COST ESTIMATION OF TRACKWORKS OF LIGHT RAIL AND METRO PROJECTS

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

BY

ERHAN ÖZTÜRK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

JANUARY 2009

Approval of the Thesis;

COST ESTIMATION OF TRACKWORKS OF LIGHT RAIL AND METRO PROJECTS

Submitted by **ERHAN ÖZTÜRK** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of
Natural and Applied Sciences _______________

Prof. Dr. Güney Özcebe
Head of Department, **Dept. of Civil Engineering** _______________

Assoc. Prof. Dr. Murat Gündüz
Supervisor, **Dept. of Civil Engineering, METU** _______________

Examining Committee Members:

Assist. Prof. Dr. Metin Arıkan (*)
Civil Engineering Dept., METU _______________

Assoc. Prof. Dr. Murat Gündüz (**)
Civil Engineering Dept., METU _______________

Assoc. Prof. Dr. Soofia Tahira Elias-Özkan
Architecture Dept., METU _______________

Instr. Dr. Engin Erant
Civil Engineering Dept., METU _______________

Assist. Prof. Dr. Latif Onur Uğur
Junior Technical College., Ahi Evran University _______________

Date: _______________

(*) Head of Examining Committee

(**) Supervisor

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

Name, Last name : Erhan ÖZTÜRK

Signature :

ABSTRACT

COST ESTIMATION OF TRACKWORKS OF LIGHT RAIL AND METRO PROJECTS

Öztürk, Erhan

M.S., Department of Civil Engineering

Supervisor: Assoc. Prof. Dr. Murat Gündüz

January 2009, 97 pages

The main objective of this work is to develop models using multivariable regression and artificial neural network approaches for cost estimation of the construction costs of trackworks of light rail transit and metro projects at the early stages of the construction process in Turkey. These two approaches were applied to a data set of 16 projects by using seventeen parameters available at the early design phase.

According to the results of each method, regression analysis estimated the cost of testing samples with an error of 2.32%. On the other hand, artificial neural network estimated the cost with 5.76% error, which is slightly higher than the regression error. As a result, two successful cost estimation models have been developed within the scope of this study. These models can be beneficial while taking the decision in the tender phase of projects that includes trackworks.

Keywords: Artificial Neural Network, Early Cost Estimation, Metro, Light Rail Transit, Regression

ÖZ

HAFİF RAYLI VE METRO PROJELERİNİN HAT İŞLERİ FİYAT TAHMİNİ

Öztürk, Erhan

Yüksek Lisans, İnşaat Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Murat Gündüz

Ocak 2009, 97 sayfa

Bu çalışmanın ana hedefi çok değişkenli regresyon analizi ve yapay sinir ağları yaklaşımlarını kullanarak Türkiye'de ön proje aşamasında bulunan hafif raylı ve metro projelerinin hat işleri için ön fiyat tahmin modelinin geliştirilmesidir. Bu iki yaklaşım tamamlanmış 16 projeye ait ön tasarım aşamasında mevcut bulunan17 parametreyi içeren bir data setine uygulanmıştır.

Her bir methodun sonuçlarına göre, regresyon analizi test örneklerinin fiyatlarını ortalama hata yüzdesi 2.32 olacak şekilde hesaplamıştır. Diğer taraftan, yapay sinir ağı fiyatı ortalama yüzde 5.76 hata payı ile tahmin etmiştir ki bu değer regresyon sonucundan bir miktar yüksektir. Sonuç olarak, çalışma kapsamı çerçevesinde iki adet başarılı fiyat tahmin modeli geliştirilmiştir. Bu modeller hat işleri içeren projelerin ihaleye giriş kararının alınması aşamasında yararlı olabilecektir.

Anahtar Kelimeler: Yapay Sinir Ağları, Ön fiyat Tahmini, Metro, Hafif Raylı Sistem, Regresyon

TO MY FAMILY

ACKNOWLEDGMENTS

This study was suggested and has been completed under the supervision of Assoc. Prof. Dr. Murat GÜNDÜZ in the Middle East Technical University (METU) in Ankara, Turkey.

I would like to express my sincere thanks and deep sense of gratitude to my supervisor Assoc. Prof. Dr. Murat GÜNDÜZ for his invaluable guidance, interest, patience and encouragement throughout this study. His wealth of knowledge and enthusiasm to share them with me is most appreciated. I also express my appreciation to H. Asım MUSLUOĞLU and Gürler DUMAN for their great aids.

My special gratitude goes to Assist. Prof. Dr. Latif Onur UĞUR for his invaluable help, support and patience for my frequent disturbances.

Finally, I would like to thank my family for their “all the time” love, understanding, support, encouragement and patience, which gave me the strength, the motivation and the inspiration to complete this study. I always try to merit you.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF SYMBOLS

AAR	AustralAsia Railway
ANN	Artificial Neural Network
APTA	American public Transportation Association.
AREMA	American Railway Engineering and Maintenance-of-Way Association
BP	Back Propagation
CEEC	Central and Eastern European Countries
ERRAC	European Rail Research Advisory Council
EU	European Union
GA	Genetic Algorithm
LRT	Light Rail Transit
m	Number of Input Layer Neurons
MAPE	Mean Absolute Percent Error
P	Significance Level
r	Correlation Coefficient
R^2	Coefficient of Determination
RMSE	Root of Mean Square Error
UITP	International Union of Public Transport
UK	United Kingdom
US	United States
USD	United States Dollar

CHAPTER 1

INTRODUCTION

1.1 GENERAL REMARKS

In today's world, due to the growing of population and its accumulation in the city centers, the public transportation comes out one of the most important issues, which will be handled by infrastructural investments to cities. When we consider the previous experiences and many other researches on ways of mass passenger transport in city centers, it is obvious that the most efficient solution to the public transportation is Light Rail Train (LRT) and Metro systems. These systems are used over centuries in developed countries of the world unlike developing countries. There is a considerable gap in terms of the availability of length of LRT or Metro line per citizen between developed and developing countries. That's why in order to compensate this gap and provide a modern service to their societies; recently the municipalities of developing countries start to make huge investments to these public transportation systems. At this point, an accurate early cost estimation of these systems becomes more critical for many parties including owners while taking investment decisions, because they have very limited budget. Besides, deviation from the pre-defined budget often brings a quick response from the public, the press, and sometimes even the state legislature. When this occurs, the municipality or state loses credibility over society and at the end the projects becomes less efficient than the design stage (Chester et al., 2005). On the other hand, if the owner can produce realistic budgets their image is enhanced and society gains.

In addition to these; when we consider budget in terms of the contractors, the accuracy of estimation of construction costs in a construction project is a critical factor in the success of the project. The cost estimation models, which in the early stage estimate the construction costs with minimum project information, are useful in the preliminary design stage of a construction project. Improved cost estimation techniques, which are available to project managers, will facilitate more effective control of time and costs in construction projects (Hegazy T., 2002).

Cost estimation is an area, which over the years has received much attention from civil and cost engineers. In an ideal situation all necessary cost information is present, allowing calculating the costs accurately. For the construction sector, this information comes from the geotechnical investigations, topographic measurements, structural design, and methods to be use. Collecting and combining of all of these components of detailed design stage, which are generating by separate specialized parties, takes considerable amount of time. However, sometimes reliable cost estimation is required within a very limited time period in order to decide the feasibility of the projects; it cannot be justified to generate detailed design drawings for every possible business development opportunity. Since these design stages are too time consuming, other fast yet accurate methods are required (Verlinden et al., 2008). Therefore, parametric cost estimation methods, which are very useful in the early stage of a project's life cycle, has been introduced, when little information is known about the project's scope. These parametric cost estimation models include historical data that are currently used in practice as well as new data specific to a new project. One of the widely used parametric modeling type is regression, or multiple regression analysis. This is a very unique technique which can be used both analytical and predictive purposes by considering the affect of potential new items to the overall estimate reliability, although it is not appropriate when describing non-linear relationships, which are multidimensional, consisting of a multiple input and output problem (Tam and Fang.,1999). Another type is artificial neural network (ANN) is a computer system that simulates the learning process of the human brain. ANNs are widely applied in many industrial areas, including construction. The applicability of ANNs to construction has been extensively studied (Boussabaine., 1996). In addition, researchers have explored the application of ANNs to improve the accuracy of cost estimating beyond that of the regression model (Garza and Rouhana., 1995). In this study, these two techniques are used in order to evaluate historical project data.

1.2 OBJECTIVES

The aim of this study is to establish and compare cost estimations models in order to assist cost prediction of trackworks of Light rail and Metro systems in Turkey regardless of the type of the infrastructure system of the project. In other word, the developed model for railway superstructure does not depend on feature or type of the section of the line such as TBM (tunnel boring machine) tunnel, depressed open/close or at grade line. For this reason, the data of completed LRT and Metro projects in which includes trackworks in their scope, were collected via site visits and related municipality and contractor interviews.

1.3 SCOPE

The study reported herein is based on realized data of actively working and under construction LRT and Metro Projects in Turkey. These data are gathered from various companies, which are responsible for construction of track works of above mentioned projects. Trackworks data of 16 projects were analyzed by parametric cost estimations models, which are regression and neural networks.

1.4 SYNOPSIS OF THE THESIS

This study is structured as 6 main chapters. First one is introductory chapter which describes the frame work and the main aim of this study. Detailed information is presented in the following chapters. In Chapter 2, the literature related with the previous cost estimation studies are covered. Chapter 3 gives brief information about the current Light rail and Metro systems in Europe and Turkey. Chapter 4 describes the main components of the trackworks construction and data collection procedure for the analysis. Chapter 5 gives a description about the estimation variables for both regression and ANNs and explains analysis stages. Finally, the summary of the study and the principal conclusions drawn from the comparison of the results of this study are provided in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

As pointed in the previous chapter, early cost estimation of construction projects is quite important due to many reasons and beneficial for parties, owners and the contractors. It is especially significant for the low budgeted developing societies. Therefore, in order to make reliable estimations, there is a necessity to understand the cost modeling techniques that can be used for the construction sector. The previous researches related with early cost estimation techniques and specific study area of this thesis, which is trackworks, should be revealed as intensively as possible so that the affecting parameters of trackworks and type of estimation model decided well. Therefore, in this chapter, the literature available related with model that will be used and cost of trackworks will be presented intensively.

Cost is defined as the total value of parts that is used until a final product is created. Construction cost is calculated by the sum of the product of quantities and corresponding unit prices of them. Because of the reason that the total quantities of the project resources are not change regardless of the construction duration, it is possible to calculate the total cost of the future similar projects (Uğur., 2007). In construction sector, when the idea of making a new structure appeared, the establishment of financial model in a correct way is very important in order to solve cash flow problems and to prevent loss of capital belongs to society. Sometimes, even for the very simple situations, early cost estimations and cost control mechanisms may be required (Polat and Çıracı, 2005). However, by nature this estimation is done, prior to the actual construction of that works. Verlinden et al. (2008) states that cost thus needs to be estimated within a specified accuracy range, although all necessary detailed information is not present yet. To overcome this lack of detailed information, cost-estimation techniques are used in order to approximate the cost within a certain accuracy range. Besides the estimation accuracy issues, cost estimation methods are also required to be rapid in order to ensure a quick response to customers, but also cheap to utilize so that a fairly accurate preliminary cost estimate can be obtained without the need for a detailed design (Caputo and Pelagagge, 2007). In literature, different cost estimation methods can be found, resulting in three main approaches as follows; these are variant-based cost estimation, generative cost estimation and Hybrid cost estimation. (Wierda., 1990). The first approach, variant – based cost estimation for a project, is based on the actual cost of similar projects realized before. That means that historical data of previous projects is needed. Deciding the degree of similarity of these projects is not easy and rather subjective, hence sensitive to mistakes. In addition, the collection of data requires special care. In second approach, generative cost estimation, the cost is estimated by analytical approach, which is based on a detailed analysis of the different ways of construction (method of statements) of the works in the project. Therefore, for each distinct way the direct and indirect cost have to be assigned. Because of this reason, generative cost estimation requires information and will mainly be based on project planning details. The third approach is hybrid cost estimation, which can be used in case some project parts will have detailed information available while others are still at early stages with insufficient data. For the parts with the required data available, generative methods can be used. For those parts still in the earliest stages, variant based methods would be used. Verlinden et al., (2008) stated that the accuracy increases as cost engineers pass from variant based techniques to generative cost estimation and the generation time will increase accordingly. Although different techniques are available for generating cost estimates (Mida et al., 2006), the research in this thesis focuses on multiple regression analysis and artificial neural network (ANN).

2.2 COST ESTIMATION STUDIES

Both regression techniques and ANNs are used frequently in cost estimation fields. Many studies can also be found in literature comparing both methods. McKim (1993) presented the usage of ANNs in cost estimating. The estimates obtained from his study were compared with the estimates produced by three other methods known as pump scaling factor estimates, exponent scaling using the 0.6 rule, exponent scaling using the best exponent, and the best equation. The prediction performances of ANNs were compared with these known methods. McKim observed that ANNs have great potential for estimating non-deterministic costing systems.

Smith and Mason (1999) showed the performance, stability and ease of cost estimation modeling for both methods. They concluded that if little knowledge about the relationships between dependent and independent variables is present, ANNs outperform the more classical regression techniques. However, if the relationship between different variables can be identified, the regression model has advantages in the evaluation of the model and creation of it.

Zhang and Fuh (1998) created an artificial neural network model for early cost estimation of packaging products. As an outcome, they revealed the cost affecting parameters of a product design. The correlation between these parameters and the final cost of the product was discovered by using a back-propagation artificial neural network algorithm depending on historical data.

Garza and Rouhana (1995) examined appropriate usage areas of ANNs for cost estimation purposes. In their study; the cost estimation of carbon steel pipes is done by parametric model and ANNs and their performances were compared. Their study revealed that ANNs have considerable estimation capabilities. However, an ANN has a number of disadvantages: complex neural network architecture design and parameter setting, which both require trial and error.

In the construction sector, Adeli and Wu (1998) formulated a regularization neural network to estimate highway construction costs which were very noisy. They observed that as the number of attributes was increased, the construction cost was estimated more accurately. In another study, a neural network model for parametric cost estimation of highway projects was proposed by using spreadsheet simulation.

Hegazy and Ayed (1998) used ANNs in order to analyze the data of 18 construction projects with ten variables to forecast the final construction cost. They tried to optimize ANN prediction performance by using back-propagation training, simplex optimization and genetic algorithms. It concluded that back propagation training were the most applicable to their data set.

Kim et al., (2004) stated that adequate estimation of construction cost is a key factor in construction projects. In their study, they examined the performance of three cost estimation models. The examinations are based on multiple regression analysis (MRA), artificial neural networks (ANNs), and case-based reasoning (CBR) of the data of 530 historical costs. ANN estimating model gave more accurate estimating results than either the MRA or the CBR estimating models.

Günaydın and Doğan (2004) investigated the utility of neural network methodology to overcome cost estimation problems in early phases of building design processes. Cost and design data from thirty projects were used for training and testing. Neural network methodology with eight design parameters utilized in estimating the square meter cost of reinforced concrete structural systems of 4–8 storey residential buildings in Turkey, an average cost estimation accuracy of 93% was achieved.

Sönmez (2004) established a conceptual cost estimation model for building projects using the data for thirty continuing care retirement projects built by a contractor in the United States. He showed the benefit of using both regression and ANNs in order to reveal the relationship between the variables, total building area, combined percentage area of health center, number of floors, percent area of structured parking etc. He constructed parsimonious models, which was defined as generating models without using the unnecessary variables, to get more satisfactory predictions. In order to eliminate the non contributing variables a step wise regression process is applied by considering the p values of each variables. After establishing the first regression model, the variables that have the highest p value were eliminated one by one. As a result, final regression model has been dev eloped with a reasonable R^2 value (closeness of fit). In addition to regression model, ANNs were established to compare the prediction performance of these two models.

Uğur (2007) studied for forecasting costs of multiple reinforced concrete residential buildings with ANNs, cost of construction of this kind of buildings has been calculated and used as data for an ANN. This network has a multi layer and back propagation structure with adviser to learn. Building elevations, unit numbers in each flat, normal flat area, heights of flats, total flats, outer surface’s empty areas, outer surfaces total areas and average areas of the units in normal flats were assumed as main criteria of the cost of each apartment. Using the data calculated with the ANN, building design parameters for minimum costs have been determined.

2.3 CLOSURE

In the above discussion we conclude that the construction cost estimation can be done with a reasonable accuracy through the methods multivariable regression and ANNs. The objective of this study is to develop and test a model of cost estimating for the trackworks system of Light rail Transportation (LRT) and Metro systems in early design stage via application of regression and artificial neural networks because, there is no previous study related with the early cost estimation of trackworks. The sample data employed for the cost prediction comes from a intensive survey done by contractors and the municipalities in Turkey. The data for the LRT and Metro projects, which have trackworks construction on their scopes, were collected. As a developing country, Turkey experiences rapid population growth especially in big cities. Parallel to, this demand for rapid transportation systems increase considerably. With these considerations, deciding the cost of the trackworks becomes increasingly important issue in developing countries such as Turkey.

CHAPTER 3

URBAN RAIL TRANSPORTATION SYSTEMS

This chapter gives intensive information about the urban rail transportation systems metro and light rail transit (LRT). The descriptions of these systems and the importances of them among other public transportation systems are given intensively. Moreover, the main differences and similarities between metro and LRT systems are demonstrated, because of the reason that the collected historical data comes from the projects, which are classified as either metro or LRT. This chapter also shows the findings of the previous studies regarding metro and LRT systems on the world. One of these studies was done by the European Rail Research Advisory Council (ERRAC, 2004). The aim of the ERRAC study was to give a general overview of urban rail networks (in operation, in construction and planned) and of the rolling stock in order to sketch some general trends for the future development, both for replacements as well as new needs (extensions or new lines requiring additional rolling stock). The research is based upon vigorous data from the most viable sources currently available, mostly first hand, and direct primary sources provided by the operators or cities themselves. The presentation of the results of ERRAC's study is considered to be beneficial in order to understand the importance of cost estimation of urban transportation systems.

