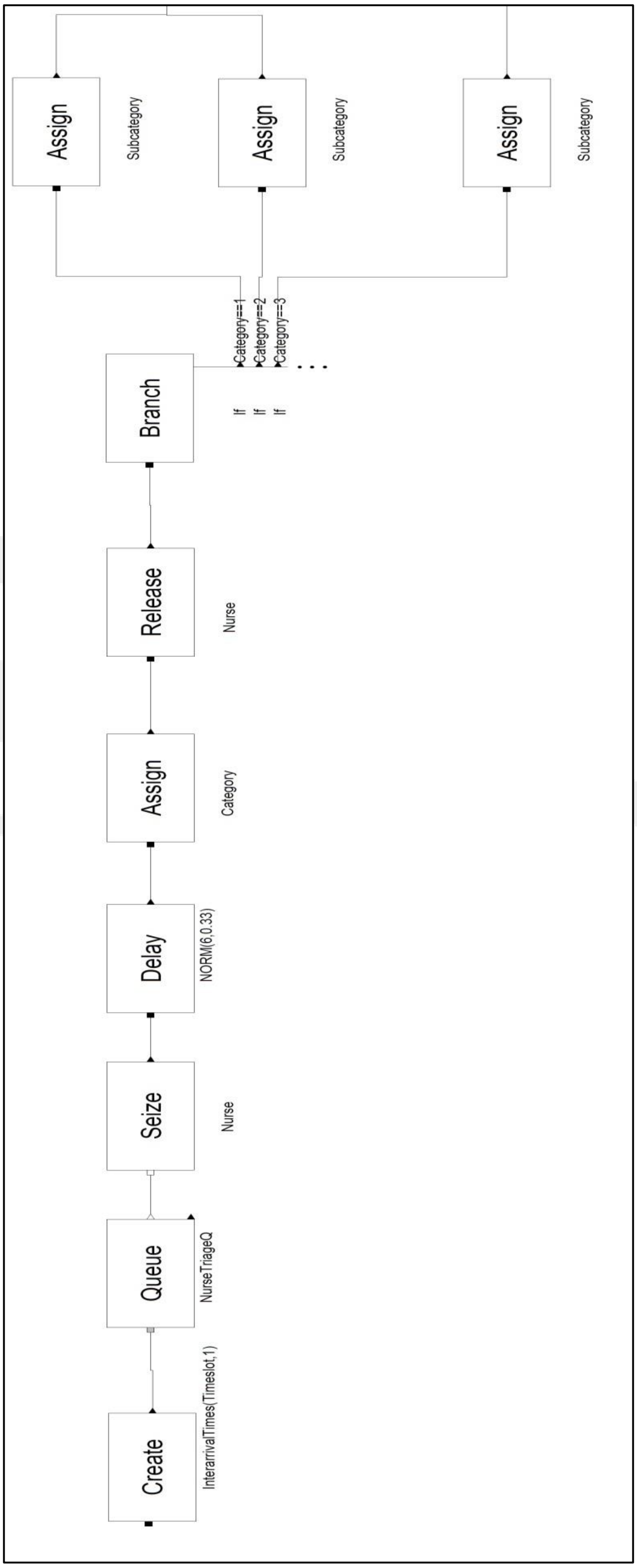


Figure 4.6 Blocks of the simulation model-2

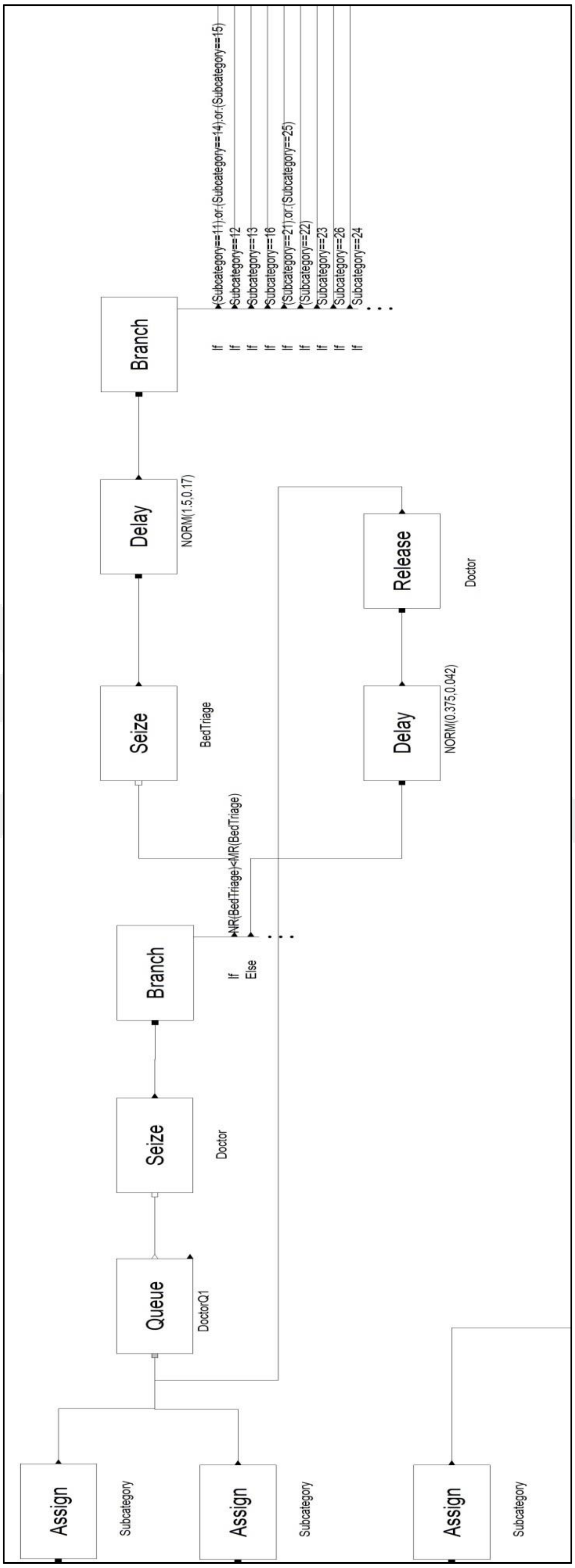


Figure 4.7 Blocks of the simulation model-3

As shown in Figure 4.8, if the patient's subcategory is 11, 14 or 15, the patient has 5% test probability. Patients who need to have tests, release the doctor, and wait the nurse for vascular access in QUEUE block. The patient has vascular access with NORM(5,0.33), releases the nurse, then waits a time for the test results derived from NORM(45,5) in DELAY block, next waits in the doctor's test control queue in QUEUE block (please look Figure 4.11). The patient with 95% who does not need to have tests, spends time with the doctor in DELAY block with NORM(1.5,0.17), then goes to the first BRANCH block denoted in Figure 4.11.

If the patient's subcategory is 12, the patient has 40% test probability. Patients who need to have tests, release the doctor, and wait the nurse for vascular access in QUEUE block. The patient has vascular access with NORM(5,0.33), releases the nurse, after that waits a time for the test results derived from NORM(45,5) in DELAY block, next waits in the doctor's test control queue in QUEUE block given in Figure 4.11. The patient with 60% who does not need to have tests, spends time with the doctor in DELAY block with NORM(1.5,0.17), then goes to the first BRANCH block depicted in Figure 4.11.

If the patient's subcategory is 13, the patient does not have tests. Therefore, these patients spends time with the doctor derived from NORM(1.5,0.17) after that goes to the first BRANCH block (please look Figure 4.11).

If the patient's subcategory is 16, the patient has 20% test probability. Patients who need to have tests, release the doctor, and wait the nurse for vascular access in QUEUE block. The patient has vascular access with NORM(5,0.33), releases nurse, next waits for the test results derived from NORM(45,5) in DELAY block, then waits in the doctor's test control queue in QUEUE block shown in Figure 4.11. The patient with 80% who does not need to have tests, spends time with the doctor in DELAY block with NORM(1.5,0.17), after it goes to the first BRANCH block given in Figure 4.11.

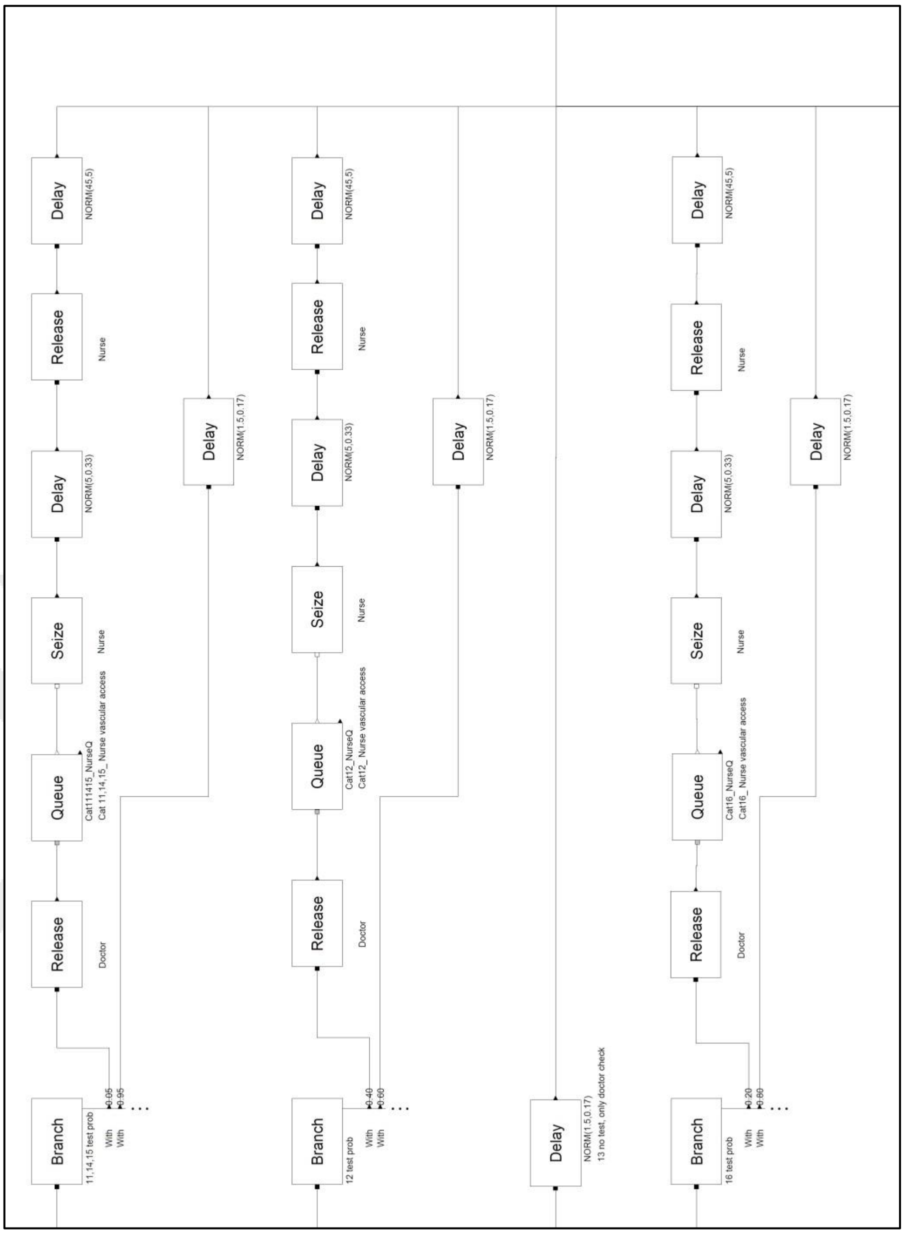


Figure 4.8 Blocks of the simulation model-4

Figure 4.9 and Figure 4.10 denotes the flow of *Category 2* patients according to their test probabilities. If the patient's subcategory is 21 and 25, the patient has 10% test probability. Patients who need to have tests, release the doctor, and wait the nurse for vascular access in QUEUE block. The patient has vascular access with NORM(5,0.33), releases the nurse, then waits a time for the test results derived from NORM(37.5,2.5) in DELAY block, after that waits in the doctor's test control queue in QUEUE block (please look Figure 4.11). The patient with 90% who does not need to have tests, spends time with the doctor in DELAY block with NORM(1.5,0.17), next goes to the first BRANCH block given in Figure 4.11.

If the patient's subcategory is 22, the patient has 80% test probability but these subcategory patients do not need vascular access. The patient waits a time for the test results derived from NORM(60,10) in DELAY block, after that waits in the doctor's test control queue in QUEUE block depicted in Figure 4.11. The patient with 20% who does not need to have tests, spends time with the doctor in DELAY block with NORM(1.5,0.17), then goes to the first BRANCH block shown in Figure 4.11.

If the patient's subcategory is 24, the patient has 80% test probability. Patients who need to have tests, release the doctor, and wait the nurse for vascular access in QUEUE block. The patient has vascular access with NORM(5,0.33), releases the nurse, next waits for a time the test results derived from NORM(60,10) in DELAY block, later waits in the doctor's test control queue in QUEUE block denoted in Figure 4.11. The patient with 20% who does not need to have tests, spends time with the doctor in DELAY block with NORM(1.5,0.17), then goes to the first BRANCH block given in Figure 4.11.

If the patient's subcategory is 23, the patient has 90% test probability. Patients who need to have tests, release the doctor, and wait the nurse for vascular access in QUEUE block. The patient has vascular access with NORM(5,0.33), releases the nurse, next waits for a time the test results derived from NORM(75,15) in DELAY block, then waits in the doctor's test control queue in QUEUE block shown in Figure 4.11. The patient with 10% who does not need to have tests, spends time with doctor in DELAY block with NORM(1.5,0.17), later goes to the first BRANCH block depicted in Figure 4.11.

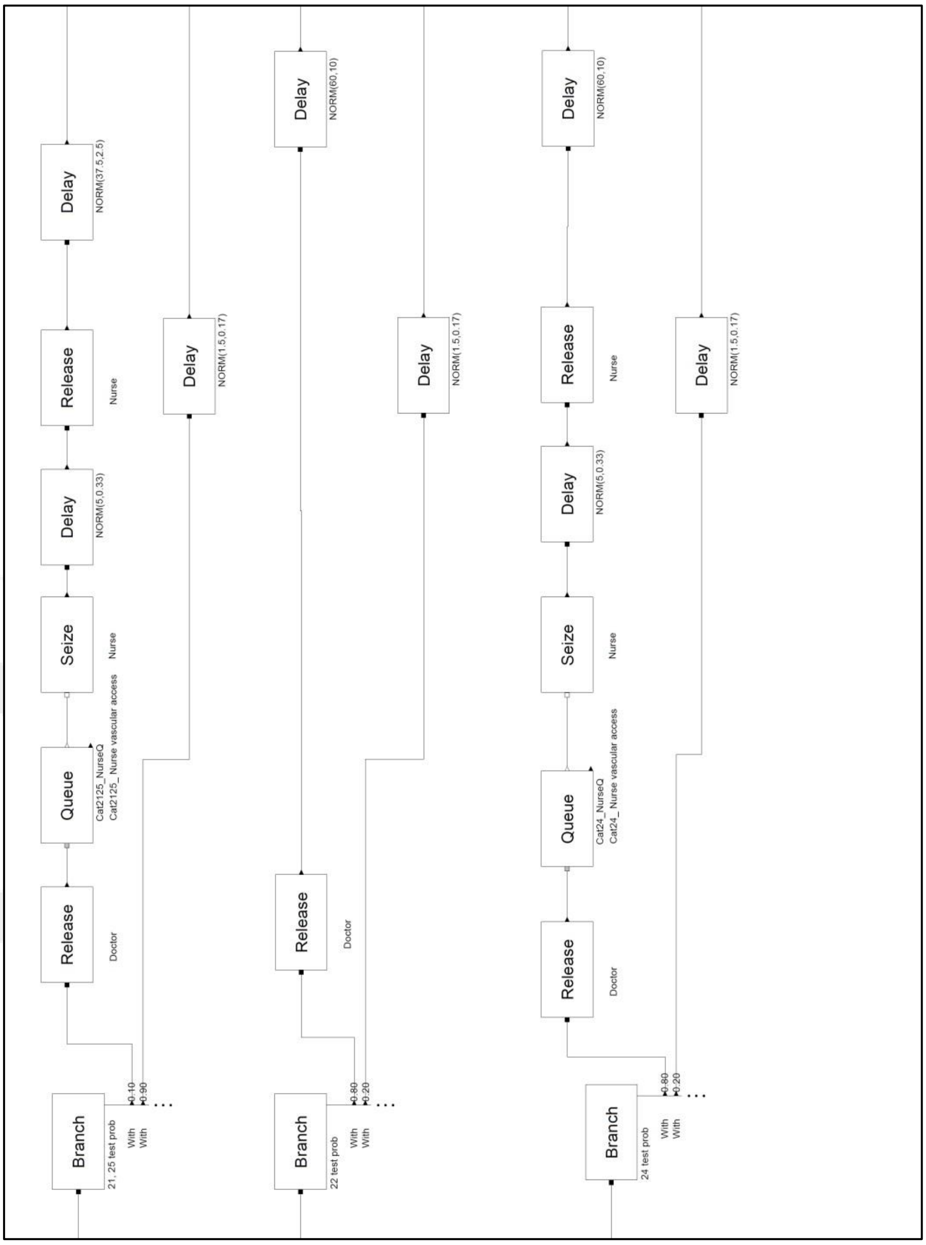


Figure 4.9 Blocks of the simulation model-5

If the patient's subcategory is 26, the patient has 70% test probability. Patients who need to have tests, release the doctor, and wait the nurse for vascular access in QUEUE block. The patient has vascular access with NORM(5,0.33), releases the nurse, then waits for the test results with duration of NORM(45,5) in DELAY block, next waits in the doctor's test control queue in QUEUE block given in Figure 4.11. The patient with 30% who does not need to have tests, spends time with the doctor in DELAY block with NORM(1.5,0.17), after that goes to the first BRANCH block denoted in Figure 4.11.

