ESTIMATING SWELLING CHARACTERISTICS OF CLAYS USING METHYLENE BLUE TEST - A MACHINE LEARNING APPROACH

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

BY
GAMZE DİDEM ÖGET

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

SEPTEMBER 2014

Approval of the thesis:

ESTIMATING SWELLING CHARACTERISTICS OF CLAYS USING METHYLENE BLUE TEST -A MACHINE LEARNING APPROACH

Submitted by **GAMZE DİDEM ÖGET** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering Department, Middle East Technical University** by**,**

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

Prof. Dr. Ahmet Cevdet Yalçıner
Head of Department, **Civil Engineering** _______________

Asst. Prof. Dr. Onur Pekcan
Supervisor, **Civil Engineering Dept., METU** _______________

Prof. Dr. Erdal Çokça
Co-Supervisor, **Civil Engineering Dept., METU** _______________

Examining Committee Members

Asst. Prof. Dr. Nejan Huvaj Sarıhan
Civil Engineering Dept., METU _______________

Asst. Prof. Dr. Onur Pekcan
Civil Engineering Dept., METU _______________

Prof. Dr. Erdal Çokça
Civil Engineering Dept., METU _______________

Instr. Dr. Nabi Kartal Toker
Civil Engineering Dept., METU _______________

Dr. Timuçin Hergül
MEGA Mühendislik _______________

Date : 04.09.2014

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: **Gamze Didem ÖGET**

Signature:

ABSTRACT

ESTIMATING SWELLING CHARACTERISTICS OF CLAYS USING METHYLENE BLUE TEST - A MACHINE LEARNING APPROACH

Öget, Gamze Didem
M.S. Department of Civil Engineering
Supervisor : Asst. Prof. Dr. Onur Pekcan
Co-Supervisor: Prof. Dr. Erdal Çokça

September 2014, 158 Pages

Clayey soils tend to increase volume when they interact with water, by a phenomenon known as swelling. It is a major problem worldwide causing excessive economical damage for the infrastructure that needs to be taken into consideration. In order to avoid the damage, the identification of swell susceptible soils and predicting their swelling potential is a must. Our study mainly focuses on the prediction of swell potential of clayey soils using methylene blue (MB) test. A set of laboratory tests containing the physical properties of clays, MB tests and oedometer tests are performed. For this purpose, 32 samples obtained from different regions of Turkey are tested to obtain Atterberg limits, clay contents and methylene blue values (MBVs). In addition, maximum dry density, optimum water content, swell percent and swell pressure tests are conducted on 20 of these 32 samples. Then the laboratory data with similar characteristics available in the literature are compiled and combined with our data set to generate a comprehensive data base. Using this database, the swelling potential is examined such that the swell percent and MBV are predicted through physical characteristics. First multivariate linear regression technique is used to understand the relationships in the database. However, the results show that the variables in the database are not linearly correlated. Then a machine learning approach is utilized such that Genetic Expression Programming (GEP) and Artificial Neural Networks (ANN) are applied to understand the relations. The results prove that the nonlinear relationship can best be modeled using ANNs. The values of MAPE for the best models of Dataset I, II, III, and IV for MBV prediction are 4.2%, 5.0%, 11.5%, and 30.6%, respectively. The ones for the determination of swell percent for Dataset I, and II are 1.8% and 20.7%, respectively.

Keywords: expansive soils, swelling potential, clay, methylene blue test, multivariate linear regression, genetic expression programming, artificial neural networks