3.1 LIGHT RAIL TRANSPORTATION SYSTEMS

As its surname indicates, Light Rail Transit (LRT) is a transit mode. Its middle name reflects the fact that it runs on rails. Boorse (2000) stated that the meaning of light comes from Britain. The term “light railway” is applied to any rail mode that is scaled down from the common size of incline railroads. In previous years, even some of the lines that operated short freight trains pulled by diminutive steam locomotives were classified as light railways. It was not until the 1970s that the term “light rail transit” came into use in the United States. There was no formal definition of LRT at that time, but it was generally understood as an urban rail transit form that is leaner and less costly than other rail modes.

A formal definition was adopted in 1989 and placed in the Transportation Research Board’s Urban Public Transportation Glossary: *“A metropolitan electric railway system characterized by its ability to operate single cars or short trains along exclusive rights-of-way at ground level, on aerial structures, in subways, or occasionally, in streets and to board and discharge passengers at track or car floor level.”*

In addition, ERRAC defined LRT systems as follows: “a tracked, electrically driven local means of transport, which can be developed step by step from a modern tramway to a means of transport running in tunnels or above ground level. Every development stage can be a final stage in itself. It should however permit further development to the next higher stage." This broad definition encompasses a wide array of situations, from conventional tramway, to tram-train solutions. Light Rail systems are thus flexible and expandable. It is not absolutely necessary to have an independent bed track over the whole route; however, the highest degree of segregation from private traffic should be aimed for. LRT systems can be developed from traditional tramway systems or planned and built as entirely new systems. The former option is likely to happen in many developed European cities and the latter option mostly in Western European countries such as Turkey.

3.1.1 HISTORY OF LIGHT RAIL TRANSPORTATION

Taplin (1998) stated that the origin of the tramway is based on horse-drawn wagons and plate ways used in mines. These systems were powered by animals in early stages. However it is obvious that it is not easy to say that animal powered system can be taken as the start of public rail transportation due to the capacity limitation of them. Although, there were systems powered by steam and petrol engines, the effective rail transportation became a reality once electric traction systems were adopted.

The first electric vehicles were battery powered, but it was the development of a practicable dynamo by Werner von Siemens, demonstrated in Berlin in 1879, which provided the way ahead for electric traction by generating power at a fixed point and supplying it to a line by conducting rail or overhead wire.

Siemens & Halske opened the first electric tramway to provide public service in Berlin in 1881, using current at 180 volts fed through the running rails.

Figure 3.1: View of a typical LRT system

Taplin (1998) states that the first lines in the United Kingdom were the Portrush and Bushmills tramway in Ireland, and Volks Railway at Brighton in 1883; the after version of seafront line still runs today. For safety reasons electrified running rails were unsuitable for a street environment, in the UK overhead wire was first used on the Bessbrook - Newry line in Ireland in 1885. Slotted tube overhead was tried in Paris in 1881, and other European cities, including Frankfurt in 1884, and the latter continuous electric street tramway has stated to operation anywhere in the world (conventional overhead wire has been used since 1906). Underground conduit was an alternative to overhead current collection, it was sometimes preferred for aesthetic reasons since poles and overhead was not required, and survived until the end of tramway operation in London in 1952, and in Washington DC until 1962. The Blackpool tramway, operated on the overhead system since 1899, opened with conduit operation in 1885, and is Britain's oldest street tramway still operating today.

3.2 HEAVY RAIL TRANSPORTATION (METRO) SYSTEM

Heavy rail refers to traditional high platform subway and elevated rapid transit lines. Principal characteristics of metros are operation over rights of way that are completely segregated from other uses, with the track placed in subway tunnels, on elevated structures, or on fenced surface rights of way, free of grade crossings with roads. Trains consist of 2 to 12 cars, each with its own motors, and drawing power from a third rail (or in some cases from overhead wire). Boarding is from high platforms that are even with the floor level of the car, allowing large numbers of people to enter and leave rapidly. (APTA, http://www.apta.com/research/stats, last access September 11, 2008)

In addition, ERRAC defines metro as “a tracked, electrically driven local means of transport, which has an integral, continuous track bed of its own (large underground or elevated sections).” This results in a high degree of freedom for the choice of vehicle width and length, and thus a large carrying capacity (above 30,000 passengers per hour per direction – pass/h/dir.). Intervals between stations would be typically more than 1 km, and because the alignment does not have to follow existing streets, curve radii and section gradient can be more generously dimensioned and thus permits for an overall higher commercial speed.

Figure 3.2: View of a typical metro system

3.2.1 HISTORY OF METRO SYSTEMS

London is the mother of all the world's metro systems. Already in 1863 the first tunnel was opened in the city centre for a rail line between Paddington (originally called Bishops Rd) and Farringdon although trains of the Metropolitan Railway were operated initially by steam engines. (Urbanrail, http://www.urbanrail.net/ eu/lon/london.htm, last access September 15, 2008). However, the first "real" metro line was the City & South London Railway, between Stockwell and King William Street (later replaced by Bank) in the City of London, which was opened 4 Nov 1890 and which is part of today's Northern Line. This was the first underground line using electric traction in the world.

Yesilada and Nielsen (1996) stated that before the end of 19th century, new lines were constructed in Glasgow (1896), Budapest (1897), Boston (1897) and Vienna (1898). After London, the next major system was the Paris Metro, whose first line was opened in 1902. Subsequently, a rapid increase on the number of metro systems in the various parts of the world observed. The number of metro systems worldwide increased over time as shown in figure 3.3 below. (Metrobits, 2008, http://mic-ro.com/metro/metrolist.html, last access June 12, 2008)

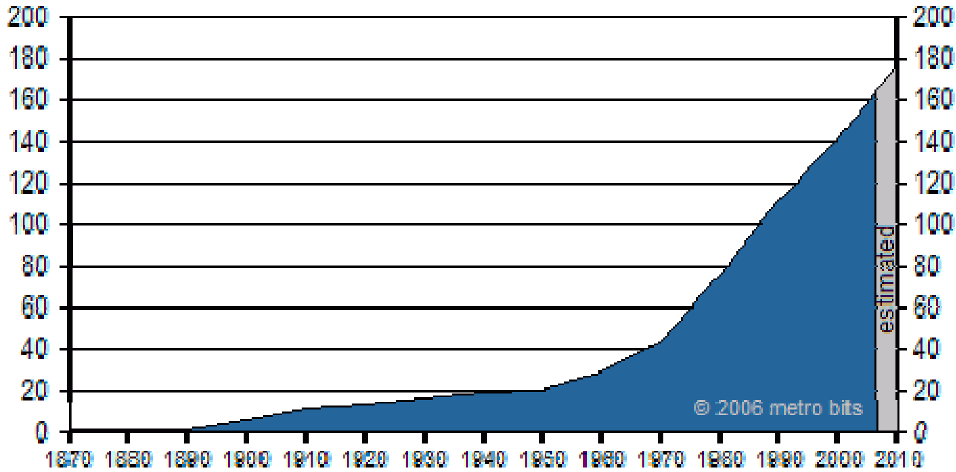

Figure 3.3: Number of metro systems in the world over time as given by Metrobits

3.3 BENEFITS OF LIGHT RAIL AND METRO

Rail transit systems provide economic, social and environmental benefits, and these benefits tend to increase as a system expands and matures. The most important benefits can be listed as follows Tennyson (2000):

- Rail transit reduces transit costs. Costs for rail transit are typically lower, in some cases, as much as 50% less per passenger-mile than for bus transit. This reduces annual transit operating costs and public subsidy.
- Rail transit adds capacity for people. Light rail will go where people live and work and will add greatly to people movement capacity in the corridor. By connecting neighborhoods, downtown and community centers, people may not need to take a car to access transit and may not even need to purchase a car. Other transit serving only a highway corridor must rely on feeder buses or cars to bring people to transit stations which are frequently located in unattractive areas.
- Rail transit is safe. Passengers traveling by rail are very safe and neighborhoods through which the system passes are also much safer. A single light rail vehicle removes 60 to 125 cars from the road, a metro vehicle much more and signal systems make the neighborhoods safer for local traffic and pedestrians.
- Rail transit reduces pollution. Electrified light rail does not pollute near our homes and may not pollute at all in areas relying on water and wind generated power. Getting people out of their cars will preserve our clean air and clean water.
- LRT fits anywhere. it can run on the street, across the street, under the street, over the street, on railroad tracks, or in canal beds. Although it serves communities best when built on the surface, light rail can be run on elevated structures or in tunnels if necessary. Stations and right-of-way are compact and efficient.

- Rail transit strengthens our downtowns. Downtown retail core areas which have struggled for years, choked with automobile traffic and losing business to suburban malls, will see customers return brought to the central business district by an attractive transit system. Non-retail businesses benefit from improved mobility for their workforce.
- Rail transit enhances property values. Areas in the vicinity of light rail stations normally see an increase in land value and new, high-quality transit-oriented development encourages vibrant community centers. New homes and businesses can also reduce the property tax rate in an area.
- Rail transit is quiet. Light rail and metro vehicles produce less noise than diesel or other fuel-burning buses and much less noise than the equivalent volume of automobile traffic. Smooth, welded rails and vibration absorbing fasteners eliminate much of the noise we associate with rail travel.

By means of economy, Alku (2002) pointed out that a Light Rail system saves in building cost when compared to a traditional metro and connecting bus solution. Roughly it can be estimated, that only a quarter of the network is situated in the city centre, where underground lines may be necessary. At 75% of the network the savings in the building cost are 60 to 80% compared to build a kilometer of metro line. The savings in building are based on several features in the line. For the first, for a Light Rail a level crossing with streets is possible saving to construct bridges. The track does not need heavy ground modifications, because same level of curvature and up and downhill are allowed as for streets. Light Rail does not need expensive terminals as stops, instead simple tram stop on street level works. Bus and car connections are easy and inexpensive to arrange, as both operate on same street level with the tram, and buses can share the platform with the tram. On streets with low traffic volume Light Rail can share the street bed with road traffic like ordinary trams, so the only extra cost for Light Rail line is the track and catenary. In operating cost, the basic advantage is the difference between bus and rail transport. The operating cost of a bus is roughly same as the operating cost of one rail unit. But rail units have remarkably higher capacity, which makes the cost per passenger kilometers in rail transport at least half of that in a bus.

Litman (2008) states that high quality transit can increase transit ridership (total number of passengers traveling by public transportation), reduce per capita vehicle traffic, and help stimulate more compact and accessible land use development patterns. A high quality transit service must be relatively fast and reliable, comfortable, convenient, safe, affordable, and it must be integrated with other transportation modes. Rail tends to provide relatively high quality transit service, but also tends to have high initial investment costs. Decision-makers must often determine whether these additional costs are repaid by rail’s greater benefits. This reveals the importance of early cost estimation.

3.4 STATISTICS ABOUT METRO AND LRT SYSTEMS IN EUROPE

Understanding of current operating systems and future plans of LRT and metros is very important to increase the attention on conceptual cost estimation. In orders to demonstrate these figures, previous studies have been investigated through the literature. As Öntepeli (2005) states that one of the most intensive research that presents the current situation of light rail transportation and metro within Europe is done by ERRAC based on the studies performed by International Union of Public Transport (UITP).

The current figure regarding the LRT is presented in ERRAC's research as follows: "In the scope of this research, 170 systems were analyzed for LRTs in Europe, 107 of them can be found within the former members of European Union (EU), 32 within the new Member States and 31 within the countries beyond the EU. (including Norway, Switzerland, but also candidate countries for the EU membership such as Turkey, or the 2nd enlargement wave as well as Western Balkan countries). This group of countries, however heterogeneous as it may seem, has been constituted in order to simplify and ensure a better understanding of results." Within the scope of the above mentioned research it is found that among the 170 tram and LRT systems (941 lines), 63% of systems (107), 48% of lines (448) and 60% of track*km (4793) are in operation within the former EU countries. Germany alone accounts for more than half of these (56 systems and 2768 track*km). It should be noted that 1 km double track described as the term Track*km. The first wave of the enlargement brought another 30 systems (349 lines and

2240 km) into the EU, increasing the total system length of the former members by 46%. Most of the systems are in operation in Poland, the Czech Republic and Hungary. Another 31 systems can be found in countries that will remain outside the borders of the enlarged EU (144 lines and 1027 km).

The current figure regarding metros is presented in ERRAC's research as follows: ERRAC analyzed 36 metro systems in Europe, 27 of them can be found within the former EU members, 3 of them within the new member states joining the EU and 6 within the countries beyond the EU-25 (including Norway and Switzerland but also candidate countries for EU membership such as Turkey as part of the second enlargement wave)". Within the scope of ERRAC research it is found that among the 36 metro systems (138 lines), 75% of systems (27), 85% of lines (117) and 88% of track*km (2072) are in operation within the former EU members. The first wave of the eastern enlargement brought another 3 systems (7 lines and 93 km) into the EU. Another 6 systems can be found in countries that are outside the borders of the EU (14 lines and 181 km). Few cities in Central and Eastern European Countries (CEECs) invested in metro systems. They have, instead, expanded their tramway systems.

3.5 ENLARGEMENT OF LRT AND METRO SYSTEMS IN EUROPE

The trend of constructing new light rail systems in medium and small-sized cities in Western Europe has become progressively stronger. Significantly, the trend has not only been confined to medium and small-sized cities, also metropolis like London, Paris, and Barcelona are pursuing ambitious light rail programs in their respective cities. It should also be noted that firms in Western Europe have acquired a large body of knowledge and experience in successfully implementing light rail. According to the ERRAC's business scenarios (2002), LRT development is expected to double the length of existing systems and increase by 50% the number of LRT systems in Western Europe by 2020. ERRAC's research demonstrates strong evidence of an increase of roughly 40% of the track length; this figure should be considered as a minimum estimate as research so far was unable to find length data (new lines or extensions) for some 30 cities that already have plans at their disposal, and new projects may arise in coming years. On the contrary, major occurrences such as economic downturn or recession and difficulty in funding (financial engineering) may postpone some projects. This data on system extensions will provide ratios to assess additional needs in rolling stock. If we have a look at the number of cities with LRT, then the increase amounts to 55%. Still this evidence demonstrates that the initial ERRAC forecast for the track length may have been overoptimistic. On the other side, figures show that the ERRAC forecast for the number of cities equipped may have been slightly conservative.

Cities in former members of EU have the lion's share of metro development with approximately 83 % of all new construction, while new Member States and beyond EU cities account for only 5 % (7 km) and 12 % (16.5 km) in the construction of extensions. For planned metro system growth, cities in former members of EU account for 60% of all planned schemes, with an additional 9

% (45 km) in the new Member States and 31 % (155 km) in the beyond EU countries. At his point, the dominant role of Turkey should be pointed out. In Turkey, among these schemes, 3 systems are being built and 5 are planned in cities which do not currently offer metro services.

3.6 CONCLUSION

It is clear that all of the countries surveyed recognize the long-term benefits of enhancing the transit infrastructure and uniformly exhibit the political will to deploy the necessary resources to capture those benefits (Bottoms, 2004). In order to manage and get these necessary resources, the required feasibility studies should be done for LRTs and metros. For an investor, one of the most important assets in feasibility stage of a new line or in extension of an existing line is the understanding of overall project cost effecting parameters, which are site conditions, number of stations, rolling stock, signalization, type of the civil structure (tunnel, at grade, elevated etc.), track works etc. By stating these parameters precisely, the ideal initial budget can be generated reasonably. In literature, there are given and applied many cost estimation studies for several cost affecting parameters. However some of them did not take the attention of researchers. Early cost estimation of trackworks, which is common for metros and LRTs, is one of those kinds that are not much in interest. That is the reason why this thesis study is objected to perform cost prediction analysis by using historical data of realized projects in Turkey. The components of trackworks and data collection procedure are discussed in the next chapter.

CHAPTER 4

TRACKWAY TYPES, COMPONENTS AND DATA COLLECTION

4.1 INTRODUCTION

A railway can be defined as an engineered structure consisting of two metal guiding rails on which cars are self-propelled or pulled by a locomotive (Arema, 2003). Armstrong (1998) defines a railway as: “A railroad consists of two steel rails which are held a fixed distance apart on a roadbed. Vehicles, guided and supported by flanged steel wheels and connected into trains, are propelled as a means of transportation”. Webster’s Dictionary defines a railroad in three different ways: “1. an assembly of rails, sleepers and fastenings over which cars, locomotives and trains are moved, 2. a complete system of such roads, including land, rolling stock, stations, etc. 3. The persons or corporation owning and managing such a system” (Webster’s Dictionary, http://www.websters-online-dictionary.org, last access 12 June 2008). Within this study, the term track refers railway infrastructure according to the first definition of Webster’s Dictionary.

The track is a basic component railway infrastructure and represents the main distinction between land transportation and all others, because it provides a fixed guidance system. “The track, which is the steering base for the train, has evolved from an ancient design of vehicle guidance (wooden rail) with origins dating, some historians have suggested, from the Sumerian culture of 2000 BC.” (Railway Technical, http://www.railway-technical.com/track, last updated 24th June 2008). In today’s world, modern railways are based on the steel wheel running on a steel rail. Different types of guided vehicle technology exist such as rubber-tyred systems, magnetic based and guided busways (metrobuses). However, these forms of guidance technologies are not considered within the scope of this study.

In order to maintain a comfortable ride with train, the track alignment has to be set to within a millimeter of the design. “Track design and construction is a complex and multi-disciplinary engineering science involving earthworks, steelwork, timber and suspension systems. There are many different systems throughout the world and many variations exist in their performance and maintenance” (Railway Technical, http://www.railway-technical.com/track, last updated 24th June 2008). Undoubtedly, the most widely used systems in urban areas are LRTs and metros.

This chapter presents the fundamental types and basic components of modern trackway and construction with drawings, photos and examples from around the world. These particular components of trackway of LRT or metro systems correspond to the majority of the variables to be used for ANN and regression analysis in the next chapter.