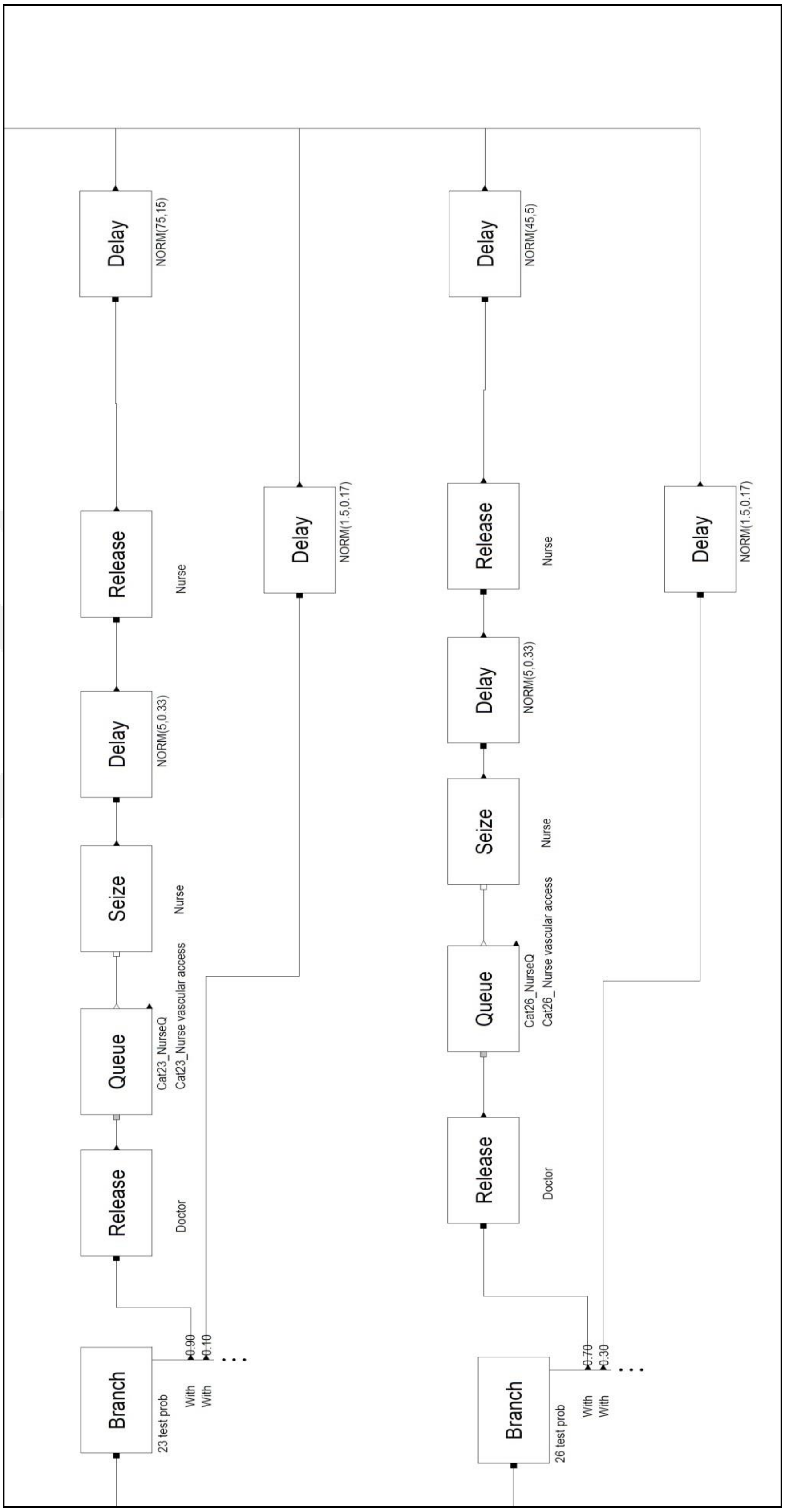


Figure 4.10 Blocks of the simulation model-6

As depicted in Figure 4.11, patients wait in QUEUE block for the doctor to check the results. The doctor checks the results in a duration derived from NORM(4,0.33) in DELAY block. *Category 1* and *Category 2* patients are controlled according to their categories in BRANCH block after the doctor's check. If the patient's category is 1, the patient has 4% going on observation probability. Therefore, the patient, who needs to have observation, releases the doctor and waits in QUEUE block for observation bed. If the *Category 1* patient does not need observation, the patient goes to RELEASE block and releases the bed triage and the doctor, and then leaves the system after passing TALLY blocks shown in Figure 4.17.

If the patient's category is 2, then the subcategories of these patients are also checked. If the subcategory is 21, 23, 24 or 26, the patient has 95% observation probability. The patient, who needs to have observation, releases the doctor and waits in QUEUE block for observation bed. If the *Category 2* patient does not need observation, the patient goes to RELEASE block and releases the bed triage and the doctor, and then leaves the system after passing TALLY blocks given in Figure 4.17.

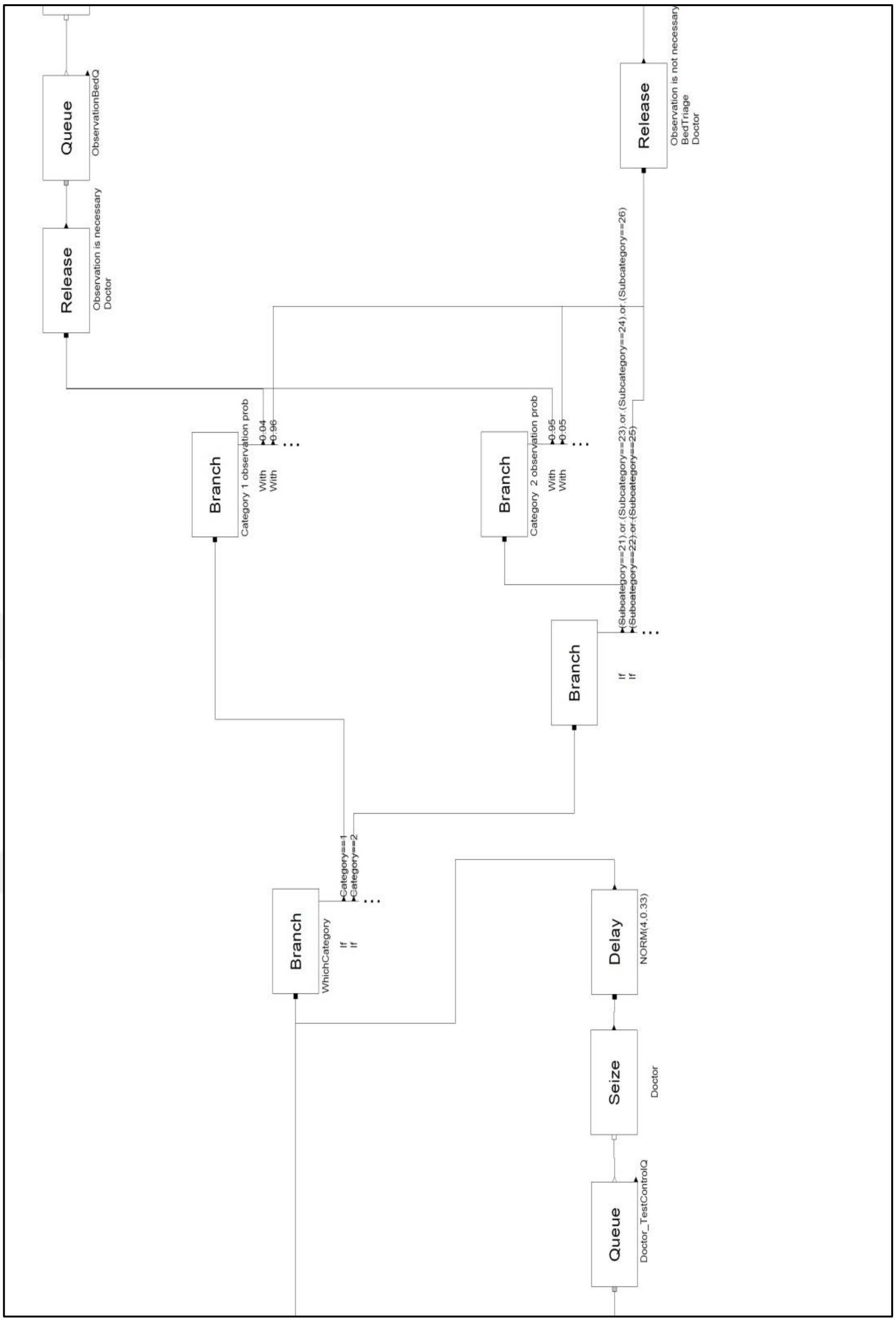


Figure 4.11 Blocks of the simulation model-7

As denoted in Figure 4.12, the patients who need observation, only after seizing the observation bed, can release the triage bed. After seizing the observation bed, they wait for the nurse to do a control. This control takes a time derived from NORM(6.5,0.17) in DELAY block, then the patients leave the nurse and wait for 15 minutes in the observation bed for the nurse's second control.

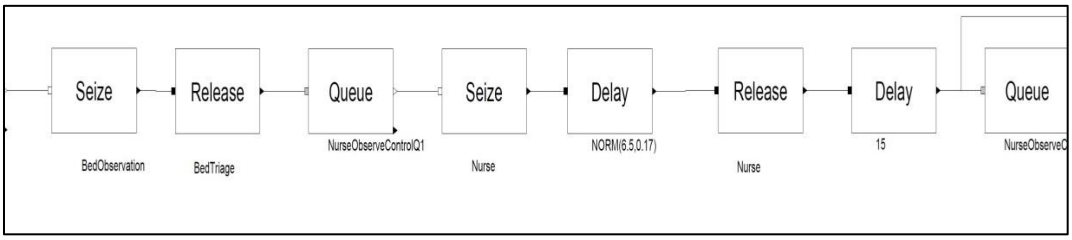


Figure 4.12 Blocks of the simulation model-8

As shown in Figure 4.13, patients wait in QUEUE block for the nurse's second control and this control takes a time derived from NORM(3.5,0.17), after that patients leave the nurse. In the BRANCH block, whether the observation will continue or not is checked. With 25% probability, the observation continues. Therefore, patients spend time in bed for additional 15 minutes and wait for the nurse to do control again join to the QUEUE block. With 75% probability, the observation does not continue. These patients, who do not need more observation, wait in QUEUE block for a doctor to do last control, spends time with the doctor in DELAY block derived from NORM(3.5,0.17), and after that releases the doctor. In BRANCH block, whether the discharge of patients will be done or not is checked. With 20% probability, the discharge cannot happen, therefore patients do not release the observation bed, they go to DELAY block to wait 15 minutes and have all these previously told steps again. With 80% probability, the discharge happens. At that time the patients leave the observation bed and then leave the system after passing TALLY blocks given in Figure 4.17.

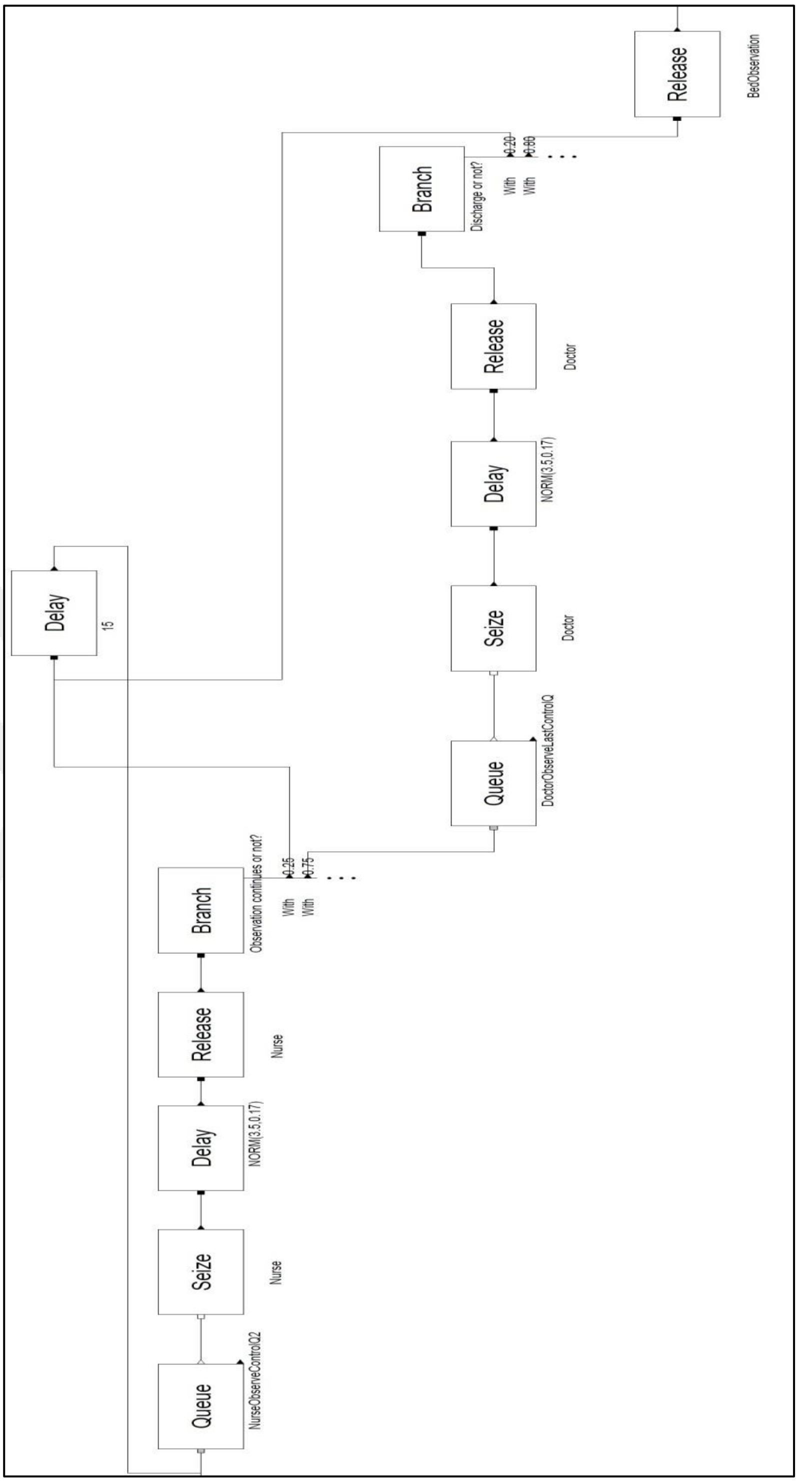


Figure 4.13 Blocks of the simulation model-9

Figure 4.14 shows the flow of *Category 3* patients in the *ED* emergency department, after their subcategories are assigned to them. After waiting in QUEUE block, *Category 3* patients seize an observation bed and a doctor. In DELAY block the doctor does control and the first treatment with a time derived from NORM(4,0.33). After releasing the doctor, the patients waits for a nurse to do vascular access takes a time derived from NORM(5,0.33), and then release the nurse.

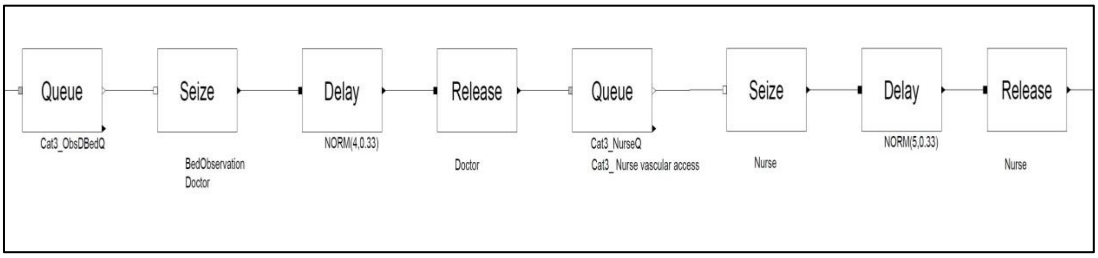


Figure 4.14 Blocks of the simulation model-10

As depicted in Figure 4.15, *Category 3* patients are checked according to their subcategories after releasing the nurse. If the subcategory is 31, 32, 34, 35, 36, the duration of tests is 105 minutes. If the subcategory is 33, the duration is 120 minutes. After waiting for tests, the patients wait for the nurse to check with a duration of NORM(3.5,0.17), and then release the nurse.

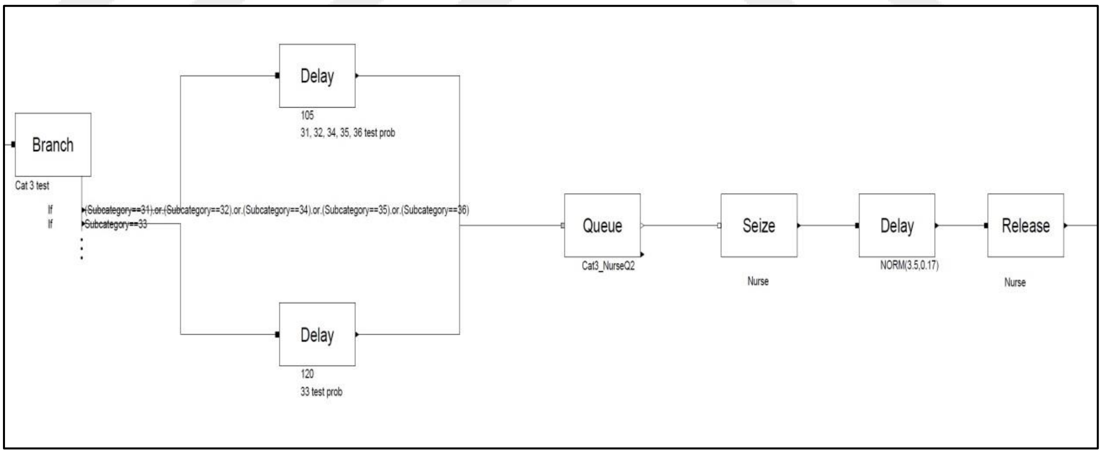


Figure 4.15 Blocks of the simulation model-11

As denoted in Figure 4.16, patients wait for the second doctor check in the observation bed and the duration of this control is derived from NORM(6,0.33). The patients release both the observation bed and the doctor lastly, then leave the system after passing TALLY blocks given in Figure 4.17 in where the average overall flow times of patients and times spent of patient depending on categories (length of stays) are recorded.

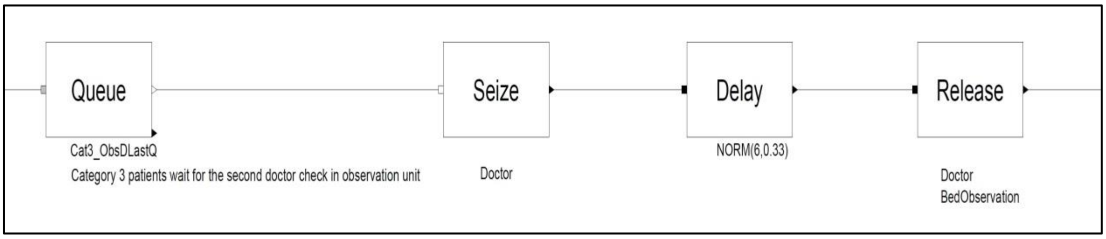


Figure 4.16 Blocks of the simulation model-12

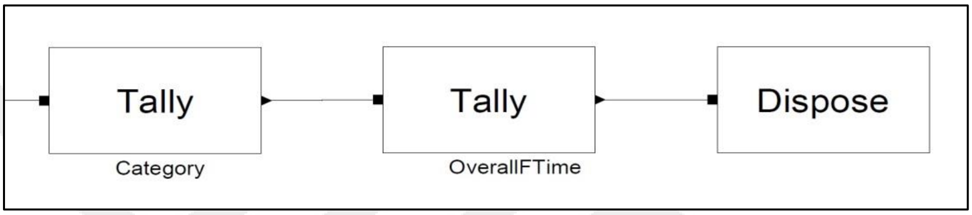


Figure 4.17 Blocks of the simulation model-13

4.2.6 Verification and Validation of the Simulation Model

> In the context of simulation modeling, the process of determining the correctness of the model operation typically consists of two separate functions: verification and validation. Verification is the process of determining that a model operates as intended. …. In contrast, validation is the process of reaching an acceptable level of confidence that the inferences drawn from the model are correct and applicable to the real-world system. (Pegden, Shannon & Sadowski, 1990, p.133).