ÖZ

KİLLERİN ŞİŞME KARAKTERİSTİĞİNİN METİLEN MAVİSİ TESTİ İLE TESPİT EDİLMESİ – BİR YAPAY ZEKA YAKLAŞIMI

Öget, Gamze Didem

Yüksek Lisans, İnşaat Mühendisliği Bölümü

Tez Yöneticisi: Asst. Prof. Dr. Onur Pekcan

Ortak Tez Yönetici: Prof. Dr. Erdal Çokça

Eylül 2014, 158 Sayfa

Killi zeminler, suyla temasları halinde hacmen artarak, şişme eğilimi göstermektedir. Killerin şişmesi, altyapı elemanları üzerinde istenmeyen zararlara sebep olmakta ve dünya çapında hesaba katılması gereken önemli bir problem oluşturmaktadır. Bu hasarı önlemek için, şişme potansiyelli zeminleri önceden tanımlamak gerekmektedir. Bu çalışma, killi zeminlerin şişme yüzdelerinin, metilen mavisi testi ile belirlenmesi üstüne odaklanmaktadır. Bu amaçla, Türkiye'nin farklı bölgelerinden toplanan 32 adet numunede gerçekleştirilen deneyler ile numunelerin Atterberg limitleri, kil muhteviyatları ve metilen mavisi değerleri belirlenmiş olup, 32 numuneden 20'si üstünde ayrıca maksimum kuru yoğunluk, optimum su muhtevası, şişme yüzdesi ve şişme basınç testleri yapılmıştır. Daha sonra, deneylerden elde edilen datalar ile literatürden derlenen benzer özellikteki datalar birleştirilerek oluşturulan veri tabanı kullanılarak şişme yüzdesi ve metilen mavisi değerlerini tahmin eden modeller oluşturulmuştur. Öncelikle, veri tabanındaki ilişkileri anlamak amacıyla çok değişkenli doğrusal regresyon analizi tekniği kullanılmıştır. Ancak, sonuçlar veri tabanındaki değişkenler arasında doğrusal bir ilişki olmadığını göstermiş olup, bunun üzerine, yapay zeka yaklaşımı değerlendirilmiş ve bu kapsamda “genetik programlama” ve “yapay sinir ağları” metodları uygulanarak veriler arasındaki ilişki incelenmiştir. Bu analizlerden çıkan sonuçlar, veriler arasındaki modellerin en iyi “yapay sinir ağları” ile modellenebileceğini göstermiştir. Yapay sininr ağları kullanılarak metilen mavisi değeri tahmin etmek içi oluşturlan modellerde, mutlak ortalama hata yüzdesi değerleri, dataset I, II, III ve IV için sırasıyla 4.2%, 5.0%, 11.5%, ve 30.6% olarak elde edilmiştir. Bu değer şişme yüzdesi analizleri için ise dataset I ve II için, 1.8% ve 20.7% olarak belirlenmiştir.

Anahtar Kelimeler: Şişen zeminler, şişme potansiyeli, kil, metilen mavisi testi, çok değişkenli doğrusal regresyon, genetik programlama, yapay sinir ağları

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my supervisor, Asst. Prof. Dr. Onur PEKCAN for his extreme patience, encouragement and advise during my graduate studies. His guidance helped me during research, experiments phase and writing this thesis. I am thankful for his great interest and help

I would like to thank my co-supervisor Prof. Dr. Erdal ÇOKÇA for his support and valuable advises. He answered all my questions selflessly and patiently. I am thankful for his valuable advises, insight, encouragement and excellent guidance.

I am grateful to all soil mechanics laboratory staff and all people for all the help they gave during my experiments. In particular, I would like to express my sincere thanks to Ali BAL and Ulaş NACAR for their valuable support and friendship. Also, I am thankful to Yılmaz Emre SARIÇİÇEK, whom never withhold his help during thesis writing and endless experiments.

I am also thankful to Olcay BİNGÖL, who has always supported and motivated me during this process. He helped almost at every stage of this thesis and without his support this thesis would have never finished. In addition I would like to express my gratitude to Yaşam KIZILDAĞ, she helped selflessly when I needed the help most. I am grateful for her help and being my friend for almost a decade.

I am thankful to my friends at MEGA Mühendislik for their patience and friendship. In particular, I am thankful to my chief Timuçin HERGÜL for his limitless support and help during every stage of the thesis. I am also thankful to my dear fellows Görkem TUNCA, Kübra BEZGİN, Ali Günalp GÖRGÜLÜ and Mustafa PAZARBAŞIOĞLU for their help, tolerations and encouragements. Especially, I would thank to Hatice ERGİN for her her help during finding soil samples, her friendship and support.