4.2 FUNDAMENTAL TYPES OF TRACKWAY

The track is the most visible part of a railway route but there is a sub-structure supporting the track which is equally as important in ensuring a safe and comfortable ride for the train and its passengers or freight. There are two types of sub structures, which are called ballasted and direct fixation (non-ballasted) trackways respectively. The figure 4.1 below is a representative sketch of double-track line with its fundamental parts. In general, the total width across the double-track alignment is about 15 m for modern LRT and metro systems. “The ‘cess’ shown each side of the alignment is the area available for a walkway or refuge for staff working on the track” (Railway Technical, http://www.railway-technical.com/track, last updated 24th June 2008).

As information; if the line is electrified on the overhead system, catenary masts are placed outside the drains and, beyond them, there is a walkway area. This may just be a cleared path for staff to walk safely, avoiding passing trains or, on modernized routes, a properly constructed path. Next to this path will be a cable passing through. These were originally concrete but are nowadays often made of plastic. Cables crossing the track are protected by a plastic tube, usually bright orange in the UK. Proper cable protection is essential to prevent damage by animals, track maintenance tools, weather and fire. The detailed information about ballasted and direct fixation types of trackway will be given in following section.

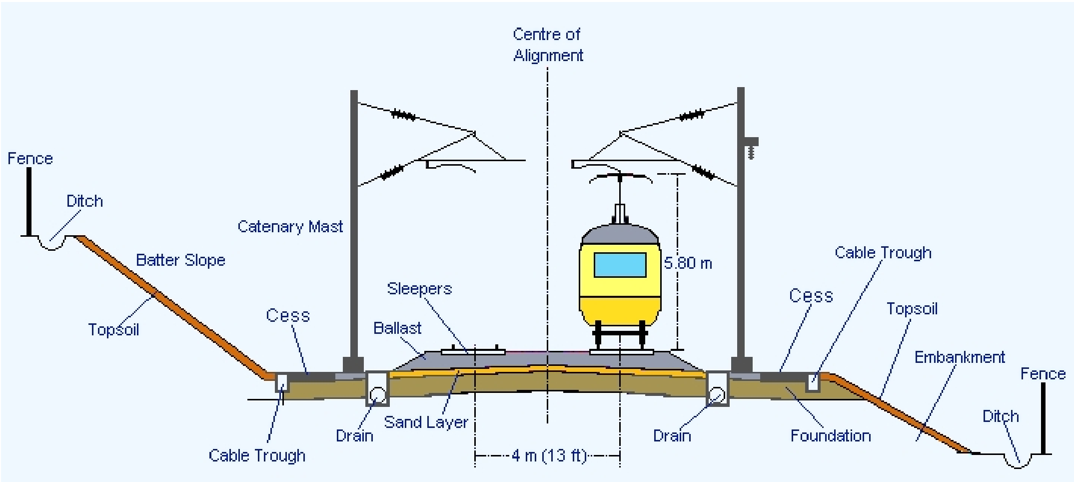

Figure 4.1: Typical cross section of double track railway

4.2.1 BALLASTED TRACKWAY

The ballasted track structure is made up of subgrade, sub-ballast, ballast, ties and rail as illustrated in Figure 4.2. “Each of these contributes to the primary function of the track structure, which is to conduct the applied loads from train traffic across the subgrade safely” (Arema, 2003).

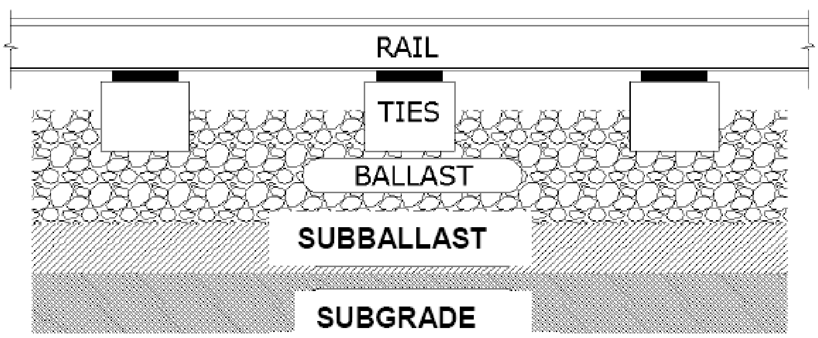

Figure 4.2: Longitudinal cross-section of ballasted trackway

The subgrade is the ground upon which the track will be laid. Natural ground level can be selected as subgrade, or it can be an embankment or cutting. “It is important that the subgrade is made of the right materials and is properly compacted to carry the loads of passing trains. The subgrade under the track has a "camber" rather like that seen on a roadway. This is to ensure ease of water run-off to the drains provided on each side of the line. The track itself is supported on "ballast", made up of stones usually granite or, in the US, basalt - below which is a layer of sand, which separates it from the subgrade” (Railway Technical, http://www.railway-technical.com/track, last updated 24th June 2008). A detailed ballasted trackway cross section can be seen in figure 4.3 below.

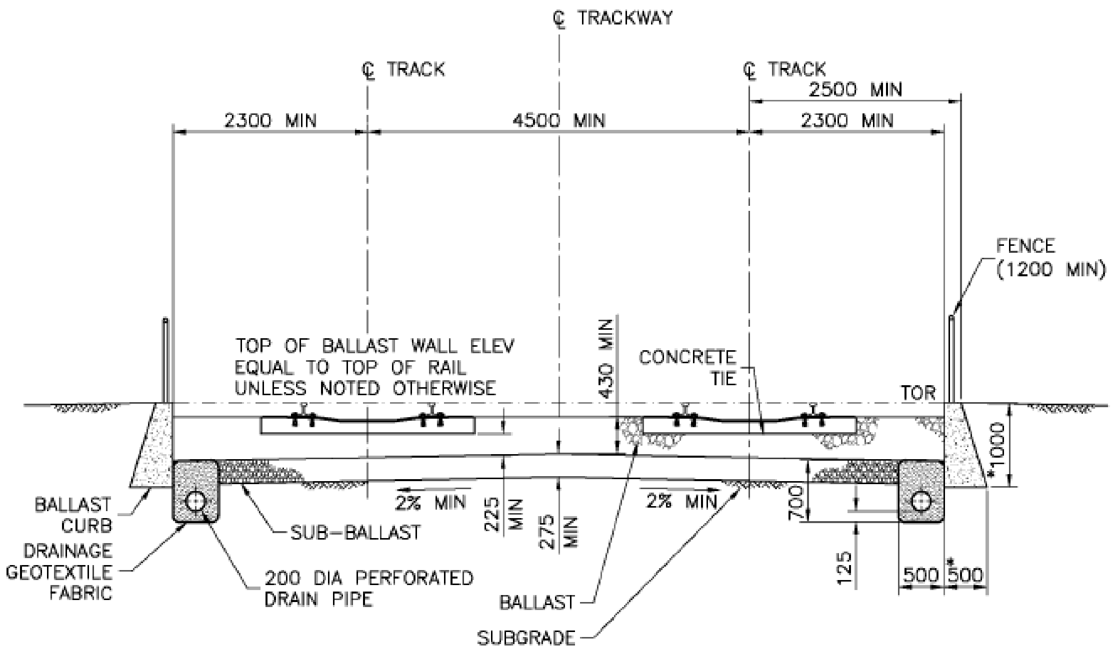

Figure 4.3: A detailed cross-section of ballasted trackway

4.2.2 DIRECT FIXATION TRACKWAY

Direct fixation term, which is a track type, refers to all rail and special track components and their fastenings. Those items embedded in concrete plinths such as inserts, embedded timber or concrete blocks specifically for track fastenings, and any special bolts for holding the fastening blocks are all part of the direct fixation track. On the other hand, Daniels and Tuten (2001) state that ballasted track is generally understood to include subballast, ballast, sleepers, fasteners and rail. There are two main reasons for implementing direct fixation track instead of ballasted track. Firstly, it is suitable for the constrained clearances cases. Secondly, it reduces the superstructure load. Therefore, it is preferred to be used on elevated structures (bridges) or tunnels. Direct fixation track is also more adaptable for street at grade crossing in the cities. Representative drawings and pictures of direct fixation line can be seen in figure 4.4, 4.5 and 4.6.

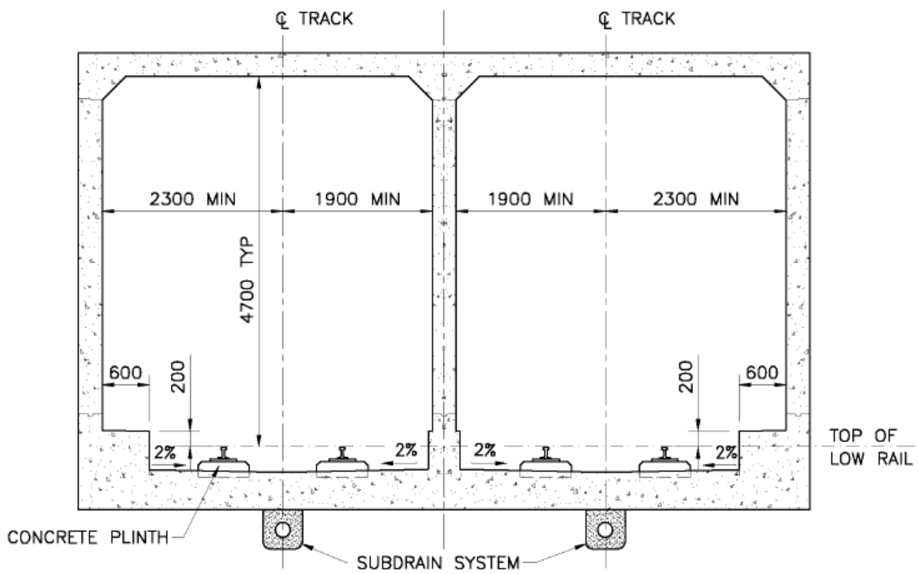

Figure 4.4: Cross-section of direct fixation trackway on tunnel

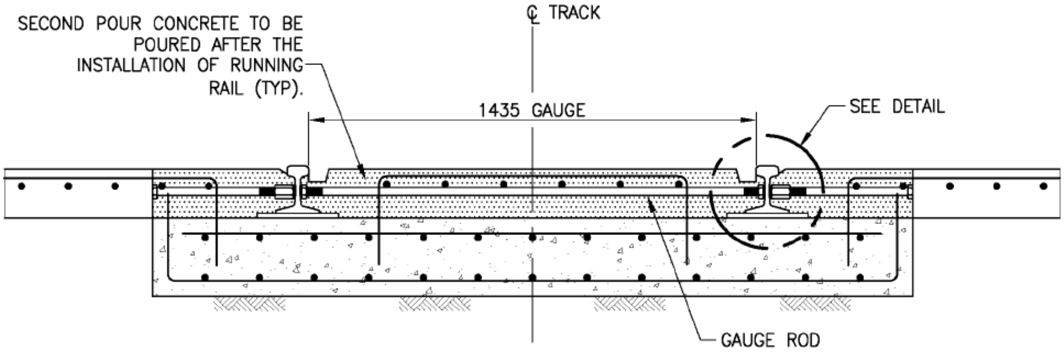

Figure 4.5: Cross-section of direct fixation trackway at street crossing

Figure 4.6: Direct fixation track on segregated open area

As mentioned previously, Daniels and Tuten, 2001 states that ballasted track is any track containing ballast. Ballastless track is everything else. So, there are subcategories of direct fixation also. These are as follows:

- Discrete fasteners bolted to the support
- Embedded rail track
- Continuously supported rail
- Embedded block track and embedded tie track

As a result, direct fixation track is a system where the rail fastens directly to the track. The most of the common form of direct fixation track is that with discrete fasteners bolted to the support. This form typically has fastener body, hold down bolts (see Figure 4.7) and rail clips. Within this study, the projects were classified as direct fixation or ballasted tracks in the lights of above considerations. It should be noted that several projects have both types of trackways in their scope.

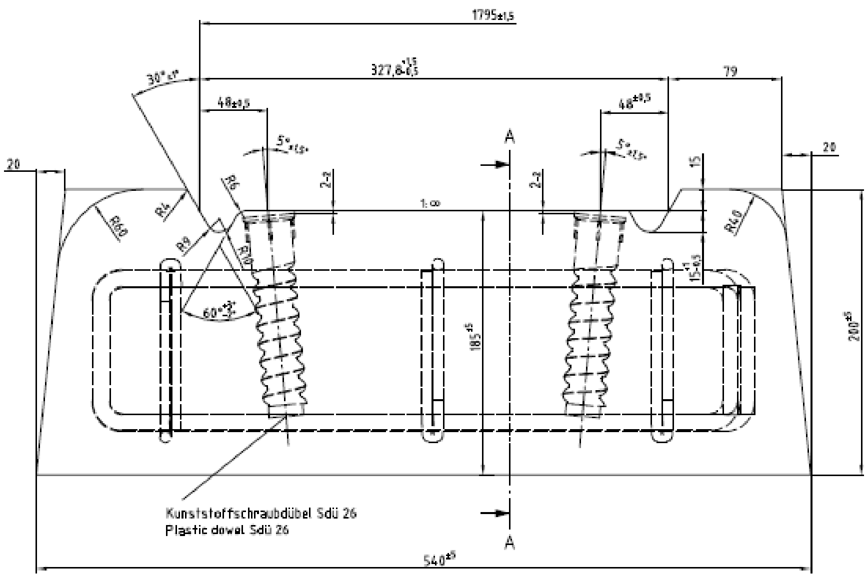

Figure 4.7: Hold down bolts

4.3 COMPONENTS OF TRACKWAY

4.3.1 CONCRETE TIES

Concrete ties (sleepers) can be described as rail support elements that transfers the static and dynamic load of trains rails to ballast section. They are placed perpendicular to rail direction with an equal spacing (see Figure 4.8), but the spacing may change from project to project. Although, there are many types of ties using on railways, concrete ties are the most widely used bearing elements. It should be pointed out that all the projects, from which the realized data are collected within the scope of this thesis, are using concrete sleepers as well. “Concrete ties (see Figure 4.9) are rapidly gaining acceptance for metros and heavy haul mainline use, (both track and turnouts), as well as for curvature greater than 2°. They can be supplied as crossties (i.e. track ties) or as switch ties. They are made of pre-stressed concrete containing reinforcing steel wires. The concrete crosstie weighs about 600 lbs. vs. the 200 lb. Timber track tie” (Arema, 2003). The concrete tie utilizes a specialized pad between the base of the rail and the plate to cushion and absorb the load, as well as to better fasten the rail to the tie. Failure to use this pad will cause the impact load to be transmitted directly to the ballast section, which may cause rail and track surface defects to develop quickly. An insulator is installed between the edge of the rail base and the shoulder of the plate to isolate the tie (electrically). An insulator clip is also placed between the contact point of the elastic fastener used to secure the rail to the tie and the contact point on the base of the rail.

Figure 4.8: Preview of a track with concrete tie

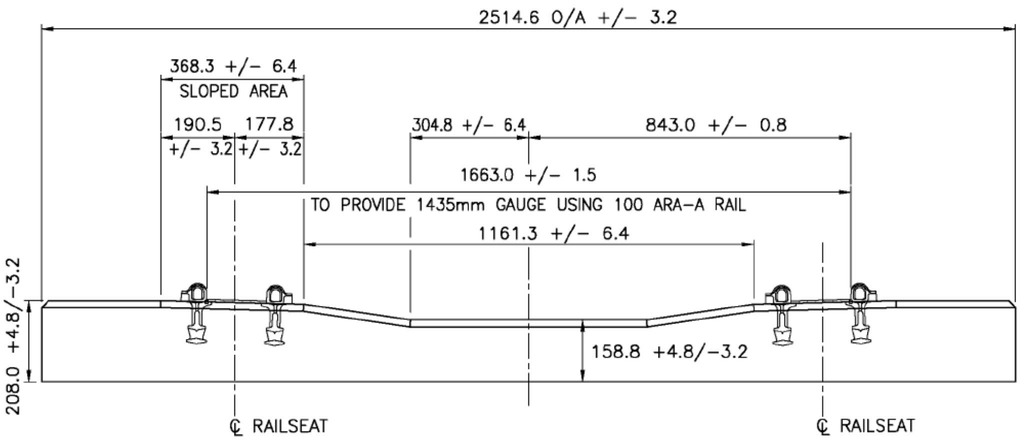

Figure 4.9: Concrete sleeper with dimensions

4.3.2 RUNNING RAILS

Running rails are the most important part of the trackway, because they are directly in contact with the train wheels. The main special duty the rail is to carry huge amounts of concentric loads. That's why they have to be produced with a special care and material. David (2004) described the running rail as hot rolled steel in the profile (cross section) of an asymmetrical I-beam is usually used as the surface on which railway wheels run. The standard form of rail used around the world is the "flat bottom" rail (see Figure 4.10). It has a wide base or "foot" and narrower top or "head". Unlike some other uses of iron and steel, railway rails are subject to very high stresses and have to be made of very high quality steel alloy. Rail is graded by weight over a standard length. Heavier rail can support greater axle loads and higher train speeds without sustaining damage than lighter rail. Therefore, it has a great effect on the standards of LRT and metro lines. Because of this reason, in order to observe the cost effectiveness of rail to the final cost of the projects, the consumption of running rail during the construction of LRT and metro projects has to be dealt with.

Figure 4.10: Flat bottom rail

4.3.3 RAIL WELDING

Rail welding made possible the installation of continuous welded rail. Before the application of welding, rails were joined by bolted plates, called fishplates, leaving a space between for expansion. However, by the invention of thermite welding, many well established associated benefits were emerged. “Reduced bolt hole rail failures and bolted joint maintenance, increased rail life, better track circuit reliability, reduced equipment wear, a better ride quality and reduced track maintenance costs are among these benefits that can be directly attributed to thermite welding” (Hauser, 1978). The development of continuously welded rail was undertaken in Europe during the 1950’s and 1960’s and has been progressively introduced into Turkey since that time until now it is the standard practice in LRTs and metros. All of the rail tracks studied in this study are constructed using this technique. The figure 4.11 demonstrates the application of thermite welding process.