Verification techniques such as developing a prototype model, using interactive debuggers, substituting constants for random variables, manually checking and tracing the model are used for verifying the simulation model.

The simulation model is built gradually in a step-by-step approach. First, a general model for all category types of patients is modeled. Then, a detailed model which includes all details for each patient category is created.

Interactive debugger and trace tools of Arena simulation modeling software, such as Command, Break, Trace and Watch are used for checking the model status.

Patient arrivals, service times, test times and observation times are random variables. Constants are substituted for random variables and the model is checked if the patient flows in the system as intended.

For validation of the model, the comparison of the model outputs of performance measures with the real system values is done. There are three performance measures in the system; the average overall flow time of patient, the average doctor utilization and the average nurse utilization. There is no recorded data about utilization rates but the average overall flow time of all patients, the average flow time of *Category 1*, *Category 2* and *Category 3* patient values can be used for validation. First confidence intervals on the average flow time of each category are constructed and depicted in Figure 4.18. Then, these confidence intervals are shown the experts working in the system. The intervals are confirmed and validated by the *ED* head doctor, the other doctors and the nurses.

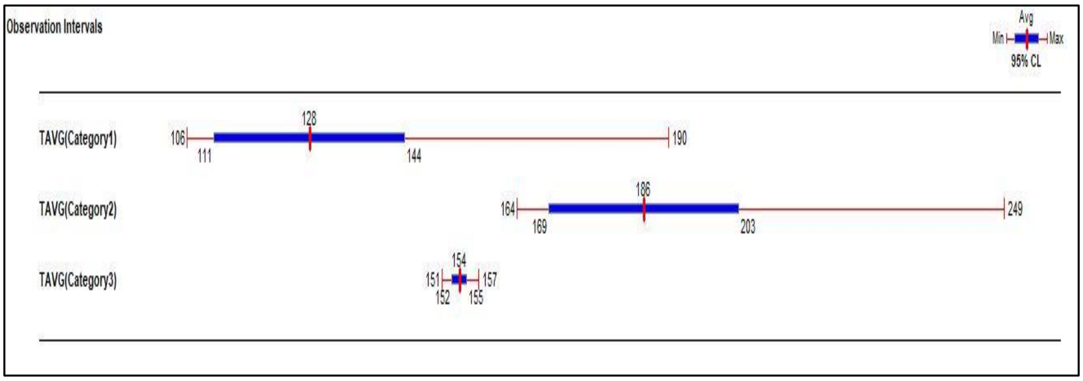


Figure 4.18 Confidence intervals of patient flow times

For the current *ED* system, the values of the performance measures obtained by 10 replications are given in Table 4.6.

Table 4.6 Current system performance measures

Performance Measures	Current System Values
Average overall flow time	152.79 min
Average doctor utilization	69.991 %
Average nurse utilization	45.318 %

4.3 Response Surface Methodology Study

MINITAB software is used for building two-level full factorial (2^k) designs, constructing normal probability plot of residuals, residuals versus predicted response plot and autocorrelation diagram, analyzing variance, obtaining first order regression models, building Central Composite, Nested Faced Center, Box-Behnken, Nested Box-Behnken Designs, obtaining second order regression models and running Derringer-Suich multi-response optimization procedure.

As mentioned before, the aim of the study is to find the best allocation of resources (input factors) minimizing the patient's average overall flow time spent in the *ED* (first response) with the requirement of 60% utilization rate of the doctors (second response) and 55% utilization rate of the nurses (third response).

The general followed steps in this section are the steps used in Yalçınkaya (2004).

4.4.1 Estimation Process

In this section the *phase zero* and *phase one* of the RSM is depicted. In *phase zero*, the objective of factor screening is to reduce the list of candidate variables to a relatively few, so that subsequent experiments will be more efficient and require fewer runs or tests. A screening experiment is referred as *phase zero* of a response surface study. After identifying important independent variables *phase one* of the response surface study begins. In phase one, the objective is to determine if the current levels or settings of the independent variables result in a value of the response that is near the optimum, or if the process is operating in some other region that is (possibly) remote from the optimum (Myers & Montgomery, 2016, p.7-8).

4.4.1.1 Phase Zero

Generally, at phase zero, a screening experiment is made for searching potential input factors, which are thought to be important in the response surface study, and for determining important factors. The screening experiment part is skipped because the input factors are the number of doctors, nurses, triage beds and observation beds and none of them will be eliminated.

The input factors are, respectively;
A: The number of doctors in the *ED*
B: The number of nurses in the *ED*
C: The number of triage beds in the *ED*
D: The number of observation beds in the *ED*

The responses are, respectively;
Y_1: The average overall flow time of patients spent in the *ED* (in minutes)
Y_2: The average utilization rate of doctors (in percentage)
Y_3: The average utilization rate of nurses (in percentage)

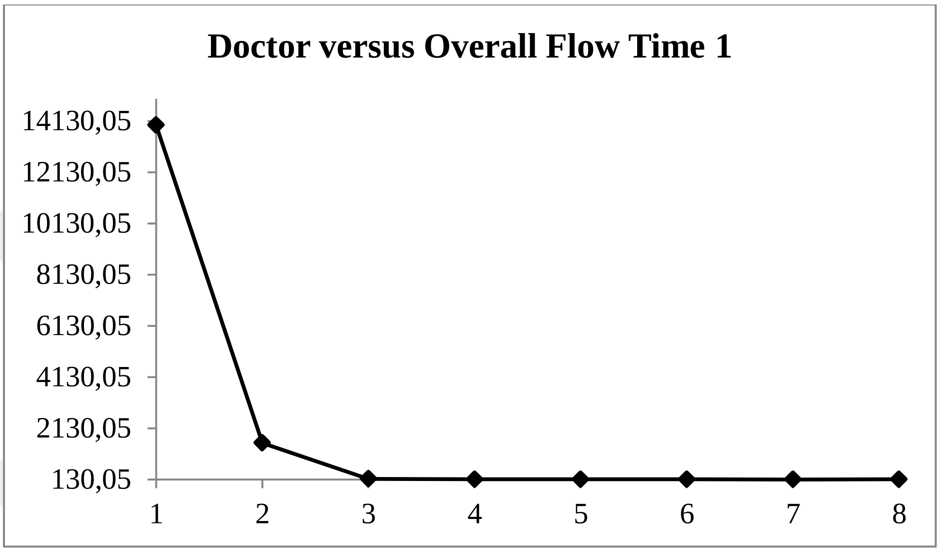
4.4.1.2 Phase One

To determine low and high levels of input factors, one-factor-at-a-time (*OFAT*) which is a strategy of experimentation is applied.

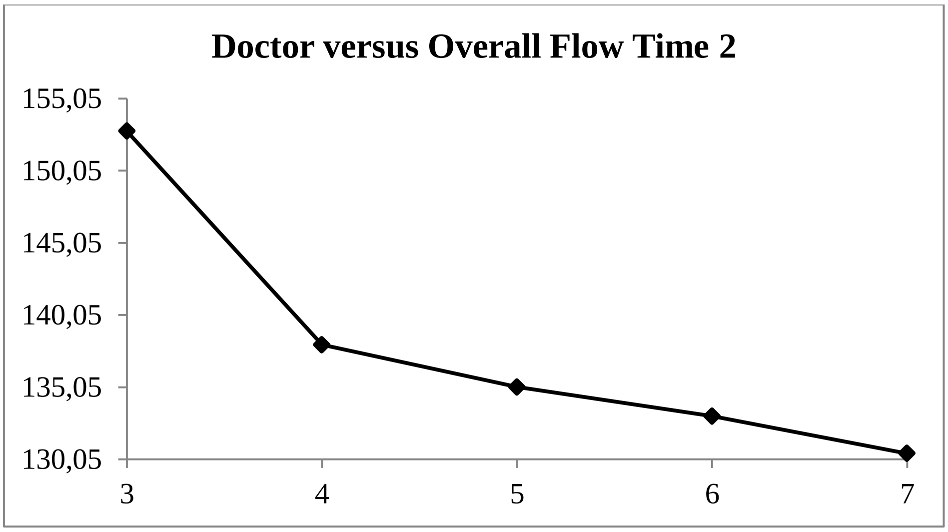
> The *OFAT* method consists of selecting a starting point, or baseline set of levels, for each factor, and then successively varying each factor over its range with the other factors held constant at the baseline level. After all tests are performed, a series of graphs are usually constructed showing how the response variable is

affected by varying each factor with all other factors held constant. (Montgomery, 2017, p.4).

To identify the input variables ranges, pilot simulation runs are conducted on the simulation model. Results of *OFAT* pilot runs are shown in Figures 4.19, 4.20, 4.21 and 4.22 where overall flow time values are obtained by taking average of 10 replications at each factor values combination.

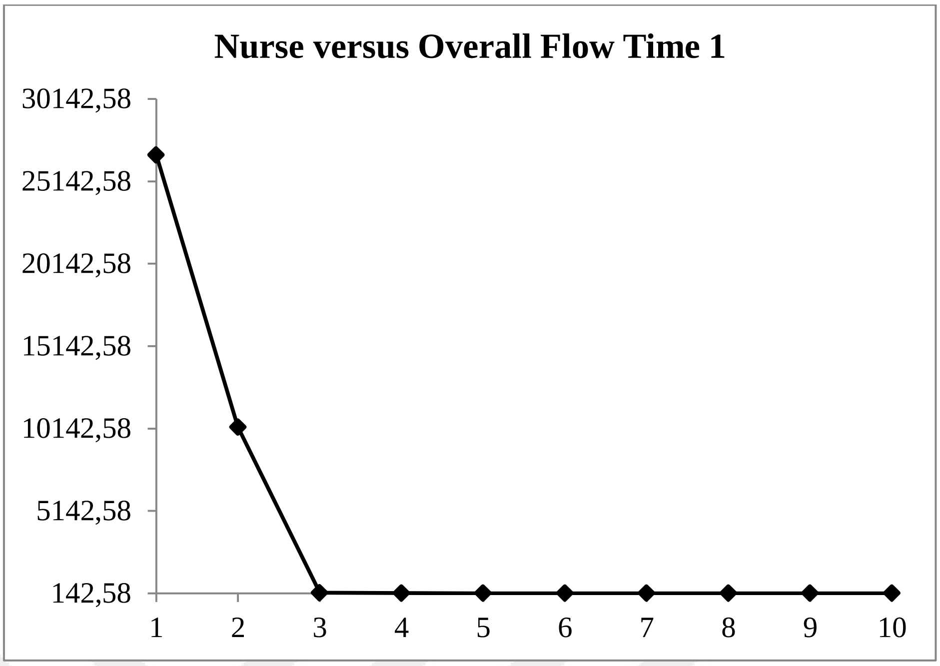


(a) The number of doctors is between 1 and 8

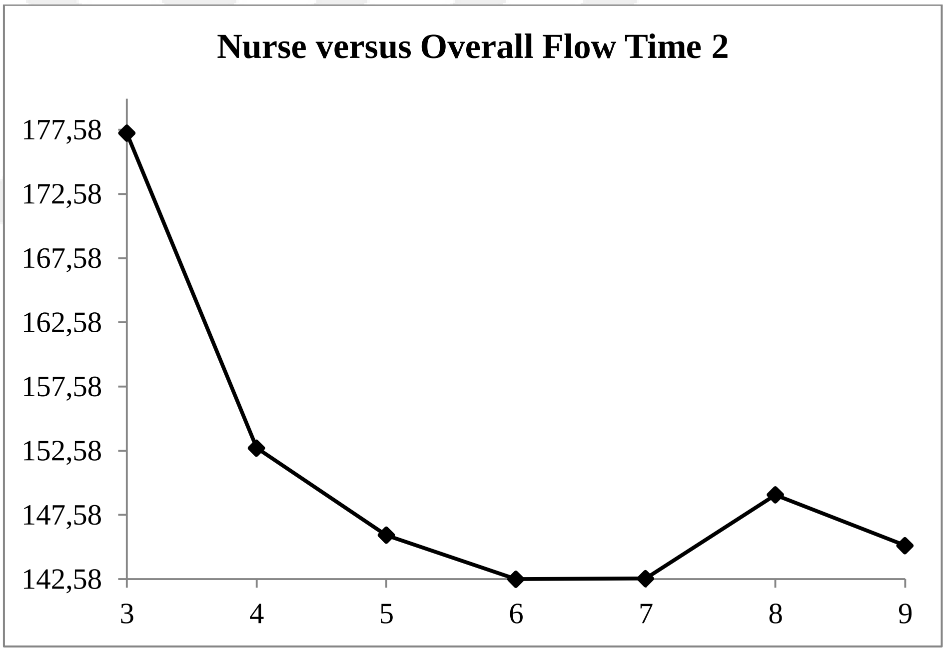


(b) The number of doctors is between 3 and 7

Figure 4.19 OFAT approach of doctor versus overall flow time

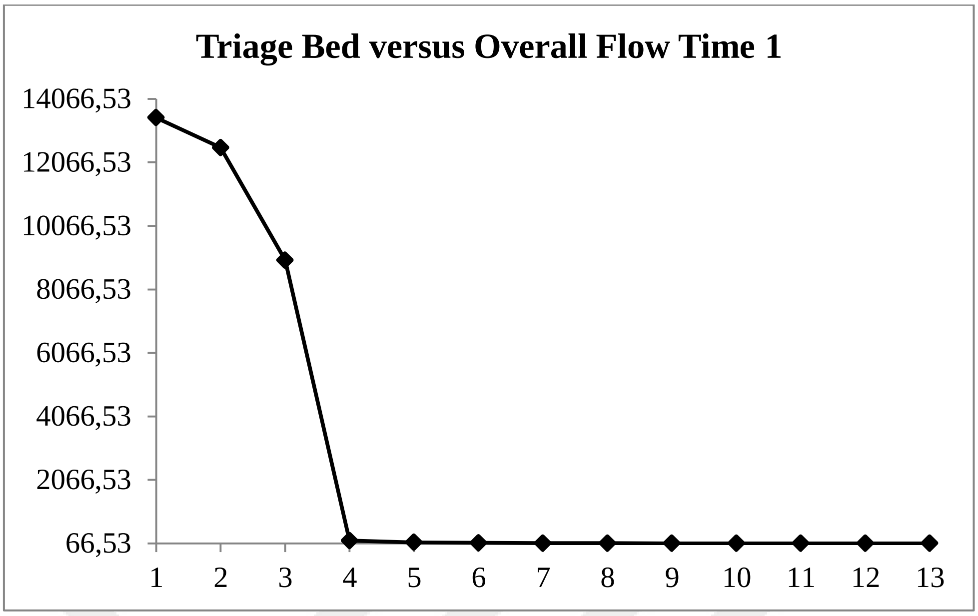


(a) The number of nurses is between 1 and 10

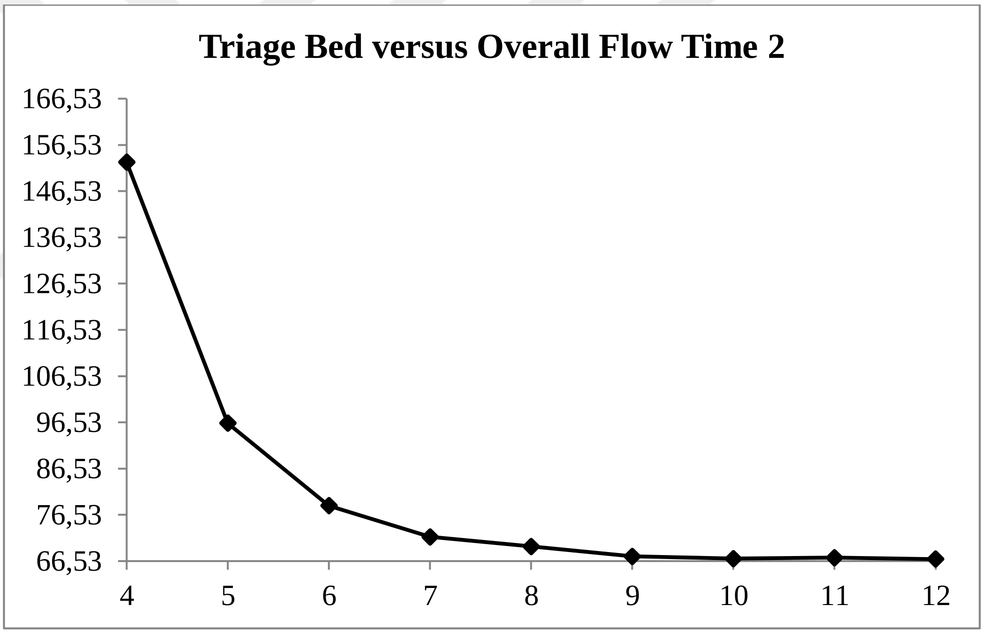


(b) The number of nurses is between 3 and 9

Figure 4.20 OFAT approach of nurse versus overall flow time

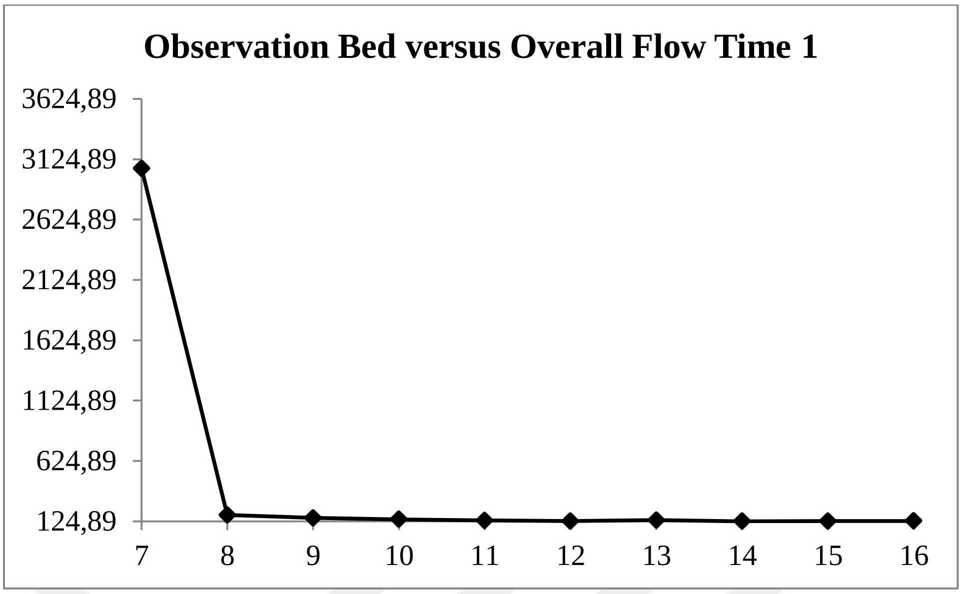


(a) The number of triage beds is between 1 and 13

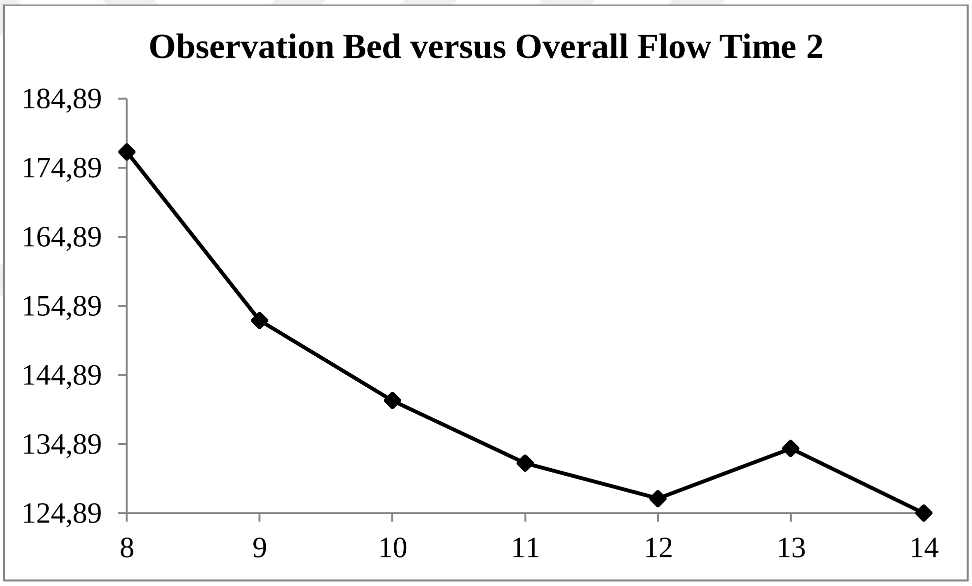


(b) The number of triage beds is between 4 and 12

Figure 4.21 OFAT approach of triage bed versus overall flow time



(a) The number of observation beds is between 7 and 16



(b) The number of observation beds is between 8 and 14

Figure 4.22 OFAT approach of observation bed versus overall flow time

After the OFAT study, the low and high levels of input factors are decided to be as given in Table 4.7. Coded values are found by using formula (2.4), as an example of the low-level of *A* is calculated as;