I would like to express my thanks also to Caner UZUN and Anıl EKİCİ for their patience, help, inspirations and their belief in me during my thesis and graduation. In addition I want to specially thank to Nurettin Murat AZAP, Ayşe Nur YAZGAN, Ali Ersin DİNÇER, Ufuk OKUMUŞ, Deniz KARAER ARIK and Ali Can ARIK who have supported me and helped me stay focused on my study.

Last but not least, I would like to thank to my dear family. This thesis is dedicated to them without their support, love, patience and belief in me I would never have accomplished this. Hence, I would like to thank my mother Ayşe Seyhan ÖGET, my father Mehmet ÖGET and my brother Veysel Mert ÖGET for their love, understanding, toleration and support.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF SYMBOLS

A : Activity

AFNOR : Française de Normalization

ANN : Artificial Neural Networks

ASTM : American Society of Testing and Materials

BET : Brauner, Emmett and Teller gas adsorption technique

CC : Clay Content

CEC :Cation Exchange Capacity

Dc : Domain Constant

ρ_{dry} : Maximum Dry Density

Δh : Percent Swell

EI : Expansion Index

F : Student t-test

f_i : Fitness

F_r : Fraction passing No.4 sieve

FS : Free Swell

GEP : Genetic Expression Programming

I_s : Plasticity Index

ISS : Initial Soil Suction

K-expression	: Karva Language Expression
LL	: Liquid Limit
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MB	: Methyelene Blue
MBV	: Methylene Blue Value
N	: Hazard Coefficient
ORF	: Open Reading Frame
P_{ij}	: The value predicted
PI	: Plasticity Index
PL	: Plastic Limit
PVC	: Potential Volume Change
Q	: Square Root
R^2	: Coefficient of Determination
RNC	: Random Numerical Constant
SL	: Shrinkage Limit
SP	: Swell Percent
V	: Final Volume
V_0	: Initial Volume
VIF	: Variance Influence Factor
w	: Water Content

w_{opt} : Optimum Water Content

T_i : The target value

CHAPTER 1

INTRODUCTION

1.1. Problem Statement

Understanding the soil behavior requires determining its type and engineering properties. The classification of soils is performed according to their particle sizes. Silts, clays, sands, gravels, cobbles all have different physical properties. Soils with particle size smaller than 0.002 mm are defined as clay (Chen, 1975). Even though clays are defined based on their particle sizes, their mineralogy is the most important factor determining their governing behavior (Chen, 1975).

Clayey soils are much likely to change their volume when they interact with water. Unsaturated clays tend to increase their volume with water addition, similarly they are inclined to shrink when the water is removed. The swelling potential of clays are defined as the tendency of unsaturated clays to change their volume in the presence of water. Various factors such as the mineralogy, water content, regional climate, etc. can affect the swelling behavior of clays.

When light structures are considered, constructing them on clayey soils susceptible of swelling can cause severe economic damage. The cost of swelling damage can be millions of dollars on single and multi-story houses, walk ways, drive ways, parking areas, highways and streets, underground utilities, airports as well as swelling induced urban landslides (Jones and Holts, 1973). Demonstrations of case histories provide more insight to the swelling problem (Li et. al., 2013, Yenes et. al., 2012, Ozer et. al., 2011). Some examples of the damage due to swelling of clayey soils are shown in Figure 1.1 and Figure 1.2. Considering these structural and therefore the economic losses, understanding swelling behavior of clays and predicting the swelling potential become really important.