Figure 4.11: Thermite welding application photo taken by Öztürk.E

Thermite welding is a welding process, which produces coalescence of metals by heating them with superheated liquid metal from a chemical reaction between metal oxide and aluminum with or without the application of pressure. Filler metal is obtained from an exothermic reaction between iron oxide and aluminum. The temperature resulting from this reaction is approximately 2500° C. The superheated steel is contained in a crucible located immediately above the weld joint. The superheated steel runs into a mould, which is built around the parts to be welded. Since it is almost twice as hot as the melting temperature of the base metal, melting occurs at the edges of the joint and alloys with the molten steel from the crucible. Normal heat losses cause the mass of molten metal to solidify, coalescence occurs, and the weld is completed. (AAR, 2003)

4.3.4 SPECIAL TRACKWORK

Special trackwork is defined as trackwork structures, trackwork components or fittings that are normally fabricated in whole, or in part, from regular rolled rail section. In general, the following components are included in special trackwork:

- Turnouts and crossovers
- Diamonds
- Guard rails
- Expansion or sliding rail joints
- Lateral restraining devices required at structural interface elements

Usually, all special trackwork are located on tangent track and constant profile grade. Special trackwork located on curves require unique customized design and are difficult to fabricate and maintain. Fabrication and on-site installation variables associated with special trackwork in curves may also compromise operating safety of the system. They are, placed on concrete ties on ballasted track and incorporated on a direct fixation system on concrete slab track.

Simple turnouts and crossover systems (as part of the special trackwork) are costly systems compared to other components of track. Therefore, the number of turnouts and the crossover in a project should also be considered for the conceptual cost estimation purposes. In this study, a simple turnout (see Figure 4.12) refers to mechanical structure enabling railway trains to be guided from one track to another at a railway junction. On the other hand, a crossover (see Figure 4.13) refers to a pair of switches that connects two parallel rail tracks, allowing a train on one track to another.

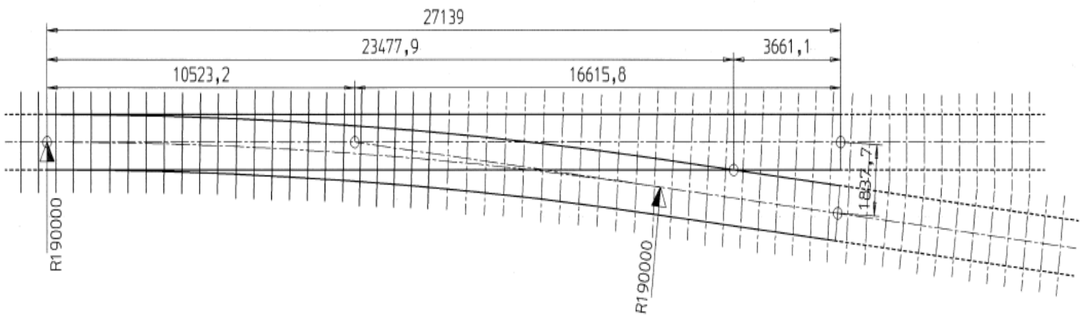

Figure 4.12: Simple turnout layout

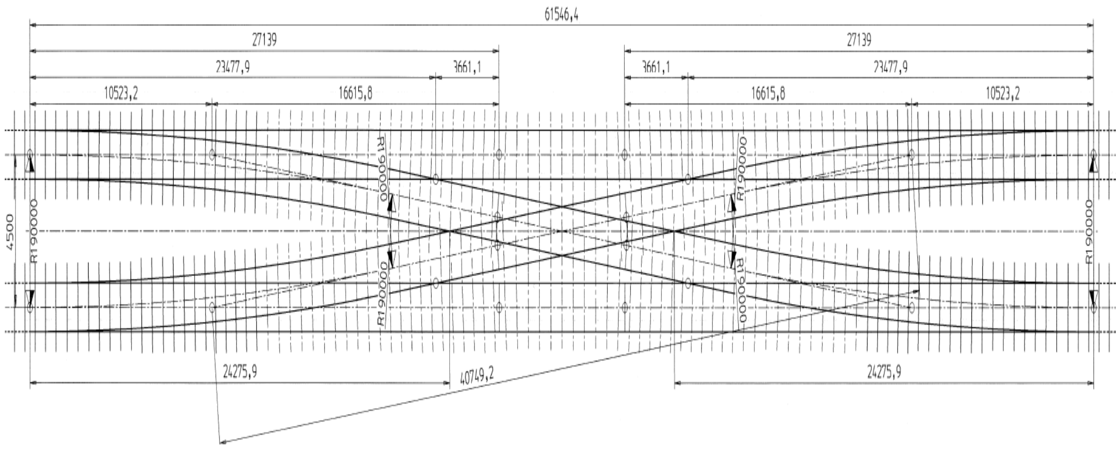

Figure 4.13: Crossover layout

Figure 4.14: A crossover picture

4.4 TRACKWAY GEOMETRY

The route upon which a train travels and the track is constructed is defined as an alignment. An alignment is defined in two fashions. First, the horizontal alignment defines physically where the route or track goes (mathematically the XY plane). The second component is a vertical alignment, which defines the elevation, rise and fall (the Z component). Alignment considerations weigh more heavily on railway design versus highway design for several reasons.

First, unlike most other transportation modes, the operator of a train has no control over horizontal movements (i.e. steering). The guidance mechanism for railway vehicles is defined almost exclusively by track location and thus the track alignment. The operator only has direct control over longitudinal aspects of train movement over an alignment defined by the track, such as speed and forward/reverse direction. Secondly, the relative power available for locomotion relative to the mass to be moved is significantly less than for other forms of transportation, such as air or highway vehicles. (Arema, 2003)

These factors result in much more limited constraints in design stage when considering alignments of small terminal and yard facilities as well as new routes between distant locations. Therefore, “when establishing alignments, the type of train traffic (freight, passenger, light rail, length, etc.), volume of traffic (number of vehicles per day, week, year, life cycle) and speed should be taken into account” (Allen, 1920). The design criteria for a new coal route across the prairie handling 15,000 ton coal trains a mile and a half long ten times per day will be significantly different than the extension of a LRT and metro line.

“Suffice it to say that in todays environment, the designer should also add to the decision model environmental concerns, politics, land use issues, economics, long-term traffic levels and other economic criteria far beyond what has traditionally been considered” (Arema, 2003). These added considerations are well beyond what is normally the designers task of alignment design, but they all affect it. The designer will have to work with these issues occasionally, dependent upon the size and scope of the project. “On a more discrete level, the designer must take the basic components of alignments, tangents, grades, horizontal and vertical curves, spirals and superelevation and construct an alignment, which may effective the cost of construction, easy to maintain, efficient and safe to operate" (AAR, 2003).

Because of these reasons, cost effectiveness of these geometrical design criteria to the final cost is also investigated in this study by adding several of them as variable to model such as, maximum horizontal and vertical curvature, maximum slope of the line and maximum super elevation.

4.5 DATA COLLECTION

The preceding studies carried out by various researchers in very wide spectrum for the early cost estimation related with the different construction projects, the reasons and outcomes of these deficiencies have been reported in Chapter 2. It is observed that in order to create a cost estimation model, majority of studies in Turkey have been carried out cost for the civil scope in other words, for structural works like construction of residential buildings, tunnels, bridges etc. However, nowadays, the growing market for urban rail transportation in Turkey requires an early cost estimation study for the construction of trackways, which is the unique part of LRT and metro systems. Data of completed projects in Turkey made this study possible. Nevertheless, it is not easy to collect the data of projects from contractors or clients, which are municipalities in Turkey. Therefore, some of the researchers made case studies instead of collecting data. The others made surveys to get the required data by sending the survey forms but, in that case, response to the questions usually takes considerable amount of time and sometimes impossible. Besides, there is always a possibility of misunderstandings about the questions.

Initially it was attempted to get realized final cost information through e-mail from the client and contractors. Because of the difficulties faced particularly in obtaining the necessary data via e-mail to carry out this sort of analysis, well planned face to face interviews have been implemented with the sources as second attempt. The reason why the idea of getting information by means of e-mail is not appreciated by the contractor companies except 2 of them out of 16 was the confidentiality issue. The administrators of these contractors were worried about leakage out this valuable cost realization information with their names. Therefore, it is guaranteed that the names of the respondents or companies would not be revealed under any circumstances in order to get answer during the face to face interviews, which were made in cities, Istanbul, Bursa, Eskisehir, Kayseri, Konya, İzmir and Ankara. In addition to this guarantees, the references of advisor professor of this study and other valuable professionals from the industry were effective on the response rate.

The first and one of the most important steps in collection of data was to decide on method, which has been explained above. The second step in such a study is to decide on sample size, which affects the analysis validity directly. Therefore, the sample size is tried to be expanded as many as possible. A target project list was formed by conducting a small investigation for the existing and under construction projects, which have trackworks in their scope.

It should be noted that projects under construction, but do not have track works in their scope such as Ankara metro extension 4 (Tandoğan – Keçiören) were not dealt with in this study, because the construction of these trackways has not started yet. The final step was to define the cost effecting parameters of the trackways of LRT and metro lines. In the light of the experiences and suggestions of professionals working on trackworks, parameters were decided.

The data for this study were collected from 16 urban rail projects physically (in place) during one year period. Several of these projects are LRT and the others are metros. Because of the reason that trackway construction is common both LRT and metro projects, these systems have been analyzed in the same manner in this study. Table 4.1 shows the lists of these projects. In this table, abbreviation M and L is used for metros and LRTs respectively. Totally, 7 metro and 9 LRT project data are achieved to gather and investigated in Chapter 5.

Table 4.1: List of projects

Number	Project	Type	Location	Lenght of the Trackway (m)
1	M1	Metro	Ankara	31081
2	M2	Metro	Ankara	29543
3	L1	LRT	Antalya	22768
4	L2	LRT	Bursa	17027
5	L3	LRT	Bursa	4856
6	M3	Metro	İzmir	23200
7	M4	Metro	İzmir	10486
8	M5	Metro	İzmir	3300
9	L4	LRT	Eskisehir	31454
10	L5	LRT	Kayseri	34460
11	L6	LRT	Konya	3269
12	L7	LRT	İstanbul	22400
13	L8	LRT	İstanbul	11496
14	L9	LRT	İstanbul	25200
15	M6	Metro	İstanbul	37000
16	M7	Metro	İstanbul	16000

CHAPTER 5

METHODOLOGY AND DATA ANALYSIS

5.1 GENERAL

The main objective of this study is to construct a cost estimation model with the help of regression and artificial neural network (ANN) methods by using the realized data and to reveal the cost effective parameters of a trackway. In Chapter 4, an overview of LRT and metro tarckway types and most important parts of trackway is given. The literature about above mentioned methods, basic parameters to construct analysis, data identification and the steps of regression and ANNs analysis will be given within this Chapter.

5.2 METHODOLOGY LITERATURE

5.2.1 MULTIPLE REGRESSION METHOD

The term regression was first used by Pearson (1908) who states that the main aim of using regression is to reveal the relationship between several independent or predictor variables and a dependent or criterion variable. Hill and Lewicki, (2007) states that researcher's one of the main the question "what is the best predictor of ..." is able to be answered by multiple regression. That's why, multiple regression procedures are very widely used in researches on both social and natural sciences.

According to the recent citations from literature its usage area includes the following diverse applications: software development costs, roads in rural part of developing countries, query costs in data bases, urban water supply projects and design for manufacturability. Mason and Smith, (1997) showed that professional cost estimators regularly use regression to build their cost models. Because of its strong mathematical background, regression analysis, being a cost estimation technique, has been used since the 1970's. However, Verlinden et al. (2008) stated that although applied frequently, some drawbacks of regression techniques should be taken into account. Firstly, there is no general approach to help the cost engineer in choosing the model best fits historical data for his specific problem. Secondly, when using regression techniques, the type of relationship between variables must be assumed a priority. Thirdly, the number of input variables is limited to some extent. Regression models should be generated by considering above mentioned facts.

When evaluating the regression models also there are parameters need to be checked. The main two important parameters are significance level (P value) and coefficient of determination (R^2 value). Because, R^2 value expresses the variability in the output that can be explained by the variables included in the model and P value shows the significance of the variables included in the model. Generally, multiple regression models can be represented in the form of;

$Y = C + b1X1 + b2X2 + \cdots + bnXn;$

where Y is the total estimated cost, and X1; X2; : : : ; Xn are measures of distinguishable variables that may help in estimating Y . C is the estimated constant, and b1; b2; : : : ; bn are the coefficients estimated by regression analysis, given the availability of some relevant data.

In this study, a statistical software, Minitab, were used to develop the regression model. Multicolinearity of the variables is checked and the step-wise technique, based on the p values limitation, followed by using this program. These steps will be presented in the following sections of this Chapter.

5.2.2 ARTIFICIAL NEURAL NETWORK MODEL

Artificial neural networks (ANNs) are currently used to generate cost estimations as an alternative for regression techniques. Application of ANNs to enhance the accuracy of cost estimation by not being stuck within the limitations of regression has discovered by considerable number of researches. Verlinden et al. (2008) observed that ANNs are applied in many fields such as financial services, biomedical applications, time-series prediction, text mining, decision making and many others. Although, there are numerous applications of ANN, they all share an important common aspect: the processes to be predicted are correlated with a large number of explanatory variables and there may be high-level non-linear relationships between those variables. The most important aim of the ANNs is to find those nonlinear relationships to achieve better estimation.

Kim et al. (2004) describes artificial neural network as a computer system that simulates the learning mechanism of the human brain. The main structure of ANNs is based on a number of neurons, which are grouped in one or several hidden layers. Neurons in these layers are connected to each other by a weighed function called transfer function. According to the contribution of the each neuron to the final output, the output weight of neurons changes in every iteration process. Figure 5.1 shows typical neural network architecture.

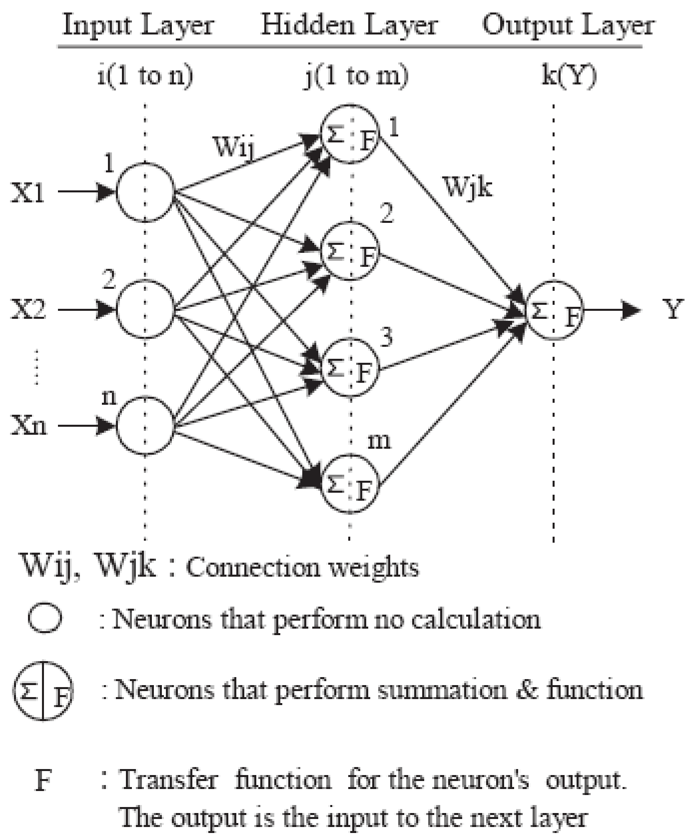

Figure 5.1: Typical ANN architecture given by Kim et al.

Above mentioned design parameters determine to the performance of the ANNs considerably and will differ depending on the field of application. The number of hidden neurons and number of hidden layers have a great influence on detection capacity of ANNs for dependency between variables. However, there is no solid rule in determining of these parameters. This feature is the one of the biggest drawbacks of ANNs shown in the literature. In addition, the parameters called learning rate and the momentum rate that affects the weight updating rule of ANNs are not also fixed values. All these parameter are decided by trial and error procedure, which takes considerable amount of time. However, in literature, several proposals are present that makes possible to limit the range of these parameters. So, training algorithm of ANNs can be chosen with a reasonable effort.

Hegazy et al. (1994) proposed that one hidden layer is sufficient to generate an arbitrary mapping between inputs and outputs and the number of neurons in the hidden layer is 0,75m, m, or 2m + 1, where m is the number of input neurons. That's why, ANN models, which contains three different numbers of hidden neurons, were performed in this study.

In literature, there are number of cost estimating models have been developed adopting the back propagation algorithm (Rumelhart, 1986). In this research, a back-propagation network (BPN) was used to estimate the trackway construction cost. A BPN usually incorporates a non-linear sigmoid transfer function to calculate the output of each neuron except for the input neurons. Kim et al (2004) states that the output of each neuron is modified by the sigmoid transfer function, which defines the output of each neuron in the given form in Equation (1) and the that of each output neuron in the form given in Equation (2) ;

$$f(x_j) = 1 \Bigg/ \left(1 + \exp\left(-\left(\sum_{1}^{n} X_i \cdot w_{ij} - \theta_{ij}\right)\right)\right), \qquad (1)$$

$$f(x_k) = 1 \Bigg/ \left(1 + \exp\left(-\left(\sum_{1}^{m} X_j \cdot w_{jk} - \theta_{jk}\right)\right)\right), \qquad (2)$$

where Xi is the value of the input variable, wij and wjk are connection weights between the input and the hidden neuron and between the hidden neuron and the output neuron, respectively, θij and θjk are the bias terms for the ith and kth neuron, respectively, and i, j, and k are the number of neurons in each layer.

“Back propagation (BP) methodology is working according to a learning algorithm called generalized delta rule, which performs a gradient descent in the error space to minimize the total error between the estimated costs and the desired costs of the output layer to update the connection weights” (Kim et al, 2004). In ANNs, the weights are updating in accordance with the coefficients of the learning rate and momentum.