$$A = \frac{3 - [7 + 3]/2}{[7 - 3]/2} = -1$$

Table 4.7 Low and high levels of input factors

Levels	Doctor	Nurse	Triage Bed	Observation Bed
-1	3	3	4	8
0	5	6	8	11
1	7	9	12	14

The first order regression models with two-factor interactions are assumed to be;

$$Y_1 = \alpha_0 + \alpha_1 A + \alpha_2 B + \alpha_3 C + \alpha_4 D + \alpha_{12} AB + \alpha_{13} AC + \alpha_{14} AD + \alpha_{23} BC + \alpha_{24} BD + \alpha_{34} CD + \varepsilon$$

$$Y_2 = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_{12} AB + \beta_{13} AC + \beta_{14} AD + \beta_{23} BC + \beta_{24} BD + \beta_{34} CD + \varepsilon$$

$$Y_3 = \delta_0 + \delta_1 A + \delta_2 B + \delta_3 C + \delta_4 D + \delta_{12} AB + \delta_{13} AC + \delta_{14} AD + \delta_{23} BC + \delta_{24} BD + \delta_{34} CD + \varepsilon \quad (4.1)$$

where α_0, β_0 and δ_0 are constants, α_1, α_2, α_3, α_4, β_1, β_2, β_3, β_4, δ_1, δ_2, δ_3 and δ_4 are coefficients corresponding to main effects, and α_{12}, α_{13}, α_{14}, α_{23}, α_{24}, α_{34}, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34}, δ_{12}, δ_{13}, δ_{14}, δ_{23}, δ_{24} and δ_{34} are coefficients corresponding to two-factor interaction effects and ε is statistical error that has a normal distribution with mean zero and variance σ^2.

4.4.1.2.1 Two-level Full Factorial Design. For fitting a first order regression model, two-level full factorial design (2^4) with central runs is designed and then the simulation model is run 10 times at each design point, and response values for each design points are found.

The values for two-level full factorial design points are the average of 10 replications, but the values of central points are obtained by one replication. All values are shown in Table 4.8.

Table 4.8 Simulation results (2^4 design with 3 central runs)

	A	*B*	*C*	*D*	Y_1	Y_2	Y_3
1	-1	-1	-1	-1	208.445	77.233	60.298
2	1	-1	-1	-1	174.013	67.056	61.060
3	-1	1	-1	-1	163.037	72.741	20.185
4	1	1	-1	-1	140.221	60.760	20.141
5	-1	-1	1	-1	91.063	35.276	60.511
6	1	-1	1	-1	92.790	20.260	60.763
7	-1	1	1	-1	69.706	32.709	20.386
8	1	1	1	-1	68.933	15.193	20.055
9	-1	-1	-1	1	150.333	69.322	60.755
10	1	-1	-1	1	136.120	57.582	60.402
11	-1	1	-1	1	121.368	65.688	20.385
12	1	1	-1	1	110.981	52.344	20.116
13	-1	-1	1	1	76.353	31.725	60.552
14	1	-1	1	1	76.515	14.095	60.384
15	-1	1	1	1	57.861	30.998	20.072
16	1	1	1	1	57.641	13.504	20.311
17	0	0	0	0	60.421	21.465	30.187
18	0	0	0	0	58.366	20.544	30.251
19	0	0	0	0	61.720	22.348	30.708

For verifying that the least squares regression assumptions are not violated, the residuals from the least squares fit are checked to see whether they are normally, identically and independently distributed with zero mean and constant variance.

The residuals are shown in Table 4.9, are calculated by the formula;

$$e_i = Y_i - \hat{Y}_i \tag{4.2}$$

where $\hat{Y}_i$ is the predicted response.

Table 4.9 Residuals for Y_1 response

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
5.301	-3.145	-1.093	-1.063	-4.592	2.436	0.384	1.772	-3.010	0.854
11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	
-1.197	3.353	2.301	-0.145	1.906	-4.062	0.252	-1.803	1.551	

A normal probability plot of residuals is used to check normality in this study. The normal probability plot of residuals for Y_1 response is shown in Figure 4.23. As it is seen, the residuals plot is approximately a straight line. Therefore, it is accepted that the normality assumption is satisfied.

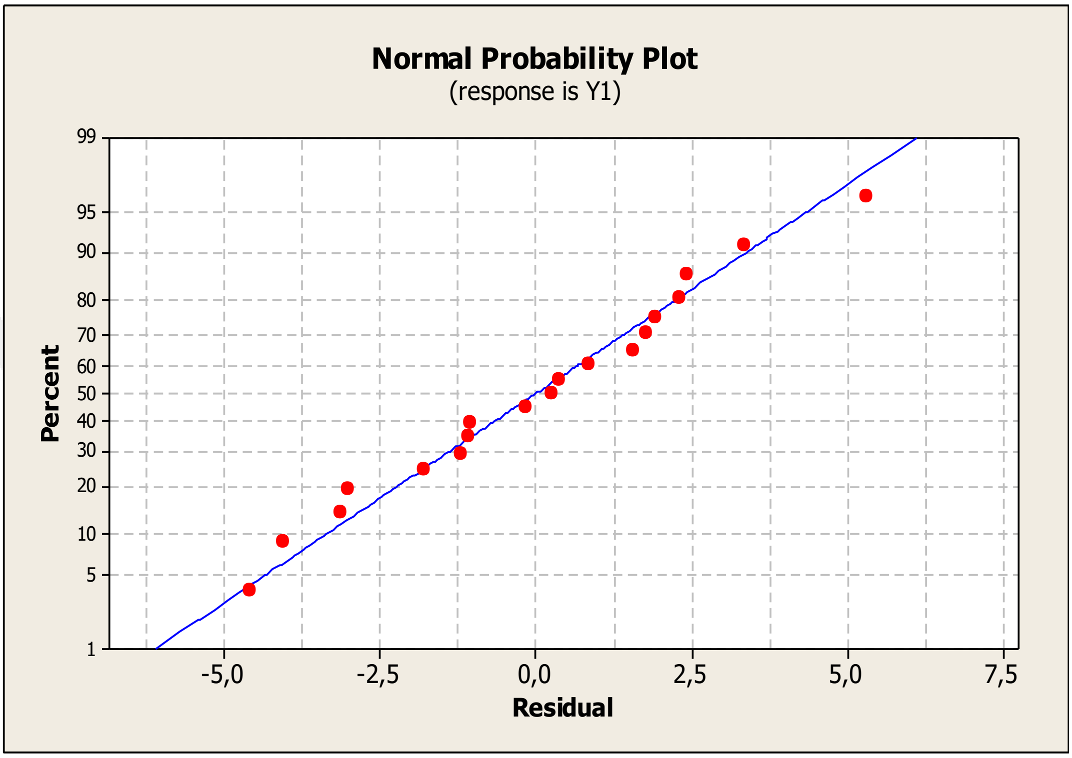


Figure 4.23 Normal probability plot of the residuals for Y_1 response

The plot of the residuals versus predicted response is also demonstrated in Figure 4.24. As it is seen, the residuals scatter randomly on the plot that indicated the variance is constant for all values of the response.

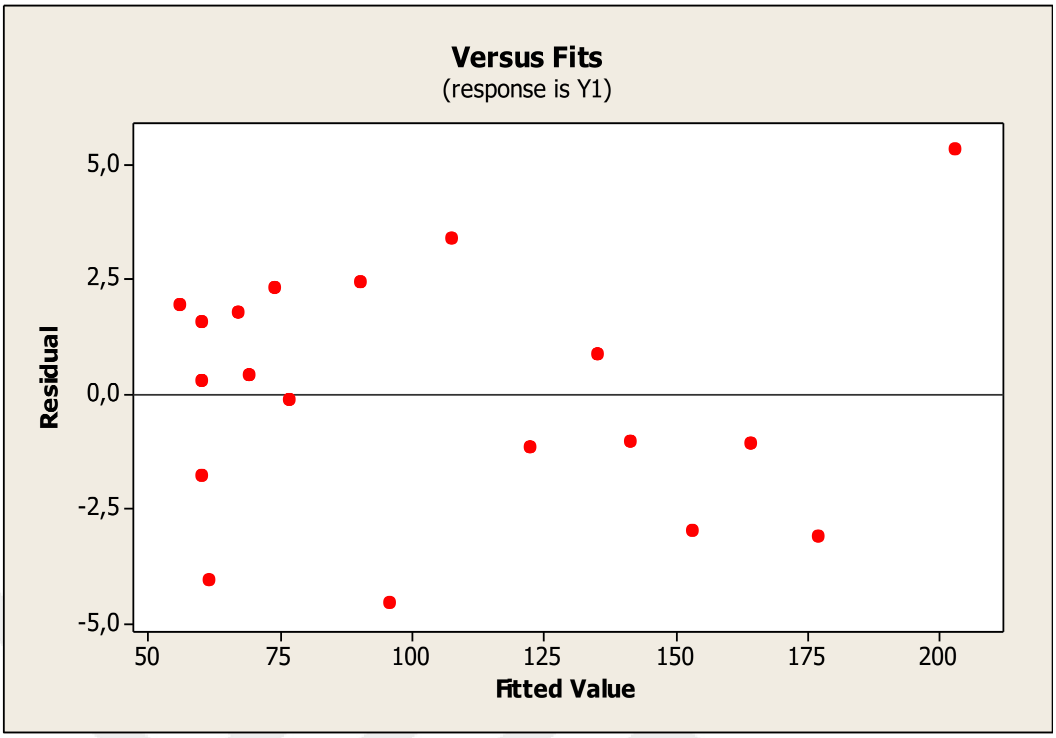


Figure 4.24 Residuals versus predicted response plot for Y_1 response

The assumption that the residuals are independent is controlled by Durbin-Watson test. "The Durbin-Watson test checks for a sequential dependence in which each error (and so residual) is correlated with those before and after it in the sequence" (Draper & Smith, 1998). The null and alternative hypotheses for positive correlation are respectively;

H_0 : *The error terms are not auto-correlated*

H_1 : *The error terms are positively auto-correlated* (4.3)

To test the H_0, the value of the test statistic d is calculated by the formula;

$$d = \frac{\left[\sum_{t=2}^{n} [e_t - e_{t-1}]^2\right]}{\sum_{t=1}^{n} e_t^{\,2}} \tag{4.4}$$

where e_t are residuals, n is the number of residuals and k is the number of independent variables.

Rejection region for positive correlation is shown as below;

i. If *$d<dL$, α, We reject H_0*

ii. If *$d>dU$, α, We do no reject H_0*

iii. If *$dL<d< dU$, α, The test is inconclusive*

α is the significance level and *dL*, *dU* are the upper and lower critical values obtained from Durbin-Watson significant tables for *n* and *k* values. For *α*=1 percent, the values of upper and lower critical values are obtained as *dL*=0.650 and *dU=1.583* (Savin & White, 1977).

According to the residuals for Y_1 response, the *d* statistic is calculated as 2.3116. After checking the rejection regions, we do not reject the null hypothesis for positive correlation. Therefore, the error terms are not positively auto-correlated.

The null and alternative hypotheses for negative correlation are respectively;

H_0 *: The error terms are not auto-correlated*

H_1 *: The error terms are negatively auto-correlated* (4.5)

Rejection region for negative correlation is as shown below;

i. If *$(4- d)<dL$, α, We reject H_0*

ii. If *$(4- d)>dU$, α, We do no reject H_0*

iii. If *dL, α $<(4- d)< dU$, α, The test is inconclusive*

After checking the rejection regions, we do not reject the null hypothesis for negative correlation. Therefore, the error terms are not negatively auto-correlated.

For the zero mean assumption a hypothesis test on the mean of the residuals with unknown variance is built. The null and alternative hypotheses are respectively;

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$ (4.6)

To test the $H_0 : \mu = \mu_0$, the value of the test statistic t_0 is calculated by the formula;

$$t_0 = \frac{\overline{X} - \mu_0}{s / \sqrt{n}} \tag{4.7}$$

where $\overline{X}$ is the sample mean, s is the sample standard deviation, and n is the sample size. H_0 is rejected if $t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$, where $t_{\alpha/2,n-1}$ and $-t_{\alpha/2,n-1}$ are the upper and lower 100α/2 percentage points of the t with $n-1$ degrees of freedom.

The values for the calculation are $\overline{X} = 0$, $\mu_0 = 0$, $s = 2.627391$, and $n = 19$, and with 5% ($\alpha = 0.05$) significance level $t_{0.025,\ 18} = 2{,}10092$, $-t_{0.025,\ 18} = -2.10092$. Substituting the values in formula (4.4) t_0 is found as 0.

Since t_0 is between -2.0294 and 2.0294 we do not reject that the mean of the residuals is zero hypothesis.

It is verified that the residuals are normally, identically and independently distributed with zero mean and constant variance for Y_2 and Y_3 response also. Next, Variance Analysis is used to see if the main effects, two-factor interactions, curvature and lack-of-fit are statistically significant. Table 4.10, Table 4.11 and Table 4.12 are the Analysis of Variance (ANOVA) tables for the response Y_1, Y_2 and Y_3, respectively. Additionally, terms and coefficients for three responses are given in Table 4.13.

In Table 4.11, *DF* denotes degrees of freedom, *Seq SS*; sequential sum of squares, *Adj SS*; adjusted sum of squares, *Adj MS*; adjusted mean square, *F*; the statistic F, *P*;

the *P*-value, *S*; standard deviation, *R-sq*, *R-sq(pred)* and *R-sq(adj)*; adjusted coefficient of determination.

S is measured in the units of the response variable and represents the standard deviation of how far the data values fall from the fitted values. The lower the value of *S*, the better the model describes the response. However, a low *S* value by itself does not indicate that the model meets the model assumptions. (Minitab Express Support, 2016).

R^2 is the percentage of variation in the response that is explained by the model. The higher the R^2 value, the better the model fits your data. The adjusted R^2 value incorporates the number of predictors in the model to help you choose the correct model. Predicted R^2 is used to determine how well your model predicts the response for new observations. (Minitab Express Support, 2016).

Table 4.10 ANOVA table for Y_1 response

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	29911.8	29911.8	7478	421.29	0.000
2-Way Interactions	6	1524.1	1524.1	254	14.31	0.001
Curvature	1	6842.3	6842.3	6842.3	385.48	0.000
Residual Error	7	124.2	124.2	17.7		
Lack of Fit	5	118.5	118.5	23.7	8.29	0.111
Pure Error	2	5.7	5.7	2.9		
Total	18	38402.4				
S = 4.21307 R-Sq = 99.68% R-Sq(pred) = 96.81% R-Sq(adj) = 99.17%						

The null and alternative hypotheses related to main effects (4.8), two-way interactions (4.9), curvature (4.10) and lack-of-fit (4.11) for variance analysis are shown below.

$$H_0 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$
$$H_1 = \alpha_i \neq 0 \text{ for at least one } i \ (i=1,2,3,4) \qquad (4.8)$$

$H_0 = \alpha_{12} = \alpha_{13} = \alpha_{14} = \alpha_{23} = \alpha_{24} = \alpha_{34} = 0$

$H_1 = \alpha_{ij} \neq 0$ *for at least one ij (i=1, 2, 3 and j=2, 3, 4)* (4.9)

$H_0 = \alpha_{11} = \alpha_{22} = \alpha_{33} = \alpha_{44} = 0$

$H_1 = \alpha_{ii} \neq 0$ *for at least one i (i=1,2,3,4)* (4.10)

H_0 = *The regression model is correct*

H_1 = *The regression model is not correct* (4.11)

where α_1, α_2, α_3, α_4 are coefficients corresponding to main effects, and α_{12}, α_{13}, α_{14}, α_{23}, α_{24}, α_{34} are coefficients corresponding to two-factor interaction effects, and α_{11}, α_{22}, α_{33}, α_{44} are coefficients corresponding to square effects.

In Table 4.10, since the *P*-value for the lack-of-fit test is greater than the significance level (0.05), the lack-of-fit is not statistically significant [H_0 (4.11) is not rejected], that is the first order regression model for Y_1 response is adequate. The main effects [H_0 (4.8) is rejected], two-way interactions [H_0 (4.9) is rejected], and curvature [H_0 (4.10) is rejected] are statistically significant (P-value<0.05).

Table 4.11 shows that the lack-of-fit is not statistically significant (P>0.05). The main effects, two-way interactions, and curvature are statistically significant (P< 0.05). The first order regression model is adequate for Y_2 response.