Figure 1.1. Damage to a Building by Swelling of Clays

(Yenes et. al., 2012)

Figure 1.2. A Damaged Pavement Due to Swelling of Soil (DiMillio, 1999)

Several methods are developed to understand the swelling behavior and its underlying mechanisms. For example, experimental methods provide insight into the swelling behavior and provide estimations for the volume change and swelling pressure. In this category, swelling potential determination using Atterberg Limit Tests (e.g.: Skempton, 1953, Holtz and Gibbs, 1956) and Potential Volume Change meter test (PVC) provide some correlations that classify swelling potential of the soil into one of the following groups; low, medium, high or very high. On the other hand, mineralogical identification such as X-ray diffraction of the clayey soil is a major tool when swelling behavior of clays is considered. When the swell percent and swell pressure of the clayey soils are pursued, soil suction measurements and oedometer tests can also be used determine those parameters. Although the above methods can predict the swelling potential of clays, they have some disadvantages such as being expensive or predictive models working well only for soils used in the experiments. In addition, some models use mathematical techniques such as linear regression which may not be capable of capturing the complex relations between the variables.

Another method to obtain swell characteristics of soils is through methylene blue (MB) test. It is an easy and practical dye test used for obtaining the maximum value of methylene blue dye that a clay sample can absorb, which is generally called methylene blue value (MBV). Relations between MBV and classification of the soil, swell percent or swell pressure are also studied in the literature. From the findings obtained so far, it is understood that there is a need to understand the relation between soil characteristics, MBV value and the swelling potential and generalize it using simple predictive models.

Although MB test is a simple and valid method for determining swelling characteristics of clayey soils, it is not commonly used in practice in Turkey. Therefore to ease the use of this test in practice, models estimating MBV from Atterberg limits and clay contents and models estimating swell percentage from MBV, Atterberg limits, dry unit weight, optimum water content and clay content are developed. This way an idea on the swelling percent of clayey soils can be obtained through MB test and simple soil laboratory tests such as Atterberg limits.

1.2. Objective

The main objective of Our study is to develop machine learning based models that can predict the swelling potential of clays reliably. With this context, the focus is given to estimate swell percent of clayey soils, using the soil characteristics obtained from the laboratory. With this aim in mind, another objective, to estimate the methylene blue (MB) values of the expansive soils have also emerged. Mehylene blue value of clayey soils is also estimated using the same characteristics and machine learning based statistical techniques.

Another objective is to compile all the laboratory data that contain MB tests and swelling characteristics together with the soil characterization tests. It is hoped that this data set will be a valuable resource for the other researches from all around the world. New techniques and tools that would be used to understand the swelling behavior of clays can be implemented using this database. The schematic representation of the methodology of the study is given in Figure 1.3 and Figure 1.4

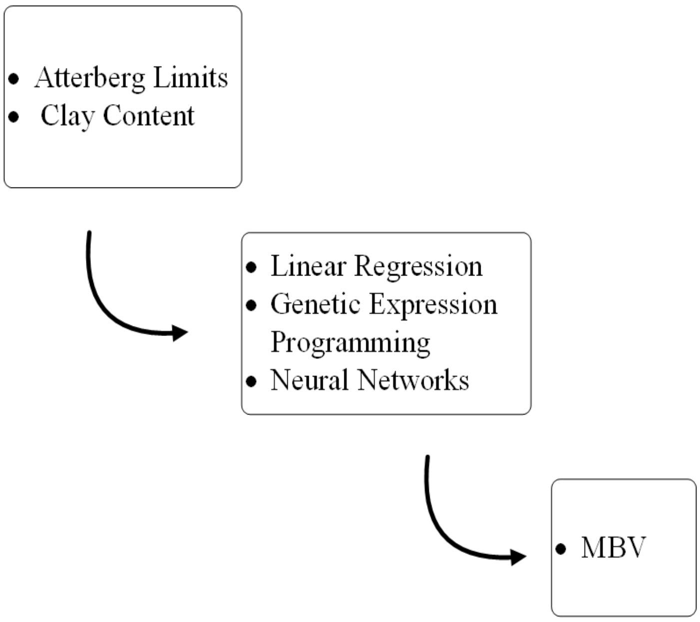


Figure 1.3. Schematic Representation of the Methodology of the Study (MBV Determination)