Hegazy et al. (1994) also proposed that the coefficients of the momentum and learning rate can be set to 0.9 and 0.7, respectively. In the light of this proposal, Kim et al, (2004) were conducted ANN analysis by changing these parameters in a range which covers Hezagy’s proposal and got reasonable results. That’s why, in this study these coefficients were set between 0.5 and 0.9 (in steps of 0.1) to examine their effect and establish the best NN model. Numerous ANN models were evaluated by changing the number of neurons in the hidden layer according to previously proposed rule and by changing the coefficients of momentum and learning in steps of 0.1.

5.3 DATA IDENTIFICATION

Variables that best describes the trackway cost is tried to select with a special attention. While selecting these variables, the experiences of the professionals working on this subject are taken into consideration. For the majority of the projects, which were selected, data of trackworks was collected successfully. In collection stage, it was very important to explain the scope of this study. Due to the reason that this study is dealing with neither the structural parts of the projects, nor the electrification and scada systems of the line. Therefore, the total cost data represents the required money to construct a trackway from the top of the subgrade level to the top of the rail. It should be noted that all cost data is taken from the companies in the same currency, which is US dollars. That’s why; the conversion of the cost to a single currency by using the exchange rate of Central Bank of Turkey was not required.

The cost of the trackway system are affected by numerous variables. The main issue is to find the independent variables. Generally, independent variables were classified into two main groups which are dummy and continuous variables (Öntepeli, 2005).

A dummy variable in statistical analysis represent each category as an integer. For example if there are categories 'small', 'medium' and 'large', it can be represented by giving integers to these like 'small' = 0, 'medium' = 1, 'large' = 2. Although, in general this idea should work fine when there are only categorical variables. In neural network analysis, it is needed to pay attention to scaling when the data set contains numerical variables also. However, there exists no categorical variable in the data set of this study. It is better to state here that the projects were not classified as metro and LRT systems. In other words, categorical variables for this type of classification were not used. This is because of the fact that there is not a considerable difference between the metro and LR system projects included in this study and listed in Table 4.1, in terms of design and construction methods applied. That's why it is avoided to use categorical variables in neural network and regression and because of the complex scaling procedure for categorical variables.

Nevertheless, the fundamental properties of the projects of which data to be analyzed were to correspond to the continuous variables, as the independent variables. The following parameters were identified to represent the main characteristics of the projects in the regression and neural network analysis, such as lengths of ballasted trackway sections, number of concrete ties/ thermite welding and the other particular sections included in a trackway construction as explained in detail in Chapter 4:

i. LTT (Total Length of Main Trackway – Meters)

This parameter is corresponding to the total length of the main line (single trackway) included in the related project.

ii. LBT (Length of Ballasted Trackway - Meters)

This parameter is corresponding to the length of the ballasted track (single trackway) included in the related project.

iii. LDF (Length of Direct Fixation Trackway - Meters)

This parameter is corresponding to the length of the direct fixation track (single trackway) included in the related project.

iv. NC (Number of Crossover)

This parameter is corresponding to the number of crossovers installed in the related project.

v. NST (Number of Simple Turnout)

This parameter is corresponding to the number of simple turnouts placed in the related project.

vi. SS (Sleeper Spacing - cm)

This parameter is corresponding to the spacing of the sleepers (ties) used in the related project.

vii. WC (Workmanship Cost - $)

This parameter is corresponding to the workmanship cost per meter of trackway in the related project.

viii. WR (Total Weight of Rail – Kg)

This parameter is corresponding to the total weight (Kg) of the rail used in the related project.

ix. NTW (Number of Thermic Welding)

This parameter is corresponding to the number of thermic welding done in the related project.

x. **NS (Number of Sleepers and Concrete Blocks)**

This parameter is corresponding to the number of sleepers and concrete blocks placed in the related project.

xi. **HPC (Hourly Passenger Capacity – passenger/hr/direction)**

This parameter is corresponding to the maximum hourly passenger capacity per direction of the related project.

xii. **MOS (Maximum Operation Speed – Km/hr)**

This parameter is corresponding to the maximum operational speed of the related project.

xiii. **CS (Commercial Speed – Km/hr)**

This parameter is corresponding to the commercial speed of the related project.

xiv. **MSL (Maximum Slope of the Line - %)**

This parameter is corresponding to the maximum allowed slope of line of the in the related project.

xv. **MS (Maximum Superelevation – Cm)**

This parameter is corresponding to the maximum allowed super elevation (tilting the trackway) the in the related project.

xvi. **MHC (Minimum Horizontal Curvature – Meters)**

This parameter is corresponding to the radius of minimum allowed horizontal curvature of the trackway the in the related project.

xvii. **MVC (Minimum Vertical Curvature - Meters)**

This parameter is corresponding to the radius of minimum allowed vertical curvature of the trackway the in the related project.

xviii. **C (Cost)**

This parameter, which is the dependent variable, is represents the total final cost of the trackway part of the related project.

5.4 APPLICATION

5.4.1 REGRESSION APPLICATION

The application of regression analysis was performed using Minitab, which is a statistical program with a spreadsheet-like data worksheet. It is capable of manipulating and transforming this data and can produce graphical and numerical summaries. Minitab also allows performing a wide variety of statistical computations.

In this study, regression analysis is used to investigate and model the relationship between a response variable and one or more predictors. Minitab provides various least-squares and logistic regression procedures. Least squares procedures are used when response variables is continuous and logistic regression when response variables is categorical (Meyer and Krueger, 1998). Due to the fact that all variables are continuous in data set, the least square procedure is applied while evaluating the data.

At this point it is better to emphasize that in order to validate the prediction performance of regression analysis, the total data of 16 projects were divided in to two groups, which are training set and the validation set. In the validation set, arbitrary chosen (by lottery) 2 project data is stored and these data were not used in the application of analysis. In other words, the training set analyzed by regression consists of remaining 14 projects data.

5.4.1.1 Correlation of Variables

The least square regression analysis is not applied if the total number of variables is greater than the number of observations, because residuals degree of freedom goes below zero. It can easily be seen that in data set of this study, the number of observations is equals to 16 and the number of the variables is 17. Moreover, when two observations are removed from the data set for the validation purpose, the gap is increased. That's why, instead of removing variables based on the experience of which may have no effect on the cost, as it is proposed in the literature, correlation of independent variables have been investigated to find the linear relationship between each other, if exists.

By using Minitab Pearson product moment, correlation coefficients between each pair of variables were calculated. Pearson product moment correlation coefficient measures the degree of linear relationship between two variables. The correlation coefficient assumes a value between -1 and +1. If one variable tends to increase as the other decreases, the correlation coefficient is negative. Conversely, if the two variables tend to increase together the correlation coefficient is positive. For two variables, the correlation coefficient r is calculated with below equation in Minitab.

For the two variables x and y,

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1)s_x s_y} \qquad (3)$$

where x bar and s_x are the sample mean and standard deviation for the first sample, and y bar and s_y are the sample mean and standard deviation for the second sample. When, the correlation coefficient is too close to either +1 or -1, it means that these two variables are highly correlated so that one of them should be removed from the analysis. The variable pairs with high correlation values, found in accordance with the Pearson correlation procedure, can be seen in Table 5.1. The full set of Pearson correlation matrix is given separately, in Appendices.

Table 5.1: List of variable pairs with high correlation values

Variable Pairs	**Pearson Correlation Value**
LTT - WR (x1 - x 8)	0.922
LTT - NTW (x1 - x 9)	0.915
WR - NTW (x8 - x9)	0.927
MS - MOS (x15 - x12)	0.902

As can be seen from the above table, several variables are highly correlated with each other, because of their high correlation values. That's why; the variables LTT (Total Length of Main Trackway), NTW (Number of Thermite Welding) and MS (Maximum Superelevation) are eliminated.

5.4.1.2 Best Subset Procedure

Best subsets regression generates regression models using the maximum R^2 criterion by first examining all one-predictor regression models and then selecting the two models giving the largest R^2. Minitab displays information on these models, examines all two-predictor models, selects the two models with the largest R^2, and displays information on these two models. This process continues until the model contains all predictors.

If m specifies the number of predictors, Minitab first selects the one-predictor regression model giving the largest R-squared. Minitab then prints information on this model and the next best one-predictor model. Next Minitab finds the two-predictor model with the largest R-squared, and prints information on it and the next best. The process continues until all m predictors are used.

"The best subsets regression procedure can be used to select a group of likely models for the analysis of variable selection. The general method is to the smallest subset that fulfills certain statistical criteria. The reason that one would use a subset of variables rather than a full set is because the subset model may actually estimate the regression coefficients and predict future responses with smaller variance than the full model using all predictors" (Gündüz, 2002).

In the data analysis of this study, the best subset regression is decided to be used instead of using the full set of data for regression analysis to reduce the steps of regression and eliminate more variables which do not contribute to the closeness of fitness (R^2) of the final model. That's why; best subset procedure is applied to data set in two parts.

In the first subset, which includes the first half of the variable set R^2 value of 79.0 % is achieved with 7 variables where 9 variables is giving R^2 value of 80.2 % Therefore, the remaining two variables, which are NST (Number of Simple Turnout) and SS (Sleeper Spacing), were considered as non significant.

In the second subset; which includes the second half of the variable set, set R^2 value of 79.9 % is achieved with 6 variables where 8 variables is giving the same R^2 value. Therefore, the remaining variable CS (Commercial Speed) was considered as non significant. The detailed best subset analysis results can be seen in appendices.

Consequently, by using correlation and best subset procedures 6 variables were eliminated (see table 5.2). Least square regression analysis will be conducted with the remaining 11 variables and 14 observations (see table 5.3).

Table 5.2: List of eliminated variables

1	2	3	4	5	6
Total Length of Main Trackway (Meters)	Number of Simple Turnout	Sleeper Spacing (cm)	Number of Thermite Welding	Commercial Speed (Km/hr)	Maximum Superelevation (Cm)
LTT	NST	SS	NTW	CS	MS
x1	x5	x6	x9	x13	x15

Table 5.3: List of remaining variables

Number	Variable Abbreviation	Number	Variable Abbreviation
1	LBT	7	HPC
2	LDF	8	MOS
3	NC	9	MSL
4	WC	10	MHC
5	WR	11	MVC
6	NS		

5.4.1.3 Regression Analysis Steps

As it is stated in the previous section the first regression analysis is performed with 11 variables. The evaluation of these variables is done by stepwise manner and unnecessary parameters, which do not fit the model well, have been dropped off the model by considering their p values. This procedure is called parsimonious modeling. Pankratz (1983) states that the principle of parsimony is important, because parsimonious models generally produce better forecasts in general.

"In parsimonious models, a backward elimination method is used for the initial RM. According to this technique, variables that were not contributing to the model is eliminated one by one at each step. The regression statistic, significance level (P value, which gives an indication of the significance of the variables included in the model) is used for determination of variables to be eliminated. In general, the variables corresponding to the coefficients with P values close to or less than 0.10 are considered to have significant contribution to the model" (Öntepeli, 2005). The same elimination procedure was followed in this study. The P value of the each eliminated variable and the coefficient of determination (R^2) of each model from R.1 to R.5 is given in the Table 5.4.

In the model R.1, the variable defines the radius of minimum allowed vertical curvature of the trackway (MVC) has the highest P value, which equals to 0.959. This value is very high when we consider the previously mentioned criteria of the P values. Therefore, MCV variable probably does not have a major contribution to the model and it is removed from the model. So, the regression model R.2 was performed by the remaining 10 parameters.

In model R.2, the coefficient for the variable the number of crossovers installed (NC) had the highest P value as 0.205 and was removed from the model. In the model R.3, R.4, R.5 and R.6, the same procedure is followed. As a result, the variables MOS (the maximum operational speed), MHC (radius of minimum allowed horizontal curvature), WC (workmanship cost per meter) and MSL (maximum allowed slope of line) are removed from the analysis. The detailed outputs of the Minitab program regarding these models can be seen in appendices.

The regression model R.7, was performed by using the remaining 5 variables. Because of the reason that P values of variables included in model R.7 are below or too close to 0.1, it is selected as final model which has a R^2 value of 0.963.

Table 5.4: List of P values of eliminated variables

Regression models	**Number of variables in the model**	**Eliminated variable**	**P value of eliminated variable**	**R^2 values of the models**
R.1	11	MVC	0.959	0.994
R.2	10	NC	0.205	0.994
R.3	9	MOS	0.208	0.988
R.4	8	MHC	0.294	0.982
R.5	7	WC	0.324	0.977
R.6	6	MSL	0.161	0.972

Minitab results of Final Regression Model (R.7) as follows:

The regression equation is

Cost = 529190 + 704 LBT + 707 LDF - 3860 WR + 293 NS + 325 HPC (4)

Predictor	Coef	StDev	T	P
Constant	529190	1796602	0.29	0.776
LBT	704.2	257.5	2.73	0.026
LDF	707.3	238.8	2.96	0.018
WR	-3860	2197	-1.76	0.117
NS	292.63	88.55	3.30	0.011
HPC	324.56	45.38	7.15	0.000

S = 3087067 R-Sq = 96.3% R-Sq (adj) = 93.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	1.95711E+15	3.91422E+14	41.07	0.000
Residual Error	8	7.62399E+13	9.52998E+12		
Total	13	2.03335E+15			

The prediction performance of final model (R.7) was tested by using the mean absolute percent error (MAPE). The MAPE formulation which is used for error measure as follows:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\mathrm{actual}_i - \mathrm{predicted}_i|}{|\mathrm{predicted}_i|}\,\mathrm{x}\,100 \qquad (5)$$

in “i” which is the project number; “actual” is the actual cost of the trackway and “predicted” is the predicted cost of the trackway by using the final regression model.

It is good to remember that these 2 project (observation 10 and observation 14) data were not used while generating the model. For testing the model, previously separated data of 2 projects were used. In accordance with the previously defined statistically significant variables and final regression equation (Eqn: 4), the predicted cost for these two projects were calculated (see Table 5.5) by entering the values of the variables present in the equation.

Table 5.5: Predicted values of testing projects

Project Number	**Fit (Predicted value - USD)**	**Standard Deviation Fit**	**95.0% Confidence Interval**	**95.0% PI**
10	27,799,471	2978780	(20,930,393 - 34,668,549)	(17,906,980 - 37,691,962)
14	21,379,299	1653483	(17,566,360 - 25,192,238)	(13,303,678 - 29,454,920)

Mean absolute percent error (MAPE) value of this 2 prediction was calculated according to above given formula. The results of the prediction performance of regression model R.7 can be seen in table 5.6 below.

Table 5.6: Prediction performance of final model

Project Number	**Predicted Values (USD)**	**Real Project Values (USD)**	**Percent Error**	**MAPE**
10	27,799,471	26,640,000	-4.17%	-2.32
14	21,379,299	21,280,000	-0.46%	

According to the prediction performance represented by the MAPE, regression model R.7, which has a R^2 value of 0.963, can be considered as acceptable. The R^2 value of 0.963 indicated that this model explained 96.5% of the variations in the dependent variable by the independent variables; whereas the MAPE value of -2.32 indicated that the model R.7 produced predictions within an average absolute error of 2.32%. Thus, regression analysis is finalized.

5.4.2 ARTIFICIAL NEURAL NETWORK APPLICATION

The application of ANN analysis was performed using Neural Power, which is a general, integrated, easiest-to-use and powerful Artificial Neural Network (ANN) program. It can be used in almost all study fields such as multi-nonlinear regression, forecasting, curve fit, pattern recognition, decision making and problem optimization, time series analysis and market predictions.

The parameters of ANN, which were defined in previous sections were reorganized and changed after each trial to find the best architecture thorough the Neural Power. The interfaces of the program for basic parameter input are expressed by following figures. The assigning of learning rate, momentum and stopping criteria (Root of Mean Square Error - RMSE) of the iteration is shown in the figure 5.2 below.

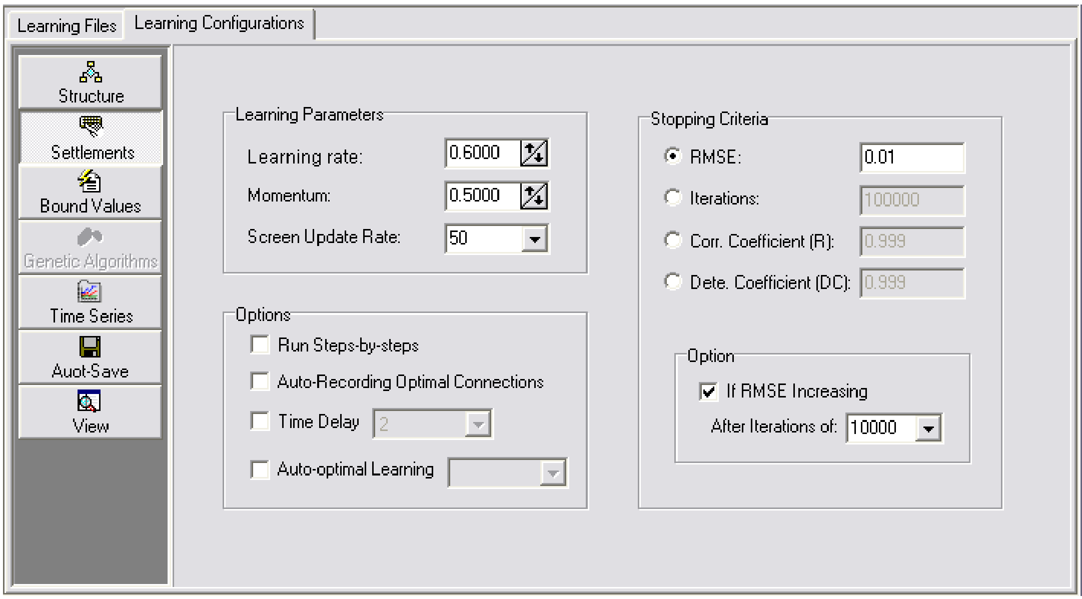

Figure 5.2: Selection of learning rate, momentum and RMSE

The RMSE is a quadratic scoring rule which measures the average magnitude of the error and shows the difference between forecast and corresponding observed values, each squared and then averaged over the sample. Then, the square root of the average is taken. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable. (Eumetcal, http://www.eumetcal.org.uk, last access 15 June 2008). In this study, RMSE value of 0.01 was used as stopping criteria of the iteration.