Table 4.11 ANOVA table for Y_2 response

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Main Effects	4	7772.08	7772.08	1943.02	3593.2	0.000
2-Way Interactions	6	65.79	65.79	10.96	20.28	0.000
Curvature	1	1374.81	1374.81	1374.81	2542.42	0.000
Residual Error	7	3.79	3.79	0.54		
Lack of Fit	5	2.16	2.16	0.43	0.53	0.755
Pure Error	2	1.63	1.63	0.81		
Total	18	9216.46				
S = 0.735356 R-Sq = 99.96% R-Sq(pred) = 99.72% R-Sq(adj) = 99.89%						

Table 4.12 shows that the lack-of-fit and two-way interactions are not statistically significant (P>0.05). The main effects and curvature are statistically significant (P< 0.05). The first order regression model is adequate for Y_3 response. Terms and coefficients for three responses are given in Table 4.13.

Table 4.12 ANOVA table for Y_3 response

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	6523.58	6523.58	1630.9	21131.9	0.000
2-Way Interactions	6	0.18	0.18	0.03	0.39	0.864
Curvature	1	253.47	253.47	253.47	3284.23	0.000
Residual Error	7	0.54	0.54	0.08		
Lack of Fit	5	0.38	0.38	0.08	0.94	0.588
Pure Error	2	0.16	0.16	0.08		
Total	18	6777.77				
S = 0.277807 R-Sq = 99.99% R-Sq(pred) = 99.94% R-Sq(adj) = 99.98%						

Table 4.13 Term and coefficient table for Y_1, Y_2 and Y_3 responses

Term	Coef for Y_1	Coef for Y_2	Coef for Y_3
Constant	112.21	44.78	40.40
A	-5.06	-7.18	0.01
B	-13.49	-1.79	-20.19
C	-38.35	-20.56	-0.02
D	-13.81	-2.87	-0.03
A*B	0.79	-0.36	-0.06
A*C	5.17	-1.28	-0.01
A*D	1.98	-0.34	-0.07
B*C	3.17	0.67	0.02
B*D	2.06	0.51	0.04
C*D	7.05	1.23	-0.02

As a result, the fitted first order models for three responses are as below;

$$Y_1 = 121.21 - 5.06\,A - 13.49\,B - 38.35\,C - 13.81\,D + 0.79\,AB + 5.17\,AC + 1.98\,AD + 3.17\,BC + 2.06\,BD - 7.05\,CD$$

$$Y_2 = 44.78 - 7.18\,A - 1.79\,B - 20.56\,C - 2.87\,D - 0.36\,AB - 1.28\,AC - 0.34\,AD + 0.67\,BC + 0.51\,BD + 1.23\,CD$$

$$Y_3 = 40.40 + 0.01\,A - 20.19\,B - 0.02\,C - 0.03\,D - 0.06\,AB - 0.01\,AC - 0.07\,AD + 0.02\,BC + 0.04\,BD - 0.02\,CD \tag{4.12}$$

Because the true response surface usually exhibits curvature near the optimum, it is understood that the determined region of experimentation is near the region of optimum (Yalçınkaya, 2004). Next, the second phase of *RSM* begins to build second order models.

4.4.2 Optimization Process

In this section, phase two of the *RSM* and the development, verification and validation of metamodels are depicted.

4.4.2.1 Phase Two

Phase two of a response surface study begins when the process is near the optimum. At this point the experimenter usually wants a model that will accurately approximate the true response function within a relatively small region around the optimum. Because the true response surface usually exhibits curvature near the optimum, a second order model (or perhaps some higher order polynomial) will be used (Myers & Montgomery, 2016, p.7-8).

4.4.2.1.1 Development, Verification and Validation of Metamodels. The designs for building metamodels and the verification and validation of these metamodels are depicted in detail in this subsection.

A. Central Composite Design (CCD)

The second order regression models are as below;

$$Y_1 = \alpha_0 + \alpha_1 A + \alpha_2 B + \alpha_3 C + \alpha_4 D + \alpha_{11} A^2 + \alpha_{22} B^2 + \alpha_{33} C^2 + \alpha_{44} D^2 + \alpha_{12} AB + \alpha_{13} AC + \alpha_{14} AD + \alpha_{23} BC + \alpha_{24} BD + \alpha_{34} CD + \varepsilon$$

$$Y_2 = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_{11} A^2 + \beta_{22} B^2 + \beta_{33} C^2 + \beta_{44} D^2 + \beta_{12} AB + \beta_{13} AC + \beta_{14} AD + \beta_{23} BC + \beta_{24} BD + \beta_{34} CD + \varepsilon$$

$$Y_3 = \delta_0 + \delta_1 A + \delta_2 B + \delta_3 C + \delta_4 D + \delta_{11} A^2 + \delta_{22} B^2 + \delta_{33} C^2 + \delta_{44} D^2 + \delta_{12} AB + \delta_{13} AC + \delta_{14} AD + \delta_{23} BC + \delta_{24} BD + \delta_{34} CD + \varepsilon \quad (4.13)$$

where α_0, β_0 and δ_0 are constant, and α_1, α_2, α_3, α_4, β_1, β_2, β_3, β_4, δ_1, δ_2, δ_3 and δ_4 are coefficients corresponding to main effects, and α_{12}, α_{13}, α_{14}, α_{23}, α_{24}, α_{34}, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34}, δ_{12}, δ_{13}, δ_{14}, δ_{23}, δ_{24} and δ_{34} are coefficients corresponding to two-factor interaction effects, α_{11}, α_{22}, α_{33}, α_{44}, β_{11}, β_{22}, β_{33}, β_{44}, δ_{11}, δ_{22}, δ_{33} and δ_{44} are coefficients corresponding to square effects, and ε is statistical error that has a normal distribution with mean zero and variance σ^2.

A *CCD* is built by augmenting the two-level full factorial design with central runs and axial runs (α=1, means design is face-centered) for fitting second order models. In Table 4.14, the *CCD* which is used for fitting second order models for three responses is given. The values for two-level full factorial design points and axial points are the average of 10 replications, and the values related to the central points are obtained by one replication.

Table 4.14 Simulation results for CCD

	A	***B***	***C***	***D***	Y_1	Y_2	Y_3		***A***	***B***	***C***	***D***	Y_1	Y_2	Y_3
1	-1	-1	-1	-1	208.445	77.233	60.298	15	-1	1	1	1	57.861	30.998	20.072
2	1	-1	-1	-1	174.013	67.056	61.060	16	1	1	1	1	57.641	13.504	20.311
3	-1	1	-1	-1	163.037	72.741	20.185	17	-1	0	0	0	60.669	33.377	30.457
4	1	1	-1	-1	140.221	60.760	20.141	18	1	0	0	0	60.247	15.952	30.564
5	-1	-1	1	-1	91.063	35.276	60.511	19	0	-1	0	0	80.878	25.611	61.139
6	1	-1	1	-1	92.790	20.260	60.763	20	0	1	0	0	59.753	20.845	20.010
7	-1	1	1	-1	69.706	32.709	20.386	21	0	0	-1	0	117.306	58.755	30.538
8	1	1	1	-1	68.933	15.193	20.055	22	0	0	1	0	58.997	18.802	30.250

Table 4.14 continues

9	-1	-1	-1	1	150.333	69.322	60.755	23	0	0	0	-1	72.649	26.493	30.178
10	1	-1	-1	1	136.120	57.582	60.402	24	0	0	0	1	58.571	20.348	30.187
11	-1	1	-1	1	121.368	65.688	20.385	25	0	0	0	0	60.421	21.465	30.187
12	1	1	-1	1	110.981	52.344	20.116	26	0	0	0	0	58.366	20.544	30.251
13	-1	-1	1	1	76.353	31.725	60.552	27	0	0	0	0	61.720	22.348	30.708
14	1	-1	1	1	76.515	14.095	60.384								

After verifying that the residuals of the second order regression models fitted for Y_1, Y_2 and Y_3 are normally, identically and independently distributed with zero mean and constant variance, the ANOVA tables are given in Table 4.15, Table 4.16 and 4.17. Additionally, terms and coefficients for three responses are given in Table 4.18.

Table 4.15 ANOVA table for Y_1 response for CCD

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Regression	14	46946.5	46946.5	3353.3	94.33	0.000
Linear	4	31645	31645	7911.2	222.54	0.000
Square	4	13777.5	13777.5	3444.4	96.89	0.000
Interaction	6	1524.1	1524.1	254	7.15	0.002
Residual Error	12	426.6	426.6	35.5		
Lack-of-Fit	10	420.9	420.9	42.1	14.72	0.065
Pure Error	2	5.7	5.7	2.9		
Total	26	47373.1				
S = 5.96233 R-Sq = 99.10% R-Sq(pred) = 94.95% R-Sq(adj) = 98.05%						

Table 4.16 ANOVA table for Y_2 response for CCD

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Regression	14	11564.8	11564.8	826.05	1049.78	0.000
Linear	4	8746.8	8746.8	2186.69	2778.93	0.000
Square	4	2752.2	2752.2	688.05	874.4	0.000
Interaction	6	65.8	65.8	10.96	13.93	0.000
Residual Error	12	9.4	9.4	0.79		
Lack-of-Fit	10	7.8	7.8	0.78	0.96	0.612
Pure Error	2	1.6	1.6	0.81		
Total	26	11574.2				
S = 0,887064 R-Sq = 99,92% R-Sq(pred) = 99,60% R-Sq(adj) = 99,82%						

Table 4.17 ANOVA table for Y_3 response for CCD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	7975.41	7975.41	569.67	8309.44	0.000
Linear	4	7369.14	7369.14	1842.29	26872.2	0.000
Square	4	606.09	606.09	151.52	2210.15	0.000
Interaction	6	0.18	0.18	0.03	0.44	0.839
Residual Error	12	0.82	0.82	0.07		
Lack-of-Fit	10	0.66	0.66	0.07	0.82	0.665
Pure Error	2	0.16	0.16	0.08		
Total	26	7976.23				
S = 0.261834 R-Sq = 99.99% R-Sq(pred) = 99.94% R-Sq(adj) = 99.98%						

In Tables 4.15 and 4.16, it is seen that linear, quadratic and two-way interaction effects are statistically significant (P<0.05), and the second order regression models are statistically significant (P>0.05) for Y_1 and Y_2 responses.

In Table 4.17, it is seen that linear, quadratic effects are statistically significant (P<0.05) and two-way interaction effects are statistically insignificant (P>0.05) for Y_3 response. The second order regression model is statistically significant (P>0.05) for Y_3 response.

Table 4.18 Term and coefficient table for Y_1, Y_2 and Y_3 responses

Term	Coef for Y_1	Coef for Y_2	Coef for Y_3
Constant	58.553	21.641	30.405
A	-4.521	-7.351	0.011
B	-13.167	-1.854	-20.234
C	-37.331	-20.496	-0.033
D	-13.062	-2.895	-0.023
A*A	2.714	2.930	0.094
B*B	12.571	1.493	10.158
C*C	30.407	17.044	-0.023
D*D	7.866	1.685	-0.234
A*B	0.785	-0.361	-0.056
A*C	5.172	-1.276	-0.007
A*D	1.977	-0.345	-0.075
B*C	3.170	0.669	0.019
B*D	2.059	0.515	0.041
C*D	7.050	1.234	-0.023

As a result, the second order regression models (metamodels) obtained with *CCD* for Y_1, Y_2 and Y_3 are respectively;

$Y_1 = 58.553 - 4.521A - 13.167B - 37.331C - 13.062D + 2.714A^2 + 12.571B^2 + 30.407C^2 + 7.866D^2 + 0.785AB + 5.172AC + 1.977AD + 3.170BC + 2.059BD + 7.050CD$

$Y_2 = 21.461 - 7.351A - 1.854B - 20.496C - 2.895D + 2.930A^2 + 1.493B^2 + 17.044C^2 + 1.685D^2 - 0.361AB - 1.276AC - 0.345AD + 0.669BC + 0.515BD + 1.234CD$

$Y_3 = 30.405 + 0.011A - 20.234B - 0.033C - 0.023D + 0.094A^2 + 10.158B^2 - 0.023C^2 - 0.234D^2 - 0.056AB - 0.007AC - 0.075AD + 0.019BC + 0.041BD - 0.023CD$ (4.14)

After building regression metamodels for three responses, they are needed to be verified and validated.

- *Verification and validation of metamodels obtained with CCD*

For verification, lack-of-fit tests are used for all three metamodels. As can be seen from Table 4.18, Table 4.19 and Table 4.20, the metamodels have no statistically significant lack-of-fit (P>0.05) for all responses with 5% significance level (α = 0.05). Additionally R^2 is 99.10%, and R^2*-adjusted* is 98.05% for response Y_1, and R^2 is 99.92%, and R^2*-adjusted* is 99.82% for response Y_2, and R^2 is 99.99%, and R^2*-adjusted* is 99.98% for response Y_3.

For validation, six points are selected randomly in experimentation region which are different from design points. These coded values of input factors are transformed to natural variables by using formula (2.4), as an example of transforming the coded value of -0.5 of B to natural values is calculated as;

$$Natural\ value\ of\ B = (-0.5) * \left[\frac{9-3}{2}\right] + \left[\frac{9+3}{2}\right] = 4.5$$

We round the fractional value 4.5 into the closest higher integer value 5 since the factors can only take integer values.

All these random points, simulation models results and metamodel results are shown in Table 4.19. The simulation model is run at these randomly selected points with 10 replications for each point. The values that are depicted at the Simulation Model column are the averages of these 10 replications. The values that are denoted Metamodel Column are predicted by fitting second order regression metamodels.

The Absolute Relative Error (*ARE*) is selected as a criterion, *ARE* value (r) is calculated by the following formula;

$$r = \left|(w - y)/w\right| \tag{4.15}$$

where simulation output is denoted by *w*, and metamodel output is denoted by *y* (Kleijnen & Sargent, 2000, p.20). In this study, the threshold (r_{max}) is accepted as 5%.

Table 4.19 Metamodels validation for CCD

	Input factors				**Simulation model**			**Metamodel**			**ARE (%)**		
	A	***B***	***C***	***D***	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
1	-0.5	-1	0	-1	111.60	34.14	60.58	96.69	36.48	61.10	15.42	6.41	0.86
2	0.25	0.5	0.75	1	47.13	13.50	22.52	57.51	15.60	22.60	18.05	13.45	0.31
3	-0.5	1	0	-1	80.36	29.76	20.09	71.39	30.43	20.24	12.57	2.21	0.75
4	-0.5	0	0.75	1	47.76	20.63	30.15	57.81	23.45	30.24	17.38	12.03	0.31
5	0.25	0	0	-1	78.03	24.65	30.22	72.86	24.30	30.50	7.08	1.46	0.91
6	-0.5	-1	0	1	79.38	27.66	60.53	79.25	27.92	60.52	0.16	0.91	0.02

Since not all the *ARE* values are smaller than 5%, the second order metamodels cannot be validated and will not be used for prediction.

Therefore, a new design is needed to be adopted. Another design which is named as Nested Faced Center Design (*NFCD*) will be used with the aim to obtain validated metamodels.

B. Nested Faced Center Design (NFCD)

Landman and Simpson et al. (2007) investigated a wind-tunnel testing of high performance aircraft. In an exploratory study, they developed a hybrid design called a “nested face-centered design”.

In this thesis, a four-factor *NFCD* by using the given example of two-factor *NFCD* in Yang (2008) is developed. A *NFCD* is built for fitting second order models. In Table 4.20, the *NFCD* which is used for three responses is given. The values related to the central points are obtained by one replication and the others are the average of 10 replications.

After verifying that the residuals of the second order regression models fitted for Y_1, Y_2 and Y_3 are normally, identically and independently distributed with zero mean and constant variance, the ANOVA tables are given in Table 4.21, Table 4.22 and 4.23. Additionally, terms and coefficients for three responses are given in Table 4.24.

Table 4.20 Simulation results for NFCD

	A	*B*	*C*	*D*	Y_1	Y_2	Y_3		*A*	*B*	*C*	*D*	Y_1	Y_2	Y_3
1	-1	-1	-1	-1	208.445	77.233	60.298	26	0.5	-0.5	-0.5	-0.5	68.882	27.222	36.192
2	1	-1	-1	-1	174.013	67.056	61.060	27	-0.5	0.5	-0.5	-0.5	69.563	34.727	22.778
3	-1	1	-1	-1	163.037	72.741	20.185	28	0.5	0.5	-0.5	-0.5	68.784	27.545	22.824
4	1	1	-1	-1	140.221	60.760	20.141	29	-0.5	-0.5	0.5	-0.5	60.978	24.285	36.311
5	-1	-1	1	-1	91.063	35.276	60.511	30	0.5	-0.5	0.5	-0.5	61.077	16.617	36.226
6	1	-1	1	-1	92.790	20.260	60.763	31	-0.5	0.5	0.5	-0.5	60.479	24.310	22.819
7	-1	1	1	-1	69.706	32.709	20.386	32	0.5	0.5	0.5	-0.5	60.721	16.676	22.765
8	1	1	1	-1	68.933	15.193	20.055	33	-0.5	-0.5	-0.5	0.5	64.694	32.074	36.308
9	-1	-1	-1	1	150.333	69.322	60.755	34	0.5	-0.5	-0.5	0.5	63.794	23.981	36.123
10	1	-1	-1	1	136.120	57.582	60.402	35	-0.5	0.5	-0.5	0.5	64.981	32.295	22.555
11	-1	1	-1	1	121.368	65.688	20.385	36	0.5	0.5	-0.5	0.5	64.525	25.277	22.808
12	1	1	-1	1	110.981	52.344	20.116	37	-0.5	-0.5	0.5	0.5	58.128	23.558	36.195
13	-1	-1	1	1	76.353	31.725	60.552	38	0.5	-0.5	0.5	0.5	58.383	15.846	36.172
14	1	-1	1	1	76.515	14.095	60.384	39	-0.5	0.5	0.5	0.5	57.919	23.837	22.807
15	-1	1	1	1	57.861	30.998	20.072	40	0.5	0.5	0.5	0.5	57.678	15.978	22.866
16	1	1	1	1	57.641	13.504	20.311	41	-0.5	0	0	0	60.098	25.609	30.192
17	-1	0	0	0	60.669	33.377	30.457	42	0.5	0	0	0	60.154	18.075	30.493
18	1	0	0	0	60.247	15.952	30.564	43	0	-0.5	0	0	60.509	21.197	36.513
19	0	-1	0	0	80.878	25.611	61.139	44	0	0.5	0	0	60.018	21.263	22.693
20	0	1	0	0	59.753	20.845	20.010	45	0	0	-0.5	0	65.962	28.984	30.327
21	0	0	-1	0	117.306	58.755	30.538	46	0	0	0.5	0	59.128	19.311	30.343
22	0	0	1	0	58.997	18.802	30.250	47	0	0	0	-0.5	62.007	22.012	30.306
23	0	0	0	-1	72.649	26.493	30.178	48	0	0	0	0.5	58.756	20.504	30.311
24	0	0	0	1	58.571	20.348	30.187	49	0	0	0	0	61.720	22.348	30.708
25	-0.5	-0.5	-0.5	-0.5	69.744	34.864	36.609	50	0	0	0	0	58.366	20.544	30.251

Table 4.21 ANOVA table for Y_1 response for NFCD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	54904.4	54904.4	3921.7	25.54	0.000
Linear	4	27735.8	27735.8	6934	45.16	0.000
Square	4	25702.2	25702.2	6425.5	41.85	0.000
Interaction	6	1466.4	1466.4	244.4	1.59	0.179
Residual Error	35	5373.7	5373.7	153.5		
Lack-of-Fit	34	5368.1	5368.1	157.9	28.07	0.149
Pure Error	1	5.6	5.6	5.6		
Total	49	60278.1				
S = 12,3909 R-Sq = 91,09% R-Sq(pred) = 79,33% R-Sq(adj) = 87,52%						

Table 4.22 ANOVA table for Y_2 response for NFCD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	13893.9	13893.9	992.42	68.42	0.000
Linear	4	8987.3	8987.3	2246.82	154.9	0.000
Square	4	4839.3	4839.3	1209.82	83.41	0.000
Interaction	6	67.3	67.3	11.22	0.77	0.596
Residual Error	35	507.7	507.7	14.51		
Lack-of-Fit	34	506.1	506.1	14.88	9.15	0.257
Pure Error	1	1.6	1.6	1.63		
Total	49	14401.6				
S = 3,80859 R-Sq = 96,47% R-Sq(pred) = 92,55% R-Sq(adj) = 95,06%						

Table 4.23 ANOVA table for Y_3 response for NFCD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	9206.12	9206.12	657.58	85.65	0.000
Linear	4	8030.47	8030.47	2007.62	261.5	0.000
Square	4	1175.51	1175.51	293.88	38.28	0.000
Interaction	6	0.14	0.14	0.02	0	1.000
Residual Error	35	268.71	268.71	7.68		
Lack-of-Fit	34	268.6	268.6	7.9	75.65	0.091
Pure Error	1	0.1	0.1	0.1		
Total	49	9474.82				
S = 2,77080 R-Sq = 97,16% R-Sq(pred) = 94,66% R-Sq(adj) = 96,03%						

In Table 4.21, Table 4.22 and Table 4.23, it is seen that linear, quadratic effects are statistically significant (P<0.05) and two-way interaction effects are statistically insignificant (P>0.05) for Y_1, Y_2 and Y_3 responses. The second order regression models are statistically significant (P>0.05) for Y_1, Y_2 and Y_3 responses.