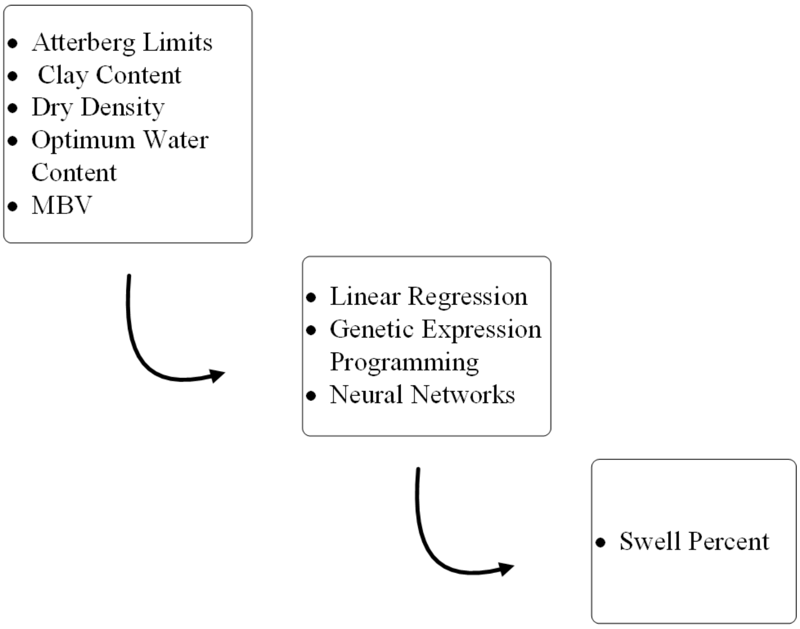


Figure 1.4. Schematic Representation of the Methodology of the Study (Swell Percent Determination)

1.3.Scope

Our study mainly utilizes the machine learning to determine the swelling potential of clays. In order for those techniques to work, the data set utilized should be as comprehensive as possible, which can only be compiled when the world literature is examined carefully. Therefore, in the scope of our study, the literature work related to estimations of swelling potential of clays is carefully examined.

A data pool combined from literature and laboratory work is used during the analysis. In the scope of our study 32 samples, obtained from different regions of Turkey are tested to obtain Atterberg limits, clay contents and MBVs. On a subset of these samples (20 samples) maximum dry density, optimum water content, swell percent and swell pressure tests are conducted. Other sophisticated laboratory experiments such as X-ray diffraction, scanning electron microscopy are kept out of the scope of our study.

In order to understand complex problems of nature, various learning techniques such as Neural Networks, Support Vector Machines, Decision Trees, Bayesian Learning Techniques, etc. could be used. Within our study, three methods are utilized for the determination of MBV and prediction of Swelling Potential; (i) Multivariate Linear Regression, (ii) Genetic Expression Programming, and (iii) Artificial Neural Networks. The reason behind the selection of these three specific methods is that they are simpler, widely used in the literature and generally produce very successful results. The other methods, although they can produce as good results as the above ones, are kept out of our study.

1.4.Thesis Organization

This thesis contains a total of five chapters: (1) Introduction, (2) Expansive Soils, (3) Machine Learning Methods (4) Experimental Study and Swellling Database, (5) Predictive Models, and (6) Summary, Conclusions and Future Studies. In Chapter 2, the literature studies related to expansive soils are presented. Within this context, the factors affecting the swelling potential of clays such as mineralogy, cation exchange capacity and diffuse double layer are provided to explain the mechanism of swelling. In addition, the factors affecting swelling characteristics, and methods to determine swelling potential are provided. In the next chapter, machine learning techniques used in our study are described. The algorithms of the methods and the important parameters which will be used in analyses are described. In chapter four, initially experimental study conducted in the scope of our study are given. The details of methylene blue and oedometer tests are given in this chapter. The results obtained from laboratory tests are also presented within this chapter. In addition, the collected data set from the literature are provided together with a discussion on the limitations of the compiled data. Then the analysis related to the characteristics of the dataset are presented.