Other important ANN parameters that are number of hidden layers, number neurons in hidden layer and transfer function type were assigned as shown in figure 5.3.

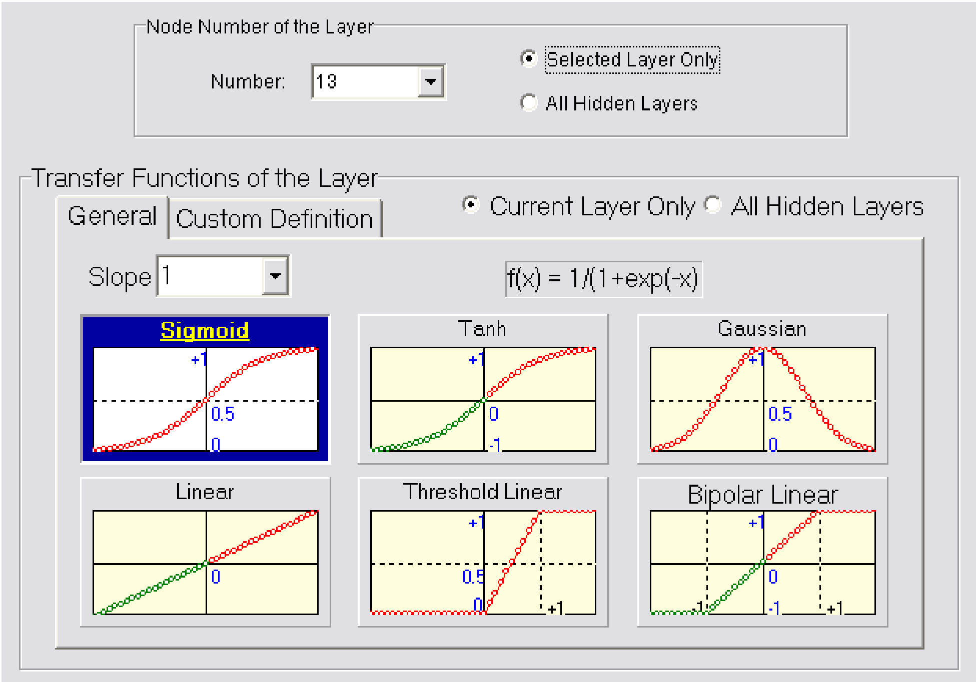

Figure 5.3: Selection of transfer function and number of neurons

Uğur (2007) states that normalization make possible to use of active region of function during the transfer of input data. Data normalization prevents the possible negative effects of high valued cumulative totals. Generally, it is recommended that data should be scaled between the intervals (0, 1) or (-1, +1).

Data are generally normalized for confidentiality and for effective training of the model being developed. The normalization of training data is recognized to improve the performance of trained networks (Siquera, 1999). That's why, in this study, the input and output values were normalized by using a scale (1/1000).

The normalized data of 14 projects (both input and cost values) were entered to the program. The data of 2 projects (project 10 and 14) are excluded from the analysis for testing the performance of the ANN configurations. The figure 5.4 shows the entered value of 14 projects.

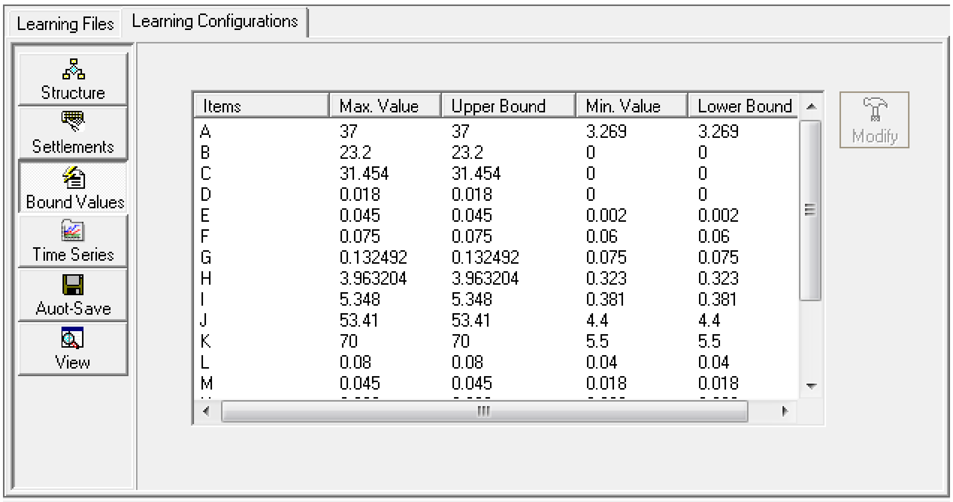

Figure 5.4: Entrance of data

After the normalized data is loaded to the program, calculations have been started. Figure 5.5 shows the approach to the desired error rate by increasing of iterations.

Figure 5.5: Graph of approach to the RMSE

5.4.2.1 Neural Network Analysis Steps

The number of hidden layer neurons has been decided according to the Hegazy's proposal mentioned in section 5.2.3 of this Chapter. In this study, three sets of ANN models (S.1, S.2, S.3) with one hidden layer were performed for the analysis and numbers of hidden layer neurons were decided in accordance with the 0,75m, m, or 2m + 1, coefficients (see Table 5.7).

Table 5.7: Number of hidden neurons in each set of model

	Number of Hidden Layer Neurons			
	S.1	**S.2**	**S.3**	
Number of Input Neurons(m)	**2m+1**	**m**	**0.7m**	**Number of Output Neurons**
17	33	17	13	1

In each set, the learning rate and the momentum parameters of ANN, were set between 0.5 and 0.9 (in steps of 0.1) to examine their effect and establish the best NN model as mentioned in section 5.2.3 of this Chapter.

In the first set (S.1), which has a configuration of 17-33-1, 25 ANN models have been developed by changing the learning rate and momentum parameters between 0.5 and 0.9 (in steps of 0.1). Among them, the best structure of an ANN (S1.A) was determined to be 17-33-1 (0.6-0.6), which means that there are 17, 33, and 1 neurons in the input, hidden, and output layers, respectively, and 0.6 and 0.6 are the learning rate and the momentum coefficient of the back-propagation algorithm, respectively (see table 5.8). The figure 5.6 below shows the ANN structure of S1.A.

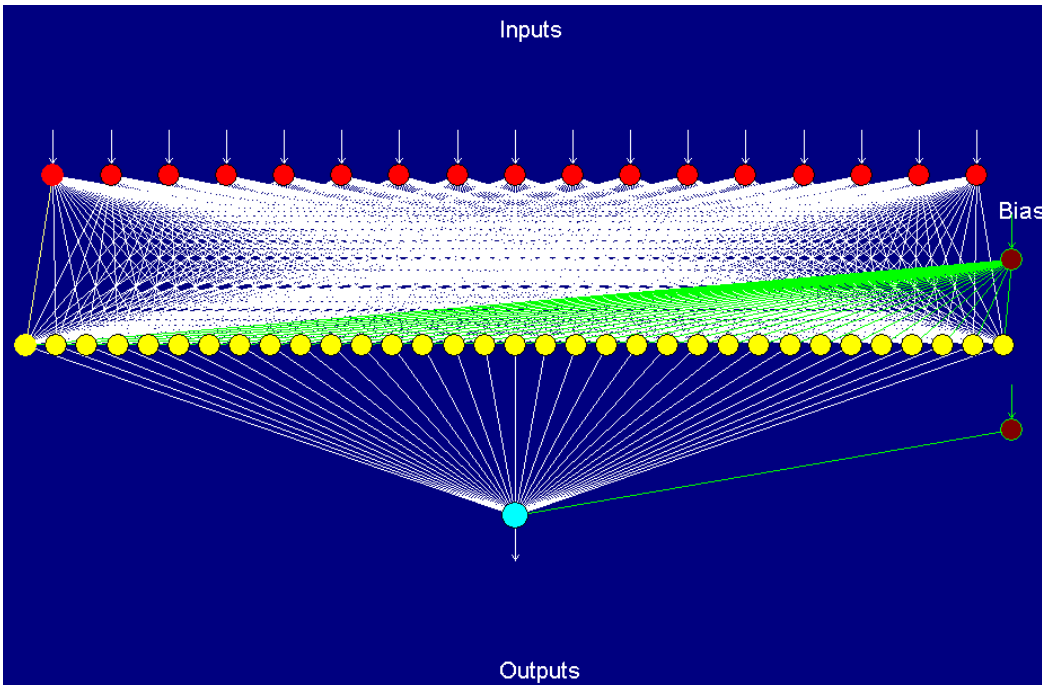

Figure 5.6: ANN architecture of S1.A

In the second set (S.2), which has a configuration of 17-17-1, 25 ANN models have been performed by changing the learning rate and momentum parameters between 0.5 and 0.9 (in steps of 0.1). Among them, the best structure of an ANN (S2.B) was determined to be 17-17-1 (0.5-0.7), which means that there are 17, 17, and 1 neurons in the input, hidden, and output layers, respectively, and 0.5 and 0.7 are the learning rate and the momentum coefficient of the back-propagation algorithm, respectively (see table 5.8). The figure 5.7 below shows the ANN structure of S2.B.

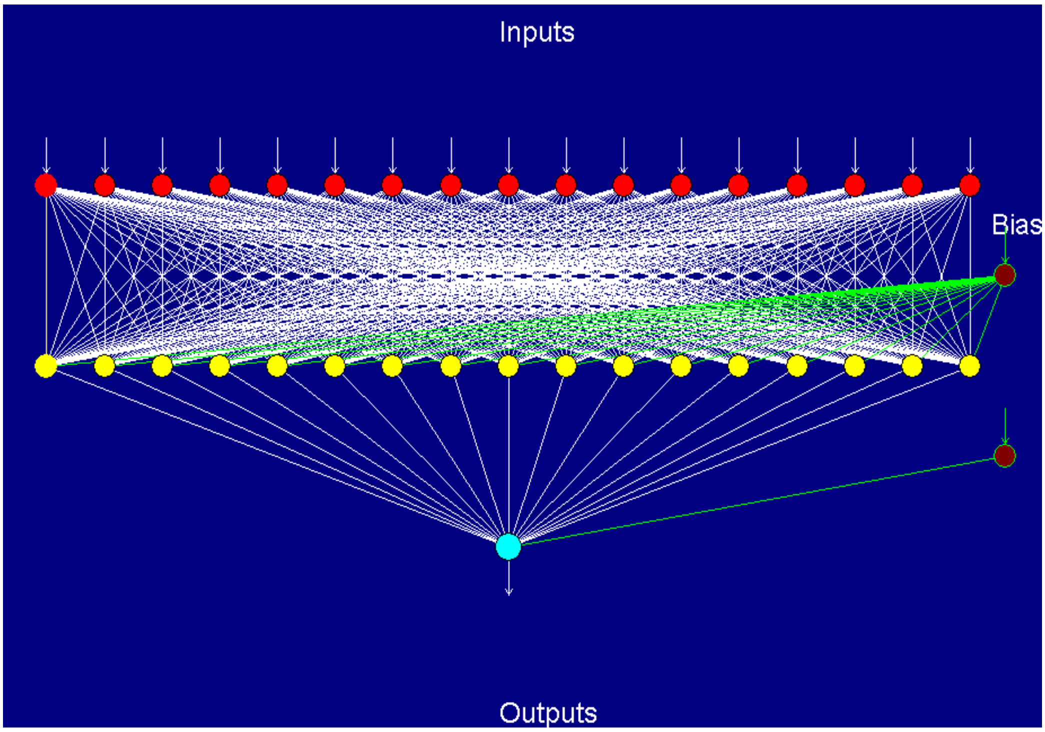

Figure 5.7: ANN architecture of S2.B

In the third set (S.3), which has a configuration of 17-13-1, another 25 ANN models have been implemented by changing the learning rate and momentum parameters between 0.5 and 0.9 (in steps of 0.1). Among them, the best structure of an ANN (S3.C) was determined to be 17-13-1 (0.5-0.5), which means that there are 17, 13, and 1 neurons in the input, hidden, and output layers, respectively, and 0.5 and 0.5 are the learning rate and the momentum coefficient of the back-propagation algorithm, respectively (see table 5.8). The figure 5.8 below shows the ANN structure of S3.C.

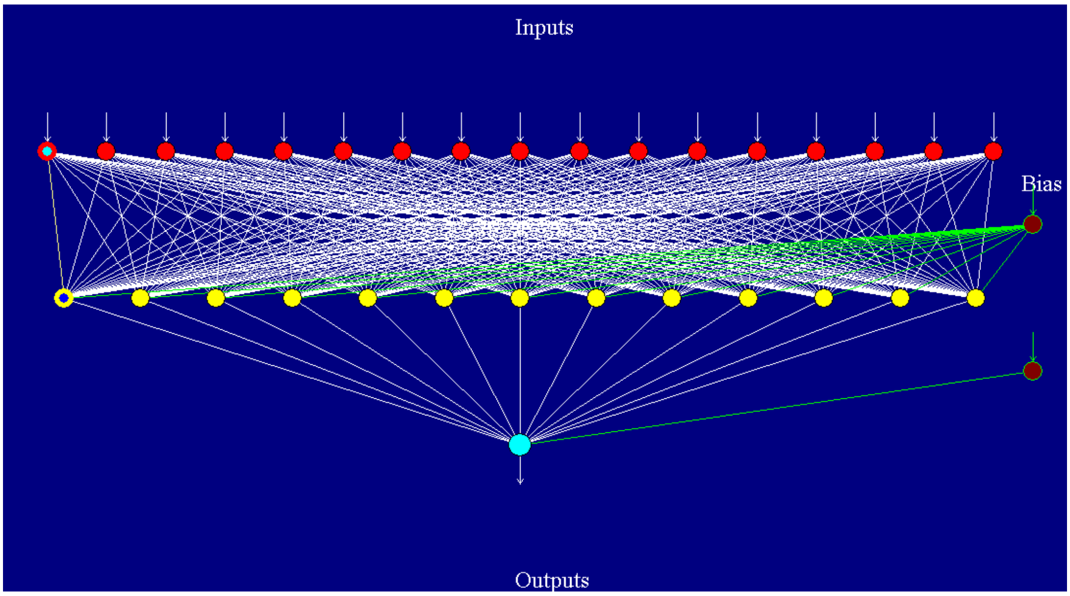

Figure 5.8: ANN architecture of S3.C

Table 5.8: Network architecture of best ANN of each group

	Network Architecture		
Network Characteristics	**S1.A**	**S2.B**	**S3.C**
Network architechture	17-33-1	17-17-1	17-13-1
Learning algorithm	BP	BP	BP
Learning rate	0.6	0.5	0.5
Momentum rate	0.6	0.7	0.5
Stopping criteria	0.01	0.01	0.01
Number of iteration	2517	3245	2983

The best architecture of each ANN group was selected by examining prediction performance of them. The predicted cost values of the best of each group for the projects number 10 and number 14 (testing projects) can be seen in table 5.9. It should be remembered that the prediction results are scaled with (1/1000). In addition, the prediction error and MAPE of S1.A S2.B and S3.C are presented in table 5.10.

Table 5.9: Predicted cost values

Project No	Actual Cost (USD*1000)	Predicted Cost (USD*1000)		
		S1.A	S2.B	S3.C
10	26,640	29,318	29,225	28,237
14	21,280	24,242	24,224	20,101

Table 5.10: Prediction performance of each model

Project No	Prediction Error		
	S1.A	S2.B	S3.C
10	9.136%	8.845%	5.656%
14	12.220%	12.152%	5.867%
MAPE	10.678	10.498	5.761

According to the prediction performance represented by the MAPE, the model S3.C was produced reasonable predictions within an average absolute error of 5.761%. Thus, the model S3.C was selected as best architecture for the data set of this study and analysis was finalized.

5.5 RESULTS OF ANALYSIS

In this Chapter, conceptual cost estimation models of trackworks, which have been generated by using regression and artificial neural networks, were presented. The most important parameters revealed thorough the regression analysis were LBT (length of ballasted trackway), LDF (length of direct fixation trackway), WR (total weight of the rail), NS (number of sleepers) and HPC (hourly passenger capacity per direction). The relationships between cost and these parameters were explained by using the final neural network model S3.C. The graphics below, which are related with these parameters are represented in figures 5.9 - 13 where, A denotes the final cost of the project and B, C, H, J, K are denotes LBT (length of ballasted trackway), LDF (length of direct fixation trackway), WR (total weight of the rail), NS (number of sleepers) and HPC (hourly passenger capacity per direction), respectively. It should be noted that these graphs show only the tendencies of cost changes with above mentioned variables.

In figures 5.9, 5.10, 5.12 and 5.13 the cost increasing with a decreasing slope as the length of ballasted trackway, length of direct fixation trackway, of sleepers and hourly passenger capacity per direction increases. However, as can be shown in figure 5.11, the cost has tendency to decrease as the total weight of the rail used (WR) or (H) decreases. This seems to be an unexpected result, which may actually presents the rate of increase of weight of rail is lower than the rate of increase of the cost. As a result, it may be concluded that as the project becomes bigger the cost effect of the rail used decreases.

In addition to these results, valuable information is provided by ANN analysis is that the most cost effective parameter is revealed as hourly passenger capacity per direction (HPC) or (K), which can be seen in figure 5.14.