Table 4.24 Term and coefficient table for Y_1, Y_2 and Y_3 responses for NFCD

Term	**Coef for Y_1**	**Coef for Y_2**	**Coef for Y_3**
Constant	53.566	20.033	28.369
A	-3.674	-7.400	0.006
B	-10.568	-1.433	-18.892
C	-31.342	-18.320	-0.027
D	-11.191	-2.648	-0.034
A*A	4.267	3.400	0.736
B*B	13.577	1.903	10.034
C*C	30.901	17.226	0.624
D*D	9.176	2.091	0.419
A*B	0.741	-0.319	-0.038
A*C	4.917	-1.215	-0.003
A*D	1.860	-0.333	-0.061
B*C	2.947	0.612	0.027
B*D	1.956	0.509	0.044
C*D	6.750	1.280	-0.014

As a result, the second order regression models (metamodels) obtained with *NFCD* for Y_1, Y_2 and Y_3 are respectively;

$$Y_1 = 53.566 - 3.674A - 10.568B - 31.342C - 11.191D + 4.267A^2 + 13.577B^2 + 30.901C^2 + 9.176D^2 + 0.741AB + 4.917AC + 1.860AD + 2.947BC + 1.956BD + 6.750CD$$

$$Y_2 = 20.033 - 7.400A - 1.433B - 18.320C - 2.648D + 3.400A^2 + 1.903B^2 + 17.226C^2 + 2.091D^2 - 0.319AB - 1.215AC - 0.333AD + 0.612BC + 0.509BD + 1.280CD$$

$Y_3 = 28.369 + 0.006A - 18.892B - 0.027C - 0.034D + 0.736A^2 + 10.034B^2 - 0.624C^2 - 0.419D^2 - 0.038AB - 0.003AC - 0.061AD + 0.027BC + 0.044BD - 0.014CD$ (4.17)

After building regression metamodels for three responses, they are needed to be verified and validated.

- *Verification and validation of metamodels obtained with NFCD*

For verification, lack-of-fit tests are used for all three metamodels. As can be seen from Table 4.27, Table 4.28 and Table 4.29, the metamodels have no statistically significant lack-of-fit (P>0.05) for both responses with 5% significance level (α = 0.05). And R^2 is 91.09%, and R^2*-adjusted* is 87.52% for response Y_1, and R^2 is 96.47%, and R^2*-adjusted* is 95.06% for response Y_2, and R^2 is 97.16%, and R^2*-adjusted* is 96.03% for response Y_3.

For validation, six points are selected randomly in experimentation region which are different from design points. All these points, simulation models results and metamodel results are shown in Table 4.25. The simulation model is run at these randomly selected points with 10 replications for each point. The values that are denoted at the Simulation Model column are the averages of these 10 replications. The values that are depicted Metamodel Column are predicted by fitting second order regression metamodels.

Table 4.25 Metamodels validation for NFCD

Coded values				**Metamodel**			**Simulation model**			**ARE (%)**		
A	***B***	***C***	***D***	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
-0.5	-1	0	-1	104.24	32.84	57.92	96.69	36.48	61.10	7.81	9.96	5.21
0.25	0.5	0.75	1	51.51	14.64	22.20	57.51	15.60	22.60	10.43	6.14	1.77
-0.5	1	0	-1	78.45	29.28	20.09	71.39	30.43	20.24	9.89	3.80	0.75
-0.5	0	0.75	1	50.62	21.56	29.29	57.81	23.45	30.24	12.43	8.07	3.16
0.25	0	0	-1	72.82	23.22	28.88	72.86	24.30	30.50	0.07	4.44	5.29
-0.5	-1	0	1	76.08	26.86	57.83	79.25	27.92	60.52	4.00	3.79	4.44

Since not all the *ARE* values are smaller than 5%, the second order metamodels cannot be validated and will not be used for prediction.

Therefore, again a new design is needed to be adopted. Another design, Box-Behnken Design (*BBD*) will be used to obtain validated metamodels.

C. Box-Behnken Design (BBD)

A *BBD* is built for fitting second order models. In Table 4.26, the *BBD* which is used for three responses is given. The values related to the central points are obtained by one replication and the others are the average of 10 replications.

After verifying that the residuals of the second order regression models fitted for Y_1, Y_2 and Y_3 are normally, identically and independently distributed with zero mean and constant variance, the ANOVA tables are given in Table 4.27, Table 4.28 and 4.29. Additionally, terms and coefficients for three responses are given in Table 4.30.

Table 4.26 Simulation results for BBD

	A	*B*	*C*	*D*	Y_1	Y_2	Y_3		*A*	*B*	*C*	*D*	Y_1	Y_2	Y_3
1	-1	-1	0	0	81.351	36.851	61.075	15	0	-1	1	0	77.571	20.487	60.550
2	1	-1	0	0	81.135	20.429	60.087	16	0	1	1	0	58.682	18.824	20.193
3	-1	1	0	0	60.829	33.400	20.314	17	-1	0	-1	0	130.260	67.216	30.507
4	1	1	0	0	59.904	15.726	20.253	18	1	0	-1	0	116.410	54.541	30.442
5	0	0	-1	-1	144.130	64.624	30.391	19	-1	0	1	0	59.423	31.131	30.167
6	0	0	1	-1	68.669	20.406	30.147	20	1	0	1	0	58.874	13.554	30.336
7	0	0	-1	1	113.200	57.377	30.358	21	0	-1	0	-1	95.684	32.856	61.005
8	0	0	1	1	57.410	18.702	30.245	22	0	1	0	-1	70.935	25.738	19.959
9	-1	0	0	-1	73.194	37.447	30.447	23	0	-1	0	1	77.183	22.844	60.063
10	1	0	0	-1	72.779	21.866	30.215	24	0	1	0	1	58.021	20.242	20.152
11	-1	0	0	1	58.987	32.557	30.207	25	0	0	0	0	60.421	21.465	30.187
12	1	0	0	1	58.325	14.986	30.394	26	0	0	0	0	58.366	20.544	30.251
13	0	-1	-1	0	146.380	62.952	60.478	27	0	0	0	0	61.720	22.348	30.708
14	0	1	-1	0	118.980	59.129	20.356								

Table 4.27 ANOVA table for Y_1 response for BBD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	20672.4	20672.4	1476.6	273.95	0.000
Linear	4	14938.1	14938.1	3734.5	692.86	0.000
Square	4	5567.2	5567.2	1391.8	258.22	0.000
Interaction	6	167	167	27.8	5.16	0.008
Residual Error	12	64.7	64.7	5.4		
Lack-of-Fit	10	59	59	5.9	2.06	0.371
Pure Error	2	5.7	5.7	2.9		
Total	26	20737.1				
S = 2,32165 R-Sq = 99,69% R-Sq(pred) = 98,30% R-Sq(adj) = 99,32%						

Table 4.28 ANOVA table for Y_2 response for BBD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	7535.24	7535.24	538.23	382.09	0.000
Linear	4	5857.06	5857.06	1464.27	1039.48	0.000
Square	4	1656.84	1656.84	414.21	294.05	0.000
Interaction	6	21.34	21.34	3.56	2.52	0.081
Residual Error	12	16.9	16.9	1.41		
Lack-of-Fit	10	15.28	15.28	1.53	1.88	0.397
Pure Error	2	1.63	1.63	0.81		
Total	26	7552.14				
S = 1,18686 R-Sq = 99,78% R-Sq(pred) = 98,79% R-Sq(adj) = 99,52%						

Table 4.29 ANOVA table for Y_3 response for BBD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	5554.5	5554.5	396.75	9400.53	0.000
Linear	4	4881.78	4881.78	1220.44	28917.05	0.000
Square	4	672.11	672.11	168.03	3981.2	0.000
Interaction	6	0.61	0.61	0.1	2.42	0.091
Residual Error	12	0.51	0.51	0.04		
Lack-of-Fit	10	0.34	0.34	0.03	0.43	0.853
Pure Error	2	0.16	0.16	0.08		
Total	26	5555				
S = 0,205439 R-Sq = 99,99% R-Sq(pred) = 99,96% R-Sq(adj) = 99,98%						

In Table 4.27, it is seen that linear, quadratic and two-way interaction effects are statistically significant ($P<0.05$), and the second order regression models are statistically significant ($P>0.05$) for Y_1 response.

In Table 4.28 and Table 4.29, it is seen that linear, quadratic effects are statistically significant ($P<0.05$) and two-way interaction effects are statistically insignificant ($P>0.05$) for Y_2 and Y_3 responses. The second order regression models are statistically significant ($P>0.05$) for Y_2 and Y_3 responses.

Table 4.30 Term and coefficient table for Y_1, Y_2 and Y_3 responses for BBD

Term	**Coef for Y_1**	**Coef for Y_2**	**Coef for Y_3**
Constant	60.169	21.452	30.382
A	-1.385	-8.125	-0.083
B	-10.996	-1.947	-20.169
C	-32.394	-20.228	-0.075
D	-8.522	-3.019	-0.062
A*A	0.586	3.242	0.017
B*B	9.984	1.963	10.022
C*C	30.401	16.897	-0.017
D*D	5.217	1.984	-0.091
A*B	-0.177	-0.313	0.232
A*C	3.325	-1.226	0.059
A*D	-0.062	-0.498	0.105
B*C	2.128	0.540	-0.059
B*D	1.397	1.129	0.284
C*D	4.918	1.386	0.033

As a result, the second order regression models (metamodels) obtained with *BBD* for Y_1, Y_2 and Y_3 are respectively;

$$Y_1 = 60.169 - 1.385A - 10.996B - 32.394C - 8.522D + 0.586A^2 + 9.984B^2 + 30.401C^2 + 5.217D^2 - 0.177AB + 3.325AC - 0.062AD + 2.128BC + 1.397BD + 4.918CD$$

$$Y_2 = 21.452 - 8.125A - 1.947B - 20.228C - 3.019D + 3.242A^2 + 1.963B^2 + 16.897C^2 + 1.984D^2 - 0.313AB - 1.226AC - 0.498AD + 0.540BC + 1.129BD + 1.386CD$$

Y_3 = $30.382 - 0.083A - 20.169B - 0.075C - 0.062D + 0.017A^2 + 10.022B^2 - 0.017C^2 - 0.091D^2 + 0.232AB + 0.059AC + 0.105AD - 0.059BC + 0.284BD + 0.033CD$ (4.19)

After building regression metamodels for three responses, they are needed to be verified and validated.

- *Verification and validation of metamodels obtained with BBD*

For verification, lack-of-fit tests are used for all three metamodels. As can be seen from Table 4.36, Table 4.37 and Table 4.38, the metamodels have no statistically significant lack-of-fit (P>0.05) for both responses with 5% significance level (α = 0.05). Additionally R^2 is 99.69%, and R^2-*adjusted* is 99.32% for response Y_1, and R^2 is 99.78%, and R^2-*adjusted* is 99.52% for response Y_2, and R^2 is 99.99%, and R^2-*adjusted* is 99.98% for response Y_3.

For validation, six points are selected randomly in experimentation region which are different from design points. All these points, simulation models results and metamodel results are given in Table 4.31. The simulation model is run at these randomly selected points with 10 replications for each point. The values that are shown at the Simulation Model column are the averages of these 10 replications. The values that are denoted Metamodel Column are predicted by fitting second order regression metamodels.

Table 4.31 Metamodel validation for BBD

	Coded values				**Metamodel**			**Simulation model**			**ARE (%)**		
	A	***B***	***C***	***D***	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
1	-0.5	-1	0	-1	97.00	35.96	61.04	96.69	36.48	61.10	0.32	1.41	0.10
2	0.25	0.5	0.75	1	52.13	13.85	22.78	57.51	15.60	22.60	9.36	11.19	0.80
3	-0.5	1	0	-1	72.40	30.12	19.90	71.39	30.43	20.24	1.41	1.02	1.65
4	-0.5	0	0.75	1	52.98	21.37	30.16	57.81	23.45	30.24	8.35	8.87	0.28
5	0.25	0	0	-1	73.61	24.75	30.31	72.86	24.30	30.50	1.03	1.87	0.62
6	-0.5	-1	0	1	77.23	28.16	60.25	79.25	27.92	60.52	2.55	0.88	0.45

Since not all the *ARE* values are smaller than 5%, the second order metamodels cannot be validated and will not be used for prediction. Therefore, a new design which is named as Nested Box-Behnken Design (*NBBD*) will be used with the aim to obtain validated metamodels.

D. Nested Box-Behnken Design (NBBD)

A *NBBD* is built for fitting second order models. In Table 4.32, the *NBBD* which is used for three responses is depicted. The values related to the central points are obtained by one replication and the others are the average of 10 replications.

After verifying that the residuals of the second order regression models fitted for Y_1, Y_2 and Y_3 are normally, identically and independently distributed with zero mean and constant variance, the ANOVA tables are given in Table 4.33, Table 4.34 and 4.35. Additionally, terms and coefficients for three responses are given in Table 4.36.

Table 4.32 Simulation results for NBBD

	A	***B***	***C***	***D***	Y_1	Y_2	Y_3		***A***	***B***	***C***	***D***	Y_1	Y_2	Y_3
1	-1	-1	0	0	81.351	36.851	61.075	26	0.5	-0.5	0	0	60.538	18.195	36.412
2	1	-1	0	0	81.135	20.429	60.087	27	-0.5	0.5	0	0	59.779	25.593	22.751
3	-1	1	0	0	60.829	33.400	20.314	28	0.5	0.5	0	0	59.992	17.975	22.732
4	1	1	0	0	59.904	15.726	20.253	29	0	0	-0.5	-0.5	67.635	29.794	30.209
5	0	0	-1	-1	144.130	64.624	30.391	30	0	0	0.5	-0.5	60.641	19.650	30.240
6	0	0	1	-1	68.669	20.406	30.147	31	0	0	-0.5	0.5	64.302	27.590	30.043
7	0	0	-1	1	113.200	57.377	30.358	32	0	0	0.5	0.5	57.820	18.949	30.152
8	0	0	1	1	57.410	18.702	30.245	33	-0.5	0	0	-0.5	62.278	26.522	30.335
9	-1	0	0	-1	73.194	37.447	30.447	34	0.5	0	0	-0.5	61.783	18.919	30.399
10	1	0	0	-1	72.779	21.866	30.215	35	-0.5	0	0	0.5	58.662	24.916	30.131
11	-1	0	0	1	58.987	32.557	30.207	36	0.5	0	0	0.5	58.413	17.118	30.197
12	1	0	0	1	58.325	14.986	30.394	37	0	-0.5	-0.5	0	66.591	28.941	36.268
13	0	-1	-1	0	146.380	62.952	60.478	38	0	0.5	-0.5	0	65.707	28.593	22.855
14	0	1	-1	0	118.980	59.129	20.356	39	0	-0.5	0.5	0	59.415	19.272	36.322
15	0	-1	1	0	77.571	20.487	60.550	40	0	0.5	0.5	0	58.899	19.173	22.625
16	0	1	1	0	58.682	18.824	20.193	41	-0.5	0	-0.5	0	66.580	33.058	30.135
17	-1	0	-1	0	130.260	67.216	30.507	42	0.5	0	-0.5	0	66.409	25.729	30.159
18	1	0	-1	0	116.410	54.541	30.442	43	-0.5	0	0.5	0	59.168	23.890	30.232
19	-1	0	1	0	59.423	31.131	30.167	44	0.5	0	0.5	0	59.221	16.283	30.394
20	1	0	1	0	58.874	13.554	30.336	45	0	-0.5	0	-0.5	62.218	21.946	36.349
21	0	-1	0	-1	95.684	32.856	61.005	46	0	0.5	0	-0.5	61.740	21.922	22.814
22	0	1	0	-1	70.935	25.738	19.959	47	0	-0.5	0	0.5	59.227	20.473	36.353
23	0	-1	0	1	77.183	22.844	60.063	48	0	0.5	0	0.5	58.680	20.447	22.688
24	0	1	0	1	58.021	20.242	20.152	49	0	0	0	0	61.720	22.348	30.708
25	-0.5	-0.5	0	0	60.939	26.141	36.672	50	0	0	0	0	58.366	20.544	30.251

Table 4.33 ANOVA table for Y_1 response for NBBD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	23378.6	23378.6	1669.9	27.67	0.000
Linear	4	13244.8	13244.8	3311.2	54.87	0.000
Square	4	9974.8	9974.8	2493.7	41.32	0.000
Interaction	6	159	159	26.5	0.44	0.848
Residual Error	35	2112.3	2112.3	60.4		
Lack-of-Fit	34	2106.7	2106.7	62	11.02	0.235
Pure Error	1	5.6	5.6	5.6		
Total	49	25490.9				
S = 7,76862 R-Sq = 91,71% R-Sq(pred) = 84,31% R-Sq(adj) = 88,40%						

Table 4.34 ANOVA table for Y_2 response for NBBD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	8751.47	8751.47	625.11	67.21	0.000
Linear	4	6012.2	6012.2	1503.05	161.61	0.000
Square	4	2717.95	2717.95	679.49	73.06	0.000
Interaction	6	21.32	21.32	3.55	0.38	0.885
Residual Error	35	325.52	325.52	9.3		
Lack-of-Fit	34	323.89	323.89	9.53	5.85	0.318
Pure Error	1	1.63	1.63	1.63		
Total	49	9076.99				
S = 3,04966 R-Sq = 96,41% R-Sq(pred) = 92,97% R-Sq(adj) = 94,98%						

Table 4.35 ANOVA table for Y_3 response for NBBD

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	14	6247.64	6247.64	446.26	92.37	0.000
Linear	4	5338.94	5338.94	1334.74	276.26	0.000
Square	4	908.1	908.1	227.03	46.99	0.000
Interaction	6	0.6	0.6	0.1	0.02	1.000
Residual Error	35	169.1	169.1	4.83		
Lack-of-Fit	34	169	169	4.97	47.6	0.114
Pure Error	1	0.1	0.1	0.1		
Total	49	6416.75				
S = 2,19806 R-Sq = 97,36% R-Sq(pred) = 95,58% R-Sq(adj) = 96,31%						

In Table 4.33, Table 4.34 and Table 4.35, it is seen that linear, quadratic effects are statistically significant (P<0.05) and two-way interaction effects are statistically insignificant (P>0.05) for all responses. The second order regression models are statistically significant (P>0.05) for all three responses.

Table 4.36 Term and coefficient table for Y_1, Y_2 and Y_3 responses for NBBD

Term	**Coef for Y_1**	**Coef for Y_2**	**Coef for Y_3**
Constant	55.065	19.722	28.404
A	-1.143	-8.030	-0.065
B	-8.935	-1.600	-18.866
C	-27.317	-18.065	-0.050
D	-7.457	-2.724	-0.076
A*A	3.580	4.211	1.176
B*B	12.424	2.841	10.425
C*C	32.189	17.550	1.117
D*D	7.928	2.848	1.056
A*B	-0.131	-0.275	0.232
A*C	3.143	-1.170	0.063
A*D	-0.044	-0.480	0.099
B*C	2.024	0.523	-0.072
B*D	1.311	1.063	0.259
C*D	4.659	1.393	0.035

The second order regression models (metamodels) for Y_1, Y_2 and Y_3 are respectively obtained with *NBBD*;

$$Y_1 = 55.065 - 1.143A - 8.935B - 27.317C - 7.457D + 3.580A^2 + 12.424B^2 + 32.189C^2 + 7.928D^2 - 0.131AB + 3.143AC - 0.044AD + 2.024BC + 1.11BD + 4.659CD$$

$$Y_2 = 19.722 - 8.030A - 1.947B - 18.065C - 2.724D + 4.211A^2 + 2.841B^2 + 17.550C^2 + 2.848D^2 - 0.275AB - 1.170AC - 0.480AD + 0.523BC + 1.063BD + 1.393CD$$

$$Y_3 = 28.404 - 0.065A - 18.866B - 0.050C - 0.076D + 1.176A^2 + 10.425B^2 - 1.117C^2 - 1.056D^2 + 0.232AB + 0.063AC + 0.099AD - 0.072BC + 0.259BD + 0.035CD \quad (4.21)$$

After building regression metamodels for three responses, they are needed to be verified and validated.

- *Verification and validation of metamodels obtained with NBBD*

For verification, lack-of-fit tests are used for all three metamodels. As can be seen from Table 4.33, Table 4.34 and Table 4.35, the metamodels have no statistically significant lack-of-fit (P>0.05) for both responses with 5% significance level (α = 0.05). And R^2 is 91.71%, and R^2-*adjusted* is 88.40% for response Y_1, and R^2 is 96.41%, and R^2-*adjusted* is 94.98% for response Y_2, and R^2 is 97.36%, and R^2-*adjusted* is 96.31% for response Y_3.

For validation, six points are selected randomly in experimentation region which are different from design points. All these points, simulation models results and metamodel results are depicted in Table 4.37. The simulation model is run at these randomly selected points with 10 replications for each point. The values that are denoted at the Simulation Model column are the averages of these 10 replications. The values that are given Metamodel Column are predicted by fitting second order regression metamodels.

Table 4.37 Metamodels validation for NBBD

	Coded values				**Metamodel**			**Simulation model**			**ARE (%)**		
	A	***B***	***C***	***D***	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
1	-0.5	-1	0	-1	94.50	35.49	59.58	96.69	36.48	61.10	2.27	2.71	2.50
2	0.25	0.5	0.75	1	57.20	15.73	23.40	57.51	15.60	22.60	0.54	0.86	3.56
3	-0.5	1	0	-1	74.14	30.44	21.09	71.39	30.43	20.24	3.85	0.02	4.23
4	-0.5	0	0.75	1	56.96	22.96	30.25	57.81	23.45	30.24	1.47	2.09	0.04
5	0.25	0	0	-1	70.40	23.67	29.57	72.86	24.30	30.50	3.38	2.58	3.05
6	-0.5	-1	0	1	77.01	28.39	58.81	79.25	27.92	60.52	2.83	1.70	2.82

All the *ARE* values are smaller than 5%, the second order metamodels are validated and can be used for prediction.

Therefore, the metamodels are decided to be used in Derringer-Suich Multi-Response Optimization phase.

4.4.2.1.2 Derringer-Suich Multi-Response Optimization Procedure. As mentioned in chapter two, Derringer-Suich method uses a desirability function in which the researcher's priorities and desires on the response values are built into one optimization procedure. Firstly, researcher must make a decision about upper limit, lower limit and target values of each response. To decide the upper and lower limits on the average overall flow time of patients spent in the *ED* response (Y_1), the simulation model is run 10 times at the high levels of all input factors. The lower limit is defined as the average of these 10 values, and is found as 57.641 minutes. To determine the upper limit, the simulation model is run 10 times at the low levels of all input factors, and the average value, 208.445 minutes, is found as upper limit. In the current problem, the target value for the utilization of doctor response (Y_2) is aimed to be 60%, the lower and upper limits are determined as 55% and 65%, respectively. On the other hand the target value for the utilization of nurse response (Y_3) is aimed to be 55%, the lower and upper limits are determined as 50% and 60%, respectively.

In order to use formulas (2.34) and (2.37) given in chapter two, the *t*, and *s* weight values must be determined. Three values 0.1, 1 and 10 for *t* and *s* will be used. Therefore, 27 situations (combination of weights) appear to be evaluated. Response Optimizer tool of MINITAB software is used for finding global optimum points for each combination of weights. The determined and used weights, the coded values for each factor and the composite desirability values calculated for the found global optimum with the related weights are given in Table 4.38.

Table 4.38 Composite desirability and coded values for the found global optimum with related weights

	Weights			Composite desirability	Coded values			
	Y_1	Y_2	Y_3		*A*	*B*	*C*	*D*
1	0.1	0.1	0.1	0.977690	-0.251664	-0.897381	-1.000000	0.377018
2	0.1	0.1	1	0.976395	-0.373737	-0.897876	-0.959887	0.111111
3	0.1	0.1	10	0.972876	-0.464300	-0.885531	-0.983581	0.696970
4	0.1	1	0.1	0.977679	-0.250179	-0.897466	-1.000000	0.372523
5	0.1	1	1	0.977607	-0.240779	-0.897972	-1.000000	0.345091
6	0.1	1	10	0.972564	-0.434343	-0.874990	-0.979798	1.000000
7	0.1	10	0.1	0.977726	-0.256245	-0.897104	-1.000000	0.391174
8	0.1	10	1	0.977619	-0.242458	-0.897886	-1.000000	0.349872
9	0.1	10	10	0.977618	-0.242320	-0.897893	-1.000000	0.349477
10	1	0.1	0.1	0.803662	-0.334030	-0.881409	-1.000000	0.902943
11	1	0.1	1	0.821552	-0.757576	-0.865074	-0.898990	1.000000
12	1	0.1	10	0.816116	-0.853657	-0.868103	-0.815674	0.919192
13	1	1	0.1	0.803382	-0.331106	-0.885398	-1.000000	0.801365
14	1	1	1	0.827007	-0.959596	-0.868549	-0.838384	0.656566
15	1	1	10	0.791370	-0.333333	-0.876549	-1.000000	1.000000
16	1	10	0.1	0.803095	-0.327409	-0.887241	-1.000000	0.750256
17	1	10	1	0.797940	-0.250408	-0.897453	-1.000000	0.373214
18	1	10	10	0.802678	-0.321735	-0.889100	-1.000000	0.695294
19	10	0.1	0.1	0.210865	-1.000000	-0.777778	-0.737374	0.898990
20	10	0.1	1	0.175102	-1.000000	-0.858586	-0.636364	-0.393939
21	10	0.1	10	0.141220	-0.373737	-0.886732	-0.919192	0.858586
22	10	1	0.1	0.154604	-0.843171	-0.757576	-0.889326	0.939394
23	10	1	1	0.154759	-0.898990	-0.858586	-0.858586	1.000000
24	10	1	10	0.105723	-0.333333	-0.898990	-0.959596	0.252525
25	10	10	0.1	0.118664	-0.374132	-0.814552	-1.000000	0.860596
26	10	10	1	0.094663	-0.212121	-0.919192	-1.000000	0.313131
27	10	10	10	0.112155	-0.332432	-0.884400	-1.000000	0.827850

The coded values of input factors shown in Table 4.38 are transformed to natural values by using the formula (2.4) given in chapter two are depicted in Table 4.39. Response Optimizer tool of MINITAB software is used for calculating the metamodel results which uses the found coded values and fitted second order regression metamodels given in 4.21. Since the optimum levels of input factors are found by using the metamodels, confirmatory simulation runs are needed at these optimum input levels. The simulation model results denoted in Table 4.39 are the average values of 10 replications in which natural values are used.

Table 4.39 Results of confirmatory runs

	Weights			Composite desirability	Natural values				Metamodel results			Simulation results		
	Y_1	Y_2	Y_3		*A*	*B*	*C*	*D*	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
1	0.1	0.1	0.1	0.977690	4	3	4	12	131.805	60.000	55.000	141.890	64.831	60.496
2	0.1	0.1	1	0.976395	4	3	4	11	131.382	59.986	55.021	141.410	65.177	60.403
3	0.1	0.1	10	0.972876	4	3	4	13	129.840	60.588	55.002	136.800	63.396	60.143
4	0.1	1	0.1	0.977679	4	3	4	12	131.832	60.000	55.000	141.890	64.831	60.496
5	0.1	1	1	0.977607	5	3	4	12	132.001	60.000	55.000	138.830	61.947	60.607
6	0.1	1	10	0.972564	4	3	4	14	129.034	60.012	54.992	141.010	64.178	60.493
7	0.1	10	0.1	0.977726	4	3	4	12	131.722	60.000	55.000	141.890	64.831	60.496
8	0.1	10	1	0.977619	5	3	4	12	131.971	60.000	55.000	138.830	61.947	60.607
9	0.1	10	10	0.977618	5	3	4	12	131.974	60.000	55.000	138.830	61.947	60.607
10	1	0.1	0.1	0.803662	4	3	4	14	130.168	60.000	55.000	141.010	64.178	60.493
11	1	0.1	1	0.821552	3	3	4	14	124.395	59.902	54.984	150.330	69.322	60.755
12	1	0.1	10	0.816116	3	3	5	14	118.330	57.326	55.009	102.210	52.123	60.795
13	1	1	0.1	0.803382	4	3	4	13	130.250	60.000	55.000	136.800	63.396	60.143
14	1	1	1	0.827007	3	3	5	13	121.248	59.898	54.993	103.240	52.602	60.805
15	1	1	10	0.791370	4	3	4	14	130.188	60.015	54.979	141.010	64.178	60.493
16	1	10	0.1	0.803095	4	3	4	13	130.334	60.000	55.000	136.800	63.396	60.143
17	1	10	1	0.797940	4	3	4	12	131.828	60.000	55.000	141.890	64.831	60.496
18	1	10	10	0.802678	4	3	4	13	130.455	60.000	55.000	136.800	63.396	60.143
19	10	0.1	0.1	0.210865	3	4	5	14	111.012	55.630	51.945	83.721	49.280	45.880
20	10	0.1	1	0.175102	3	3	5	10	115.792	55.373	54.540	111.170	55.516	61.199
21	10	0.1	10	0.141220	4	3	4	14	123.457	56.337	55.003	141.010	64.178	60.493
22	10	1	0.1	0.154604	3	4	4	14	121.062	59.989	51.212	124.890	65.507	45.518
23	10	1	1	0.154759	3	3	5	14	122.200	59.978	54.974	102.210	52.123	60.795
24	10	1	10	0.105723	4	3	4	12	129.682	59.047	55.017	141.890	64.831	60.496
25	10	10	0.1	0.118664	4	4	4	14	128.345	60.000	52.551	120.660	61.518	45.527
26	10	10	1	0.094663	5	3	4	12	132.795	59.996	55.766	138.830	61.947	60.607
27	10	10	10	0.112155	4	3	4	13	130.218	60.000	55.000	136.800	63.396	60.143
28	Current				3	4	4	9				152.790	69.991	45.318

In Table 4.39, first 27 rows reflect the global optimum solutions found by Derringer-Suich multi-response optimization method according to weight combinations, and confirmatory simulation results. The last row shows the current situation and validated simulation results.

In order to determine optimum results all three responses are checked. It is seen that the minimum value for average overall flow time of patients spent in the *ED* (length of stay) with satisfied doctor utilization rate and nurse utilization rate requirements are obtained at the natural factor levels A= 4, B = 3, C = 4, D = 13 in the 3rd, 13rd, 16th, 18th and 27th rows are demonstrated in grey box colour in the Table 4.39.

As a result, the obtained optimum point at natural values A= 4, B = 3, C = 4, D = 13 calculated as; 136.800 minutes average overall flow time of patients (length of stay), 63.396% average utilization rate of doctor and 60.143% average utilization rate of nurse.

After the optimization study, it is seen that the average overall flow time of patient spent in the *ED* (Y_1) is decreased by 15.990 minutes (10.465%), the utilization rate of doctor (Y_2) is decreased to acceptable level, and the utilization rate of nurse (Y_3) is increased to acceptable level.

CHAPTER FIVE
CONCLUSIONS

Nowadays healthcare systems become more important and the medical levels of hospitals have greatly improved. However, unnecessary and long waiting times and delays in hospitals and especially in *EDs* still remain as a problem to be solved. When an unnecessary waiting and delay occurs, this situation affects not only patients but also doctors and personnels. Because patient satisfaction is one of the most important aspects in healthcare systems and it tightly depends on waiting and delay times, it's really vital to analyze the key process to eliminate the bottlenecks, unnecessary waiting times and optimize the process of healthcare system. Improving the healthcare systems and achieving the patient satisfaction require appropriate allocation of human and material resources in order to improve its effectiveness and efficiency.

In this thesis, for an *ED* of a hospital in Izmir, the aim is to find the optimal allocation of resources while minimizing the average overall flow time of patients (length of stay) and keeping the utilization ratios of doctors and nurses at the required levels. A simulation optimization in the *ED* is studied with three main objectives; (1) to review the relevant literature, (2) to develop a simulation model of a real *ED* system and (3) to do an optimization study considering multiple responses simultaneously.

First of all, the studies on simulation optimization in *ED*s and also *RSM* studies in healthcare systems have been published in the last decade are reviewed. They are classified according to the problem type, aim, key performance index, modeling approach, software used and data gathering method and data sources.

RSM is a collection of mathematical and statistical techniques, which are useful for developing, improving, and optimizing processes. It also has important applications in the design, development, and formulation of new products, as well as in the improvement of existing product designs, and it is an effective tool for constructing optimization models (Myers & Montgomery, 2016, p.1). *RSM* is generally used for optimizing the performance (or a model) of an unknown system, which is subject to controllable, uncontrollable and unknown variables. It can be applied to any system that has key elements like a criterion of effectiveness known as the response of the system, which is measurable on continuous scale, and quantifiable independent variables that affect the system's performance. At these conditions, *RSM* is a group of techniques for finding the optimum response of the system in an optimum fashion. Although from a historical viewpoint *RSM* has been applied primarily to Operations Management, its greatest value to decision scientists is its potential application to simulation studies (Brightman, 1978, p.481). Because it is a useful tool in optimization models, in this thesis, *RSM* is used to optimize the multiple performance measures of the real *ED* system.

In the application part, at first, the problem in the *ED* is defined in detail. The problem is the bottlenecks which cause delays, long length of stays, patient and personnel dissatisfaction in the hospital in İzmir. This problem causes more intensity in the *ED* and decrease in the quality of service. Therefore, determining the optimum personnel and resource numbers plays a vital role for improving the current system. By aiming this purpose, after depicting the features of the *ED* system, patient acceptance, assessment and treatment processes, assumptions and the inputs of the simulation model, the simulation model of the real *ED* is developed by Arena simulation modeling software. Next, the verification and validation of the simulation model are depicted. In the studied current system, the resources (input factors) are three doctors, four nurses, four triage beds and nine observation beds. The values of the three performance measures are; the average overall flow time is 152.79 minute, average doctor utilization ratio is 69.991% and average nurse utilization ratio is 45.318 % calculated by the simulation model.

Secondly, the data obtained from the simulation model is used for building verified and validated metamodels for the three responses. The aim of the study is to find the best allocation of resources (input factors) minimizing the patient's average overall flow time spent in the *ED* (first response) with the requirement of 60%

utilization rate of the doctors (second response) and 55% utilization rate of the nurses (third response). And the input factors are, respectively;

A: The number of doctors in the *ED*;

B: The number of nurses in the *ED*;

C: The number of triage beds in the *ED*;

D: The number of observation beds in the *ED*.

After making an OFAT, the appropriate levels of input factors are decided. Next, for fitting first order regression models, two-level full factorial design with central runs are designed. Because the true response surface usually exhibits curvature near the optimum, the second order models are built. *CCD*, *NFCD*, *BBD* are used to build the second order models. Since not the all calculated *ARE* values are smaller than the threshold in the validation part, the second order metamodels cannot be validated so cannot be used for the prediction. Therefore, a new design named as Nested Box-Behnken Design (*NBBD*) is developed with the aim to obtain validated metamodels for the first time in the reached literature. Since the all *ARE* values are smaller than the threshold value, the second order metamodels obtained by *NBBD* are said to be validated and are used for the prediction.