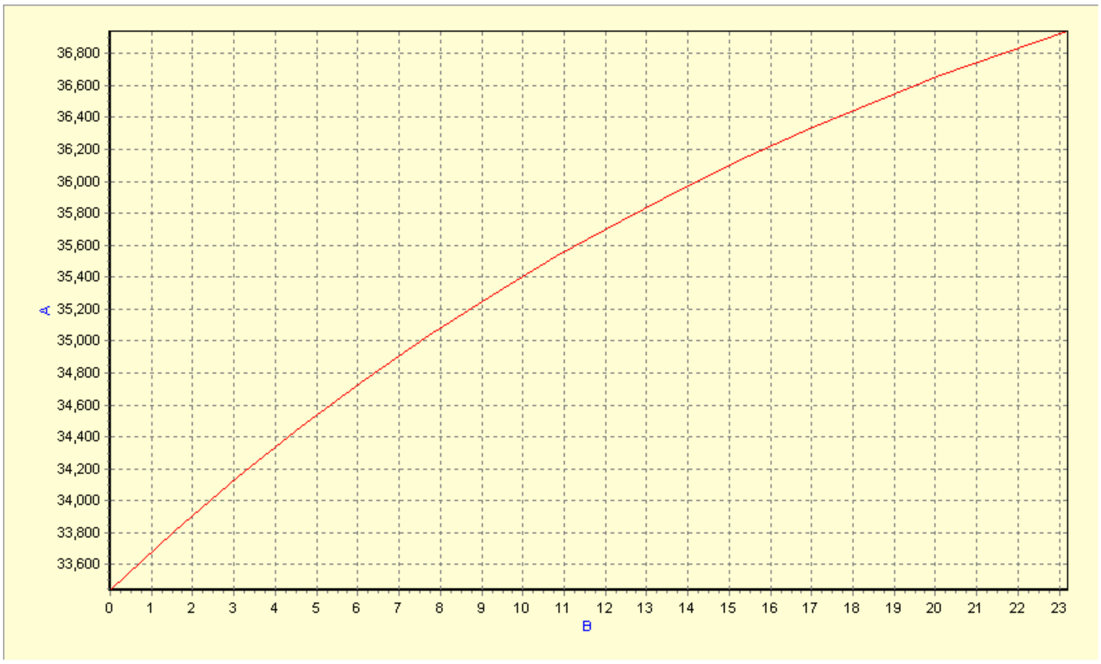

Figure 5.9: Cost (A) vs. LBT (B)

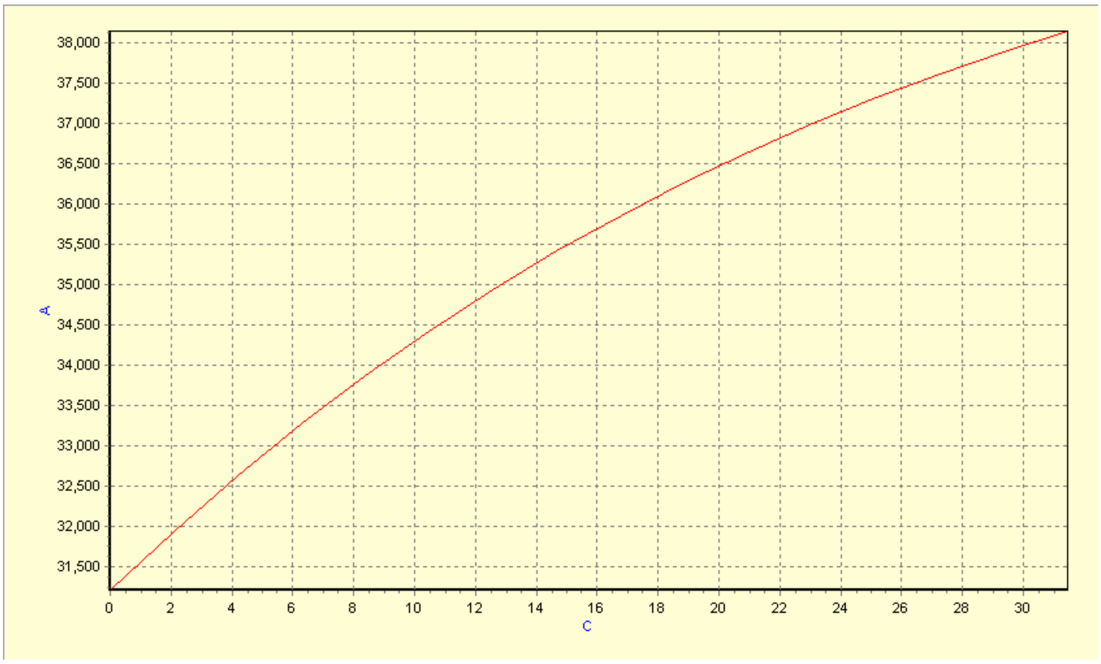

Figure 5.10: Cost (A) vs. LDF (C)

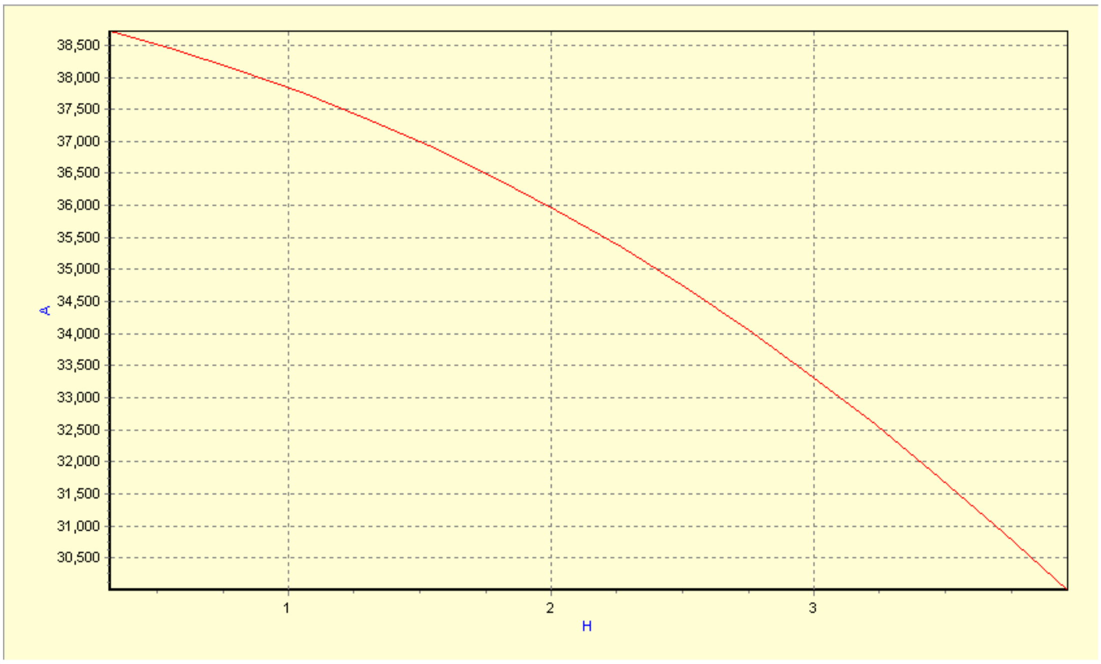

Figure 5.11: Cost (A) vs. WR (H)

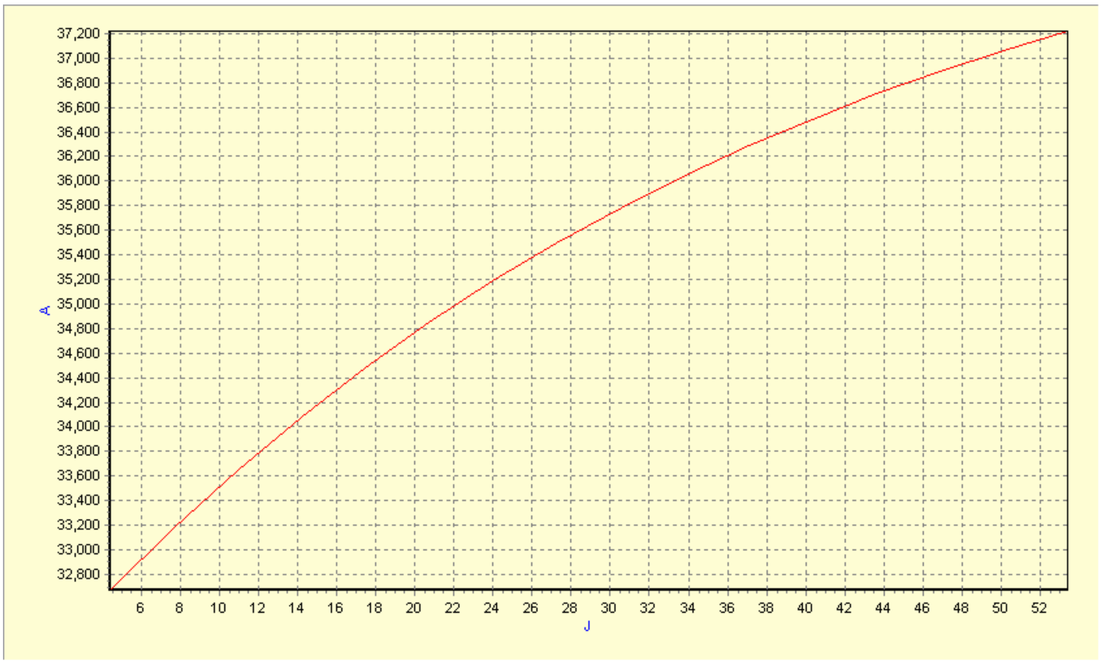

Figure 5.12: Cost (A) vs. NS (J)

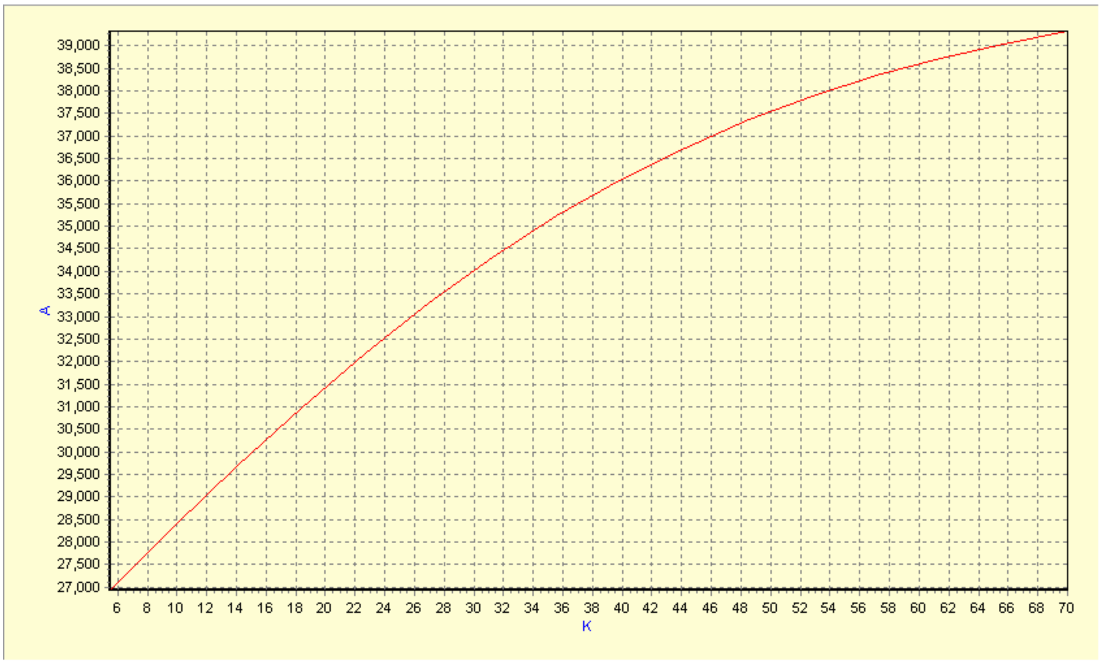

Figure 5.13: Cost (A) vs. HPC (K)

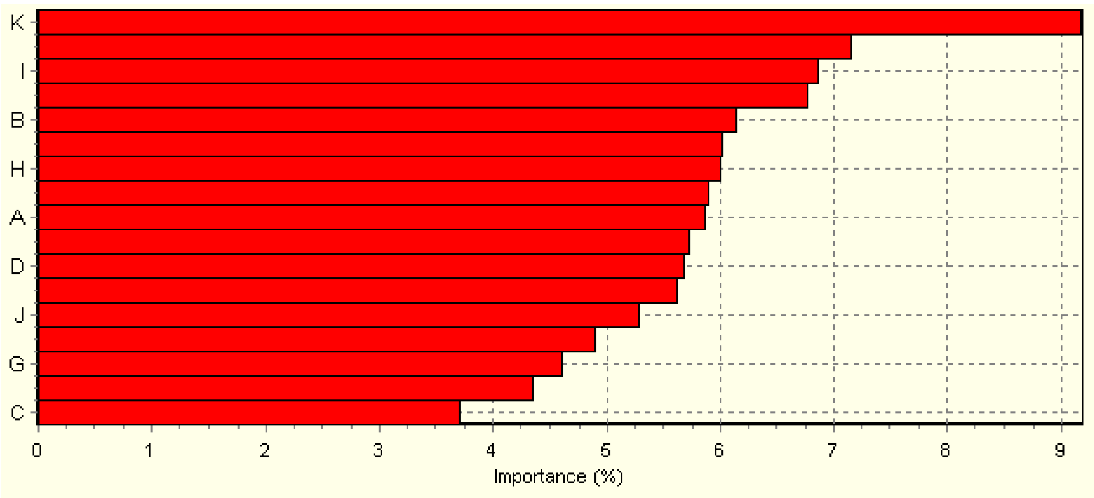

Figure 5.14: Importance of variables

CHAPTER 6

CONCLUSIONS AND SUGGESTIONS

The main objective of this work was to develop models using multivariable regression and artificial neural network approaches for cost estimation of the construction costs of trackworks of Turkish light rail transit and metro projects at the early stages of the construction process. These two approaches used a data set from 16 projects. The approach was shown to be capable of providing accurate estimates of trackworks cost by using seventeen parameters available at the early design phase.

In content of this study; Chapter 1 gives the scope and purpose of the study and the importance of the LRT and metro systems in developing countries. Chapter 2 presents the available literature, which covers the early cost estimation studies in different fields and various techniques to handle the estimation process. Chapter 3 introduces the LRT and metro system, gives definitions and basic information about them. Chapter 4 presents the main components of the trackworks construction and data collection procedure.

Finally, Chapter 5 gives background information about the techniques that was used in this study and analysis steps including results of each of them. According to the results of each method, regression analysis was estimated the cost of testing samples with an error of MAPE of 2.32%. On the other hand, artificial neural network was estimated the cost with 5.761% error, which is slightly higher than the regression error. As a result, two successful models have been developed within the scope of this study. These models can be beneficial while taking the decision in the tender phase of projects that includes trackworks.

The MAPE results have showed us, the regression has fit to the data set well. In addition to this, the prediction performance of the ANN is highly satisfactory also. According to many studies present in literature, the estimation performances of ANNs are usually presented as superior to regression analysis. That's why; the results of analysis of this research may be case specific due to the number of LRT and metro projects available in Turkey. In order to ensure the performances of these to model in this specific subject of area, further studies should be done with an expanded data set in future. Because of the reason that neural networks train themselves by using observations and the performance of a neural network model of cost estimation inevitably depends on the quality and the quantity of data. As the number of observations increases the estimation error of ANNs decreases. Therefore, it is possible to develop a solid estimation model with ANN for trackway projects in Turkey with trustworthy, high quality, full-scale cost data of various projects. However; establishing the best ANN model needed considerable amount of time, because of the the trial and error procedure, while defining the ANN parameters to find architecture for best estimation. Therefore; to be more effective in finding the parameters of the ANN than the trial and error method, other applications such as back propagation network model incorporating a genetic algorithm (GA) may be used in future to estimate the trackway cost in the early project stage.

REFERENCES

AAR. (AustralAsia Railway), (2003). Contunious rail welding.

Adeli H. and Wu M., (1998). Regularization neural network for construction cost estimation. Journal of Construction Engineering Management; 124(1):18–24.

Alku A., (2002). Modern Light Rail Basics. Helsinkin city planning committee.

Allen C. Frank, (1920). Railway Curves and Earthwork, Sixth edition, McGraw – Hill Book Co.

APTA., http://www.apta.com/research/stats, last access September 11, 2008. American Public Transportation Association.

Arema. (The American Railway Engineering and Maintenance-of-Way Association), (2003). Practical Guide to Railway Engineering.

Armstrong J., (1998). The Railroad: What It Is, What It Does. Omaha, NE: Simmons-Boardman, 4th ed.

Boorse Jack W., (December, 2000). Transportation Research Board, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Bottoms G.D., (2004). Continuing Developments in Light Rail Transit in Western Europe., Federal Transit Administration.

Boussabaine A.H., (1996). The use of artificial neural networks in construction management: a review. Construction Management and Economics; 14(5):427–36.

Caputo A.C., Pelagagge, P.M., (2007). Parametric and neural methods for cost estimation of process vessels. International Journal of Production Economics doi:10.1016/j.ijpe.2007.08.002.

Chester G. W., M.ASCE and Bing M., (2005). Neural Network Modeling of Highway Construction Costs. ASCE, Journal of Construction Engineering and Management. Vol. 131, No. 7.

Daniels L.E. and Tuten J.M., (2001). APTA Rail Transit Conference, Boston.

David E., (2004). A Metallurgical History of Railmaking Slee. Australian Railway History, pp43-56.

ERRAC., The European Rail Research Advisory Council, (April, 2004). Light Rail and Metro Systems in Europe – Current market, perspectives and research implication, Brussels.

Eumetcal., http://www.eumetcal.org.uk, last access 15 June 2008.

Garza J. and Rouhana K., (1995). Neural networkversus parameter-based application. Cost Engineering;37(2):14–8.

Günaydın H.M., Doğan Z.S, (2004). A neural network approach for early cost estimation of structural system of building. International Journal Project Management, 22: 595 – 602.

Gündüz M., (2002). Change order impact assessment for labor intensive construction. Phd Thesis, University of Wisconsin Madison.

Hauser, D., (1978). Welding of Railroad Rails-A Literature and Industry Survey, ASTM Special Technical Publication 644, edited by D. H. Stone and G. G. Knupp, pp. 118-144, American Society for Testing and Materials, Philadelphia, PA.