These metamodels are used in Derringer-Suich Multi-Response Optimization phase of the thesis. Derringer-Suich multi response optimization procedure is applied to minimize average overall flow time of patients (length of stay) with keeping utilization ratios of doctors and nurses at target levels. As a result, the obtained optimum point at natural values $A= 4$, $B = 3$, $C = 4$, $D = 13$ for which; 136.800 minutes average overall flow time of patients, 63.396% average utilization rate of doctors and 60.143% average utilization rate of nurses are calculated.

After the optimization study, it is seen that the average overall flow time of patient spent in the *ED* (Y_1) is decreased by 15.990 minutes (10.465%), the utilization rate of doctors (Y_2) is decreased to an acceptable level, and the utilization rate of nurses (Y_3) is increased to an acceptable level.

The contributions of this thesis is in two points:

- In this thesis multi-responses (three responses) are optimized simultaneously. To the best of our knowledge, this study is the first one that uses multi-response surface methodology to optimize a problem in an *ED* and the second one (the first one is Liang, Guo & Fung (2015) dealt with surgery scheduling in operating theaters) in all healthcare systems.
- In the design of experiments part, this study has a novel approach because *NBBD* is developed and applied to obtain second order metamodels again to the best of our knowledge it is the first time.

This study will provide help for healthcare decision makers to find the optimal allocation of resources in an *ED* while considering to satisfy multi objectives simultaneously.

REFERENCES

Ahmed, M.A., & Alkhamis, T.M. (2009). Simulation optimization for an emergency department healthcare unit in Kuwait. *European Journal of Operational Research*, *198*(3), 936-942.

Akcan, S., & Kokangul, A. (2013). A new approximation for inventory control system with decision variable lead-time and stochastic demand. *International Journal of Industrial Engineering*, 20.

Akteke-Öztürk, B., Weber, G.W., & Köksal, G. (2015). Desirability functions in multiresponse optimization, optimization in the natural sciences. *Communications in Computer and Information Science*, *499*, 129-146.

Akteke-Öztürk, B., Weber, G.W., & Köksal, G. (2017). Optimization of generalized desirability functions under model uncertainty. *Optimization*, *66*(12), 2157-2169.

Almagooshi, S. (2015). Simulation modelling in healthcare: Challenges and trends. *Procedia Manufacturing*, *3,* 301-307.

Angün, E., Kleijnen, J.P.C., Hertog, D.D., & Gürkan, G. (2002). Response surface methodology revisited. *In Proceedings of Winter Simulation Conference*, 377-383.

Aytuğ, H., Doğan, C.A., & Bezmez, G. (1996). Determining the number of kanbans: A simulation metamodeling approach, *Simulation*, *67*:1, 23-32.

Azadeh, A., Farahani, M.H., Torabzadeh, S., & Baghersad, M. (2014). Scheduling prioritized patients in emergency department laboratories. *Computer Methods and Programs in Biomedicine*, *117*(2), 61-70.

Azadeh, A., Ahvazi, M.P., Haghighii, S. M., & Keramati, A. (2016). Simulation optimization of an emergency department by modeling human errors. *Simulation Modelling Practice and Theory*, *67*, 117-136.

Baesler, F.F. (2016, December). Multiobjective emergency room capacity planning using simulation and response surface methodology. In *Proceedings of the 2016 Winter Simulation Conference* (3706-3707). IEEE Press.

Batmaz, I., & Tunali, S. (2003). Small response surface designs for metamodel estimation. *European Journal of Operational Research, 145*, 455-470.

Blanning, R.W. (1975). The construction and implementation of metamodels. *Simulation*, *24*(6), 177-184.

Box, G.E.P., & Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society, Series B*, *13,* 1–45.

Box, G. E. P., & Behnken, D.W. (1960). Some new three-level designs for the study of quantitative variables. *Technometrics*, *2*, 455–475.

Brightman, H.J. (1978). Optimization through experimentation: Applying response surface methodology. *Decision Sciences, 9*, 481-495.

Cabrera, E., Taboada, M., Iglesias, M.L., Epelde, F., & Luque, E. (2011). Optimization of healthcare emergency departments by agent-based simulation. *Procedia Computer Science*, *4*, 1880-1889.

Cabrera, E., Taboada, M., Iglesias, M.L., Epelde, F., & Luque, E. (2012). Simulation optimization for healthcare emergency departments. *Procedia Computer Science*, *9*, 1464-1473.

Cabrera, E., Luque, E., Taboada, M., Epelde, F., & Iglesias, M.L. (2012, December). ABMS optimization for emergency departments. In *Proceedings of the Winter Simulation Conference*, 89, Winter Simulation Conference.

Cabrera, E. (2013). *Optimisation via simulation for healthcare emergency departments*. Doctoral dissertation, Universitat Autònoma de Barcelona, Departament d'Arquitectura de Computadors i Sistemes Operatius.

Candioti, L.V., De Zan, M.M., Cámara, M.S., & Goicoechea, H.C. (2014). Experimental design and multiple response optimization. Using the desirability function in analytical methods development. *Talanta*, *124*, 123-138.

Castillo, E.D., Montgomery, D.C., & McCarvile, D.R. (1996). Modified Desirability Functions for multiple response optimization. *Journal of Quality Technology*, 28, 337-345.

Chen, T.L. (2014, June). Decision support system based on distributed simulation optimization for medical resource allocation in emergency department. *In International Conference on HCI in Business* (p.15-24). Springer, Cham.

Chen, T.L., & Wang, C.C. (2016). Multi-objective simulation optimization for medical capacity allocation in emergency department. *Journal of Simulation*, *10*(1), 50-68.

Costa, N.R., Lourenço, J., & Pereira, Z.L. (2011). Desirability function approach: a review and performance evaluation in adverse conditions. *Chemometrics and Intelligent Laboratory Systems*, *107*(2), 234-244.

Costa, N.R., & Lourenço, J. (2016). Multiresponse problems: desirability and other optimization approaches. *Journal of Chemometrics*, *30*(12), 702-714.

Derringer, G.C. (1994). A balancing act: Optimizing a product's properties. *Quality Progress*, *27*, 51-58.

Derringer, G.C., & Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology*, *12*, 214-219.

Draper, N.R. (1960) A third order rotatable design in four dimensions. *Annals of Mathematical Statistics*, *31*(4), 875–877.

Draper, N.R. (1962). Third order rotatable designs in three factors. *Technometrics*, *4*(2), 219–234.

Draper, N.R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). John Wiley & Sons.

Elalouf, A., & Wachtel, G. (2015). An alternative scheduling approach for improving patient-flow in emergency departments. *Operations Research for Health Care*, *7*, 94-102.

El-Rifai, O., Garaix, T., Augusto, V., & Xie, X. (2015). A stochastic optimization model for shift scheduling in emergency departments. *Health Care Management Science*, *18*(3), 289-302.

El-Zoghby, J., Farouk, H.A., & El-Kilany, K.S. (2016). An integrated framework for optimization of resources in emergency departments. In *6th International Conference on Industrial Engineering and Operations Management in Kuala Lumpur, IEOM*, 1621-1632.

Feili, H. (2013). Improving the health care systems performance by simulation optimization. *Journal of Mathematics and Computer Science 7 (2013) 73, 79*.

Feng, Y.Y., Wu, I.C., & Chen, T.L. (2017). Stochastic resource allocation in emergency departments with a multi-objective simulation optimization algorithm. *Health Care Management Science*, *20*(1), 55-75.

Friedman, L.W. (1995). *The simulation metamodel.* Massachusets: Kluwer Academic Publishers.

Fuller, D., & Scherer, W. (1998). The desirability function: Underlying assumptions and application implications. *IEEE International Conference*, *4*, 4016-4021.

Ghanes, K., Jouini, O., Jemai, Z., Wargon, M., Hellmann, R., Thomas, V., & Koole, G. (2014, December). A comprehensive simulation modeling of an emergency department: A case study for simulation optimization of staffing levels. In *Proceedings of the 2014 Winter Simulation Conference*, 1421-1432, IEEE Press.

Ghanes, K., Wargon, M., Jouini, O., Jemai, Z., Diakogiannis, A., Hellmann, R., Thomas V., & Koole, G. (2015). Simulation-based optimization of staffing levels in an emergency department. *Simulation*, *91*(10), 942-953.

Grummels, C.A. (2012). *Simulation and optimization of an emergency department.* Master's Thesis, Eindhoven University of Technology Department of Mechanical Engineering, Systems Engineering Group, The Netherlands.

Gul, M., & Guneri, A.F. (2015). A comprehensive review of emergency department simulation applications for normal and disaster conditions. *Computers & Industrial Engineering*, *83*, 327-344.

Guo, H., & Tang, J. (2015). Integrating simulation with optimization in emergency department management. *In Engineering Asset Management-Systems, Professional Practices and Certification*, 1483-1496, Springer, Cham.

Guo, H., Goldsman, D., Tsui, K.L., Zhou, Y., & Wong, S.Y. (2016). Using simulation and optimisation to characterise durations of emergency department service times with incomplete data. *International Journal of Production Research*, *54*(21), 6494-6511.

Guo, H., Gao, S., Tsui, K.L., & Niu, T. (2017). Simulation optimization for medical staff configuration at emergency department in Hong Kong. *IEEE Transactions on Automation Science and Engineering*, *14*(4), 1655-1665.

Harrington, E.C. (1965) The desirability function. *Industrial Quality Control*, *21*, 494-498.

He, Y., Cai, B., & Wang, M. (2016). Research on Optimization of Registration Procedure in Emergency Department Based on System Simulation. *Procedia Computer Science*, *91*, 37-46.

Hood, S.J., & Welch, P.D. (1993). Response surface methodology and its application in simulation. *In Proceedings of Winter Simulation Conference*, 115-122.

Ibrahim, I.M., Liong, C.Y., Bakar, S.A., Ahmad, N., & Najmuddin, A.F. (2015, December). A proposed simulation optimization model framework for emergency department problems in public hospital. *In AIP Conference Proceedings, 17*(1), 030011, AIP Publishing.

Ibrahim, I. M., Liong, C.Y., Bakar, S.A., Ahmad, N., & Najmuddin, A.F. (2017, April). Minimizing patient waiting time in emergency department of public hospital using simulation optimization approach. *In AIP Conference Proceedings, 1803*(1), 060005, AIP Publishing.

K, M., & Harris, P. (2016). A novel approach for resource allocation in emergency department of hospitals. *International Journal of Advanced Research in Computer Science and Software Engineering, 6*(3).

Kang, H., & Lobo, J.M. (2016, December). Identifying the optimal configuration of an express care area in an emergency department: a DES and metamodeling approach. In *Winter Simulation Conference (WSC), 2016*, 1961-1969, IEEE.

Kasaie, P., & Kelton, W.D. (2013a). Simulation optimization for allocation of epidemic-control resources. *IIE Transactions on Healthcare Systems Engineering, 3*(2), 78-93.

Kasaie, P., & Kelton, W.D. (2013b). Resource allocation for controlling epidemics: Calibrating, analyzing, and optimizing an agent-based simulation. *IIE Transactions on Healthcare Systems Engineering, 3*(2), 94-109.

Keshtkar, L., Salimifard, K., & Faghih, N. (2015). A simulation optimization approach for resource allocation in an emergency department. *QScience Connect*, 8.

Kleijnen, J.P., & Sargent, R.G. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, *120*(1), 14-29.

Kokangul, A., Akcan, S., & Narli, M. (2017). Optimizing nurse capacity in a teaching hospital neonatal intensive care unit. *Health Care Management Science, 20*(2), 276-285.

Lal, T.M., Roh,T., & Huschka, T. (2015, December). Simulation based optimization: Applications in healthcare. In *Winter Simulation Conference (WSC), 2015*, 1261-1271, IEEE.

Landman, D., Simpson, J., Mariani, R., Ortiz, F., & Britcher, C. (2007). Hybrid Design for Aircraft Wind-Tunnel Testing Using Response Surface Methodologies. *Journal of Aircraft*, *44*(4).

Liang, F., Guo, Y., & Fung, R.Y. (2015). Simulation-based optimization for surgery scheduling in operation theatre management using response surface method. *Journal of medical systems*, *39*(11), 159.

Liu, Z. (2017). Modeling and Simulation for healthcare operations management using high performance computing and agent-based model. *Journal of Computer Science & Technology*, *17*.

Liu, Z., Rexachs, D., Epelde, F., & Luque, E. (2017a). A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity. *Computers & Industrial Engineering*, *103*, 300-309.

Liu, Z., Rexachs, D., Epelde, F., & Luque, E. (2017b). An agent-based model for quantitatively analyzing and predicting the complex behavior of emergency departments. *Journal of Computational Science*, *21*, 11-23.

Madu, C.N. (1990). Simulation in manufacturing: A regression metamodel approach. *Computers & Industrial Engineering*, *18*, 381-389.

Minitab Express Support. (2016) *Interpret the key results for multiple regression.* Retrieved March 13 2018, from http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/how-to/multiple-regression/interpret-the-results/key-results/

Montgomery, D.C. (2012). *Design and analysis of experiments* (8^{th} ed.). New York: John Wiley & Sons, Inc.

Montgomery, D.C. (2017). *Design and analysis of experiments* (9^{th} ed.). John Wiley & Sons.

Myers, R.H., & Montgomery, D.C. (1995). *Response surface methodology: Process and product optimization using designed experiments.* New York: John Wiley & Sons, Inc.

Myers, R.H., Montgomery, D.C., Vining, G.G., Borror, C.M., & Kowalski, S.M. (2004). Response surface methodology: A retrospective and literature survey. *Journal of Quality Technology*, *36*, 53-77.

Myers, R.H., & Montgomery, D.C. (2016). *Response surface methodology: Process and product optimization using designed experiments* (4th ed.). New York: John Wiley & Sons, Inc.

Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Boston: McGraw-Hill, Inc.

Niessner, H., Rauner, M.S., & Gutjahr, W.J. (2018). A dynamic simulation–optimization approach for managing mass casualty incidents. *Operations Research for Health Care, 17,* 82-100.

NIST/SEMATECH. (2003). *Engineering Statistics Handbook*. Retrieved May 6 2018, from NIST/SEMATECH e-Handbook of Statistical Methods' web site. http://www.itl.nist.gov/div898/handbook/

Oh, C., Novotny, A.M., Carter, P.L., Ready, R.K., Campbell, D.D., & Leckie, M.C. (2016). Use of a simulation-based decision support tool to improve emergency department throughput. *Operations Research for Health Care*, *9*, 29-39.

Osborne, D.M., Armacost, R.L., & Edwards, J.P. (1997). State of the art in multiple response surface methodology. *IEEE International Conference*, *4,* 3833-3838.

Özler, C. (1999). Çekicilik (desirability) fonksiyonları üzerine bir inceleme. *D.E.Ü.İ.İ.B.F. Dergisi, 14*(2), 111-121.

Paul, S.A., Reddy, M.C., & DeFlitch, C.J. (2010). A systematic review of simulation studies investigating emergency department overcrowding. *Simulation, 86*(8-9), 559-571.

Pegden, C.D., Shannon, R.E., & Sadowski, R.P. (1990). *Introduction to simulation using SIMAN.* New York: McGraw-Hill, Inc.

Pinto, L.R., de Campos, F.C.C., Perpétuo, I. H.O., & Ribeiro, Y.C.N.M.B. (2014, December). Analisys of hospital bed capacity via queuing theory and simulation. In *Simulation Conference (WSC), 2014 Winter*, 1281-1292, IEEE.

Rabbani, M., Farshbaf-Geranmayeh, A., & Yazdanparast, R. (2018). A simulation optimization approach for integrated resource allocation in an emergency department, pharmacy, and lab. *Intelligent Decision Technologies,* (Preprint), 1-26.

Ribardo, C., & Allen, T.T. (2003). An alternative desirability function for achieving "six sigma" quality. *Quality and Reliability Engineering International, 19*, 227-240.

Santos, M.I.R, & Nova, A.M.O.P. (1999). The main issues in nonlinear simulation metamodel estimation. *In Proceedings of Winter Simulation Conference,* 502-509.

Savin, N.E., & White, K.J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica: Journal of the Econometric Society*, 1989-1996.

Sridharan, R., & Babu, A.S. (1998). Multi-level scheduling decisions in a class of FMS using simulation based metamodels. *Journal of the Operational Research Society, 49,* 591-602.

Sun, Y., & Li, X. (2013). Response surface optimisation of surgery start times in a single operating room using designed simulation experiments. *International Journal of Healthcare Technology and Management*, *14*(1-2), 61-72.

Tan, K.W., Lau, H.C., & Lee, F.C.Y. (2013). Improving patient length-of-stay in emergency department through dynamic queue management. (2013). WSC '13: *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World, December 8-11, 2013*, Washington DC. , 2362. Research Collection School of Information Systems.

Uriarte, A.G., Zúñiga, E.R., Moris, M.U., & Ng, A.H. (2015). System design and improvement of an emergency department using simulation-based multi-objective optimization. *In Journal of Physics: Conference Series, 616*(1),012015, IOP Publishing.

Uriarte, A.G., Zúñiga, E.R., Moris, M.U., & Ng, A.H. (2017). How can decision makers be supported in the improvement of an emergency department? A simulation, optimization and data mining approach. *Operations Research for Health Care, 15*, 102-122.

Weng, S.J., Cheng, B.C., Kwong, S.T., Wang, L.M., & Chang, C.Y. (2011, December). Simulation optimization for emergency department resources allocation. In *Proceedings of the Winter Simulation Conference*, 1231-1238, Winter Simulation Conference.

Yalçınkaya, Ö. (2004). *An optimization study for urban public transport system,* MSc Thesis, Dokuz Eylül University, Department of Industrial Engineering, Izmir, Turkey.

Yang, Y. (2008). Multiple criteria third-order response surface design and comparison (Doctoral dissertation, Florida State University).

Yazdanparast, R., Hamid, M., Azadeh, M.A., & Keramati, A. (2018). An Intelligent Algorithm for Optimization of Resource Allocation Problem by Considering Human Error in an Emergency Department. *Journal of Industrial and Systems Engineering*, *11*(1), 0-0.

Yıldız, G. (2003). *Simulation optimization of dual resource constrained CONWIP controlled lines using response surface methodology*, PhD Thesis, Dokuz Eylül University, Department of Industrial Engineering, Izmir, Turkey.

Yousefi, M., Yousefi, M., Ferreira, R.P.M., Kim, J.H., & Fogliatto, F.S. (2017). Chaotic genetic algorithm and Adaboost ensemble metamodeling approach for optimum resource planning in emergency departments. *Artificial Intelligence In Medicine, 84*, 23-33.

Yu, B., & Popplewell, K. (1994). Metamodels in manufacturing: A review. *International Journal of Production Research*, *32,* 787-796.

Zeinali, F., Mahootchi, M., & Sepehri, M.M. (2015). Resource planning in the emergency departments: A simulation-based metamodeling approach. *Simulation Modelling Practice and Theory*, *53*, 123-138.