Hegazy T., Ayed A., (1998). Neural network model for parametric cost estimation of highway projects. Journal of Construction Engineering and Management; 124(3):210–8.

Hegazy T., (2002). Computer-based Construction Project Management. Upper Saddle River, NJ: Prentice-Hall Inc.

Hegazy T., Fazio P, Moselhi O. (1994). Developing practical neural network applications using back-propagation. Microcomputer in Civil Engineering; 9:145–59.

Hill T. & Lewicki P. (2007). Statistics Methods and Applications. StatSoft, Tulsa, OK.

Kim G.H, Sung-Hoon A., Kyung-In K., (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. Building and Environment 39, pp 1235 – 1242.

Litman T., (2008). Evaluating Rail Transit Criticism. Victoria Transport Policy Institute.

Mason A.K and Smith, A.E (1997) Cost estimation predictive modeling. Regression versus neural networks. Engineering Economist, 42 (2), pp.137-161.

McKim R. (1993). Neural network application to cost engineering. Cost Engineering;35(7):31–5.

Metrobits., http://mic-ro.com/metro/metrolist.html, last access June 12, 2008.

Meyer R., Krueger D., (1998). A Minitab Guide To Statistics, Prentice Hall Inc.

Mida H., F., Martin, P., Vernadat, F., (2006). Cost estimation in mechanical production: The Cost Entity approach applied to integrated product engineering. International journal of Production Economics 103, 17–35.

Öntepeli M.B., (2005). Conceptual cost estimating of urban railway system projects. Ms. Thesis presented to Middle East Technical University (METU), Ankara, Turkey.

Pankratz A., (1983). Forecasting with univariate Box-Jenkins models, John Wiley and Sons, New York.

Pearson K. (1908). On the generalized probable error in multiple normal correlation. Biometrika, 6, 59-68.

Polat, A. P., Çıracı, M., (2005). Türkiye'de tasarım öncesinde maliyet tahmini için veri tabanı modeli, İTÜ Dergisi (Magazine) /a, 4 (2): 59-69.

Railway Technical., http://www.railway-technical.com/track, last updated 24th June 2008.

Rumelhart D., Hinton G., Williams R. (1986). Learning internal representations by error propagation. Parallel distributed processing. Cambridge, MA: MIT Press.

Smith A.E., Mason A.K., (1999). Cost estimation predictive modeling: Regression versus neural networks. Engineering economist 42 (2), 137–161.

Sönmez R., (2004). Conceptual cost estimation of building projects with regression analysis and neural networks. Canadian Journal of Civil Engineering,31: 677-683.

Squeira I. (1999). Neural Network-Based Cost Estimating. Master's Thesis. Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec, Canada.

Tam C.M, Fang C.F., (1999). Comparative cost analysis of using high performance concrete in tall building construction by artificial neural networks. ACI Structural Journal; 96(6):927–36.

Taplin M.R., (1998). The History of Tramways and Evolution of Light Rail, LRTA (Light Rail Transit Association).

Tennyson E.L., (2000). TTMA'S Position On Proposed Rapid Rail System For Trinidad and Tobago. Trinidad and Tobago Manufacturers' Association.

Uğur L.O., (2007). Analysis of Construction of Costs With Artificial neural networks. Thesis presented to Gazi University, Ankara, Turkey, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Urbanrail., http://www.urbanrail.net/eu/lon/london.htm,last access September 15, 2008.

Verlinden B., J.R. Duflou, P. Collin, D. Cattrysse. (2008). Cost estimation for Sheet Metal Parts Using Multiple Regression and Artificial Neural Networks: A Case Study. International Journal of Production Economics 111. 484 - 492.

Webster's Dictionary., http://www.websters-online-dictionary.org, last access 12 June 2008.

Wierda, L., (1990). Cost Information Tools for Designers. Delft University Press, Delft, The Netherlands.

Yesilada E., and Nielsen, Y., (1996). Railway and Metro Tunnels - CE 439 Lecture Notes, METU, Ankara.

Zhang Y.F., Fuh JYH. (1998). A neural network approach for early cost estimation of packaging products. Comput Ind Eng;34(2):433.

APPENDIX A

REGRESSION OUTPUTS

Table A.1 – Pearson Correlation Matrix

	x1	x2	x3	x4	x5	x6	x7	x8
x2	0.392							
	0.165							
x3	0.714	-0.365						
	0.004	0.200						
x4	0.501	-0.004	0.511					
	0.068	0.988	0.062					
x5	0.470	0.186	0.334	-0.108				
	0.090	0.525	0.243	0.714				
x6	0.482	-0.273	0.696	0.379	-0.015			
	0.081	0.346	0.006	0.182	0.958			
x7	-0.361	0.013	-0.375	0.207	-0.317	-0.354		
	0.205	0.966	0.186	0.478	0.270	0.214		
x8	0.922	0.383	0.642	0.399	0.463	0.401	-0.537	
	0.000	0.177	0.013	0.158	0.095	0.155	0.048	
x9	0.915	0.536	0.386	0.334	0.432	0.189	-0.448	0.927
	0.001	0.048	0.172	0.244	0.123	0.517	0.108	0.000
x10	0.589	0.518	0.202	-0.125	0.657	0.090	-0.470	0.717
	0.027	0.058	0.489	0.670	0.011	0.760	0.090	0.004
x11	0.484	0.254	0.296	0.641	-0.042	0.365	0.360	0.311
	0.080	0.381	0.304	0.013	0.886	0.199	0.206	0.280
x12	0.330	0.508	-0.053	0.305	0.044	-0.229	0.534	0.208
	0.249	0.064	0.858	0.289	0.882	0.430	0.049	0.475
x13	0.071	0.448	-0.269	0.298	-0.136	-0.419	0.735	-0.085
	0.809	0.108	0.352	0.300	0.644	0.136	0.003	0.771
x14	-0.310	-0.118	-0.224	-0.432	0.183	-0.278	-0.211	-0.190
	0.280	0.688	0.441	0.123	0.532	0.336	0.468	0.516
x15	0.119	0.412	-0.193	0.323	-0.104	-0.097	0.721	0.028
	0.685	0.144	0.509	0.260	0.725	0.742	0.004	0.924
x16	0.156	0.352	-0.110	0.400	-0.098	-0.110	0.699	0.001
	0.593	0.217	0.709	0.156	0.740	0.709	0.005	0.998
x17	-0.022	0.491	-0.396	0.217	-0.023	-0.568	0.582	-0.112
	0.941	0.075	0.161	0.457	0.939	0.034	0.029	0.704

Table A.1 (cont.) – Pearson Correlation Matrix

	x9	x10	x11	x12	x13	x14	x15	x16
x10	0.796							
	0.001							
x11	0.273	0.033						
	0.345	0.912						
x12	0.280	0.112	0.557					
	0.333	0.704	0.039					
x13	-0.018	-0.156	0.551	0.620				
	0.950	0.595	0.041	0.004				
x14	-0.150	0.094	-0.692	-0.514	-0.525			
	0.610	0.748	0.006	0.060	0.054			
x15	0.117	0.104	0.509	0.902	0.792	-0.338		
	0.691	0.724	0.063	0.002	0.006	0.237		
x16	0.083	-0.084	0.806	0.718	0.875	-0.700	0.698	
	0.779	0.776	0.000	0.004	0.000	0.005	0.005	
x17	0.050	-0.101	0.408	0.585	0.669	-0.450	0.444	0.786
	0.864	0.730	0.147	0.028	0.000	0.107	0.111	0.001

Best Subsets Regression - 1

```
                                                      x
              Adj.                    x x x x x x x x 1
Vars   R-Sq   R-Sq    C-p         s   2 3 4 5 6 7 8 9 0

   1   43.8   39.1    1.3   9757833               X
   1   37.9   32.7    2.5  10259092                 X
   2   66.7   60.7   -1.3   7844192   X X
   2   61.1   54.1   -0.2   8476011             X X
   3   74.8   67.2   -0.9   7165221   X X       X
   3   71.1   62.5   -0.2   7662511   X X         X
   4   75.9   65.2    0.9   7377255   X X         X X
   4   75.9   65.1    0.9   7383729   X X       X     X
   5   78.0   64.2    2.4   7484938   X X X       X   X
   5   77.7   63.7    2.5   7532840   X X   X X X
   6   79.0   61.0    4.2   7810853   X X X     X X   X
   6   78.5   60.0    4.3   7906631   X X       X X X X
   7   79.6   55.9    6.1   8309400   X X   X X X X X
   7   79.5   55.5    6.1   8341917   X X     X X X X X
   8   79.9   47.8    8.1   9038454   X X X   X X X X X
   8   79.8   47.4    8.1   9070243   X X   X X X X X X
   9   80.2   35.6   10.0  10035871   X X X X X X X X X
```

Remove variables x5 ve x6

Best Subsets Regression - 2

Vars	R-Sq	Adj. R-Sq	C-p	s	x11	x12	x13	x14	x15	x16	x17
1	68.9	66.3	-0.7	7261994	X						
1	30.1	24.3	10.9	10883409						X	
2	73.0	68.1	0.1	7063793	X					X	
2	71.6	66.5	0.5	7239405	X			X			
3	76.4	69.3	1.1	6929559	X	X				X	
3	75.3	67.9	1.4	7089632	X					X	X
4	78.6	69.0	2.4	6958609	X	X				X	X
4	77.6	67.7	2.7	7108035	X	X		X		X	
5	79.6	66.9	4.1	7193263	X	X		X		X	X
5	79.0	65.9	4.3	7299436	X	X		X	X	X	
6	79.9	62.7	6.0	7637247	X	X		X	X	X	X
6	79.7	62.4	6.1	7670109	X	X	X	X		X	X
7	79.9	56.6	8.0	8243542	X	X	X	X	X	X	X

Remove variable x13

Regression Analysis (R.1)

The regression equation is
y = 6259890 + 640 x2 + 857 x3 + 446683 x4 - 455117 x7 - 8586 x8 + 292 x10
+ 302 x11 + 236018 x12 + 5308596 x14 + 55682 x16 + 179 x17

Predictor	Coef	StDev	T	P
Constant	6259890	15706394	0.40	0.729
x2	640.0	257.3	2.49	0.131
x3	857.1	237.1	3.61	0.069
x4	446683	585646	0.76	0.525
x7	-455117	301128	-1.51	0.270
x8	-8586	4061	-2.11	0.169
x10	291.81	92.47	3.16	0.087
x11	301.9	180.7	1.67	0.237
x12	236018	149831	1.58	0.256
x14	5308596	2631170	2.02	0.181
x16	55682	68824	0.81	0.503
x17	179	3092	0.06	0.959

S = 2519597   R-Sq = 99.4%   R-Sq(adj) = 95.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	11	2.02065E+15	1.83696E+14	28.94	0.034
Residual Error	2	1.26967E+13	6.34837E+12		
Total	13	2.03335E+15			

Source	DF	Seq SS
x2	1	4.22382E+14
x3	1	9.34122E+14

x4	1	5.81382E+13
x7	1	1.05756E+14
x8	1	7.08282E+12
x10	1	7.88017E+13
x11	1	3.57097E+14
x12	1	73180882738
x14	1	3.01910E+13
x16	1	2.69863E+13
x17	1	21293342514

Regression Analysis (R.2)

The regression equation is
y = 6976400 + 647 x2 + 863 x3 + 473494 x4 - 468926 x7 - 8748 x8 + 292 x10
+ 293 x11 + 240755 x12 + 5417630 x14 + 59303 x16

Predictor	Coef	StDev	T	P
Constant	6976400	7907274	0.88	0.443
x2	647.2	184.0	3.52	0.039
x3	863.0	174.5	4.95	0.016
x4	473494	293127	1.62	0.205
x7	-468926	150295	-3.12	0.052
x8	-8748	2411	-3.63	0.036
x10	291.81	75.57	3.86	0.031
x11	292.60	68.51	4.27	0.024
x12	240755	102587	2.35	0.101
x14	5417630	1502062	3.61	0.037
x16	59303	23505	2.52	0.086

S = 2058966 R-Sq = 99.4% R-Sq(adj) = 97.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	2.02063E+15	2.02063E+14	47.66	0.004
Residual Error	3	1.27180E+13	4.23934E+12		
Total	13	2.03335E+15			

Source	DF	Seq SS
x2	1	4.22382E+14
x3	1	9.34122E+14
x4	1	5.81382E+13
x7	1	1.05756E+14
x8	1	7.08282E+12
x10	1	7.88017E+13
x11	1	3.57097E+14
x12	1	73180882738
x14	1	3.01910E+13
x16	1	2.69863E+13

Regression Analysis (R.3)

The regression equation is
y = 357534 + 648 x2 + 804 x3 - 305243 x7 - 6046 x8 + 225 x10 + 350 x11 + 158011 x12 + 4324606 x14 + 39121 x16

Predictor	Coef	StDev	T	P
Constant	357534	8008424	0.04	0.967
x2	648.2	217.9	2.97	0.041
x3	804.1	202.1	3.98	0.016
x7	-305243	131443	-2.32	0.081
x8	-6046	2057	-2.94	0.042
x10	224.82	74.81	3.01	0.040
x11	350.33	69.22	5.06	0.007
x12	158011	105255	1.50	0.208
x14	4324606	1588020	2.72	0.053
x16	39121	23576	1.66	0.172

S = 2438215 R-Sq = 98.8% R-Sq(adj) = 96.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	2.00957E+15	2.23285E+14	37.56	0.002
Residual Error	4	2.37796E+13	5.94489E+12		
Total	13	2.03335E+15			

Source	DF	Seq SS
x2	1	4.22382E+14
x3	1	9.34122E+14
x7	1	1.63441E+14
x8	1	6.04280E+12
x10	1	4.20701E+13
x11	1	3.95176E+14
x12	1	12903200494
x14	1	2.99538E+13
x16	1	1.63689E+13

Regression Analysis (R.4)

The regression equation is
y = 2317264 + 750 x2 + 801 x3 - 179970 x7 - 5104 x8 + 234 x10 + 345 x11
+ 3137641 x14 + 29781 x16

Predictor	Coef	StDev	T	P
Constant	2317264	8836528	0.26	0.804
x2	750.2	231.5	3.24	0.023
x3	800.5	226.0	3.54	0.017
x7	-179970	113578	-1.58	0.174
x8	-5104	2191	-2.33	0.067
x10	233.92	83.39	2.81	0.038
x11	345.09	77.31	4.46	0.007
x14	3137641	1540199	2.04	0.097
x16	29781	25432	1.17	0.294

S = 2726807 R-Sq = 98.2% R-Sq(adj) = 95.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	1.99617E+15	2.49521E+14	33.56	0.001
Residual Error	5	3.71774E+13	7.43548E+12		
Total	13	2.03335E+15			

Source	DF	Seq SS
x2	1	4.22382E+14
x3	1	9.34122E+14
x7	1	1.63441E+14
x8	1	6.04280E+12
x10	1	4.20701E+13
x11	1	3.95176E+14
x14	1	2.27418E+13
x16	1	1.01960E+13

Regression Analysis (R.5)

The regression equation is
y = - 505885 + 738 x2 + 708 x3 - 68413 x7 - 4444 x8 + 262 x10 + 405 x11
+ 1789751 x14

Predictor	Coef	StDev	T	P
Constant	-505885	8760359	-0.06	0.956
x2	738.3	238.4	3.10	0.021
x3	708.0	218.1	3.25	0.018
x7	-68413	63726	-1.07	0.324
x8	-4444	2182	-2.04	0.088
x10	261.91	82.32	3.18	0.019
x11	405.17	59.60	6.80	0.000
x14	1789751	1054561	1.70	0.141

S = 2809904 R-Sq = 97.7% R-Sq(adj) = 95.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	1.98598E+15	2.83711E+14	35.93	0.000
Residual Error	6	4.73734E+13	7.89556E+12		
Total	13	2.03335E+15			

Source	DF	Seq SS
x2	1	4.22382E+14
x3	1	9.34122E+14
x7	1	1.63441E+14
x8	1	6.04280E+12
x10	1	4.20701E+13
x11	1	3.95176E+14
x14	1	2.27418E+13

Regression Analysis (R.6)

The regression equation is
y = - 7810092 + 692 x2 + 692 x3 - 3535 x8 + 267 x10 + 372 x11 + 1657088 x14

Predictor	Coef	StDev	T	P
Constant	-7810092	5578163	-1.40	0.204
x2	692.5	237.1	2.92	0.022
x3	691.7	220.0	3.14	0.016
x8	-3535	2032	-1.74	0.126
x10	267.47	83.05	3.22	0.015
x11	372.44	51.76	7.20	0.000
x14	1657088	1058643	1.57	0.161

S = 2840349   R-Sq = 97.2%   R-Sq(adj) = 94.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	1.97688E+15	3.29479E+14	40.84	0.000
Residual Error	7	5.64731E+13	8.06758E+12		
Total	13	2.03335E+15			

Source	DF	Seq SS
x2	1	4.22382E+14
x3	1	9.34122E+14
x8	1	8.97041E+13
x10	1	2.34906E+13
x11	1	4.87410E+14
x14	1	1.97668E+13

Regression Analysis (R.7)

The regression equation is
y = 529190 + 704 x2 + 707 x3 - 3860 x8 + 293 x10 + 325 x11

Predictor	Coef	StDev	T	P
Constant	529190	1796602	0.29	0.776
x2	704.2	257.5	2.73	0.026
x3	707.3	238.8	2.96	0.018
x8	-3860	2197	-1.76	0.117
x10	292.63	88.55	3.30	0.011
x11	324.56	45.38	7.15	0.000

S = 3087067   R-Sq = 96.3%   R-Sq(adj) = 93.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	1.95711E+15	3.91422E+14	41.07	0.000
Residual Error	8	7.62399E+13	9.52998E+12		
Total	13	2.03335E+15			

Source	DF	Seq SS
x2	1	4.22382E+14
x3	1	9.34122E+14
x8	1	8.97041E+13
x10	1	2.34906E+13
x11	1	4.87410E+14