In chapter five, the development of predictive models to estimate (i) methylene blue value of given clays and (ii) Swelling potential are described. Predictive models developed using linear regression, genetic expression programming and neural network analysis are presented. The results obtained using these methods are given and a discussion is provided at the end.

The last chapter summarizes the whole work presented in the thesis. It includes the summary of the thesis as well as the major findings. The conclusions of the theses are given in this chapter. Lastly, the chapter concludes with the provisions for the future studies and gives directions to possible future work fields.

CHAPTER 2

LITERATURE REVIEW

2.1.Expansive Soils

1.1.1. Swelling Mechanism of Clays

Swelling occurs when there is an environmental change; pressure release due to excavation, temperature increase or introduction of moisture around the clay in subject. However, the main reason of swelling is the introduction of water to the system. The swelling mechanism of clays is closely related to cation exchange capacity and diffuse double layer concepts. When water is introduced, attractive and repulsive forces are developed in clays. On wetting, the cations tend to diffuse due to osmotic pressures. Even though clay minerals tend to attract them, soil water chemistry is disturbed. If there is no external force to balance the corresponding change, then the particle spacing of the soil also changes (Nelson and Miller, 1992).

Even though clays are classified according to their particle sizes, their mineralogy is the most important factor in determining the behavior of the clay particles (Chen, 1975). Clay particles are formed by physical or/and chemical weathering of rocks. The rate, the chemical reaction type and the ratio of water (with respect to rock) determine the mineralogy of clayey soils (Velde, 1995).

2.1.1.1. Mineralogy

Clay minerals are chemically active minerals usually having sheet-shaped structures and accordingly high surface areas.

Basic unit of a clay particle is either silica tetrahedron or alumina octahedron. Isomorphous substitution of these elements is possible which creates different types of clay minerals. The basic unit of clay minerals is given in Figure 2.1.

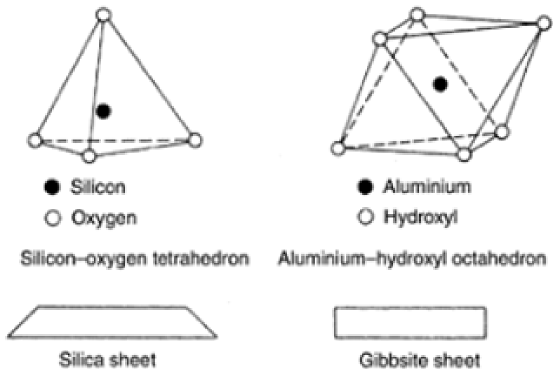


Figure 2.1. Clay Minerals Basic Units (Craig,2004)

Silica tetrahedrons or alumina octahedrons form sheets, where two or three layers of these sheets form clay minerals (Raj, 2008). Kaolinite unit is a two-layer unit, which is composed of gibbsite sheets and silica sheets on top of each other forming lattice of the mineral. Basic kaolinite units do not expand when saturated because they are formed by hydrogen bonding which results in a very stable bond that does not allow water to enter the lattice (Raj, 2008). Serpentine and halloysite minerals are two layer sheet minerals just like kaolinite except the water presence between the sheets.

Montmorillonite and illite are three-layer sheet minerals. Three-layer sheets are composed of silica and gibbsite sheets; gibbsite sheet being located between silica sheets. They are formed by the same minerals, but montmorillonite has water and exchangeable ions in between the sheets which makes the bond between silica sheets very weak. Swelling of montmorillonite mostly occurs due to the additional water adsorbed between the combined sheets (Craig, 2004).

Illite mineral has the same structure as montmorillonite mineral, only the space between combined sheets is occupied by non-exchangeable potassium ions. Also there is partial substitution of silicon by aluminum (Craig, 2004). Symbolic structure of the minerals are given in Figure 2.2 and Figure 2.3.

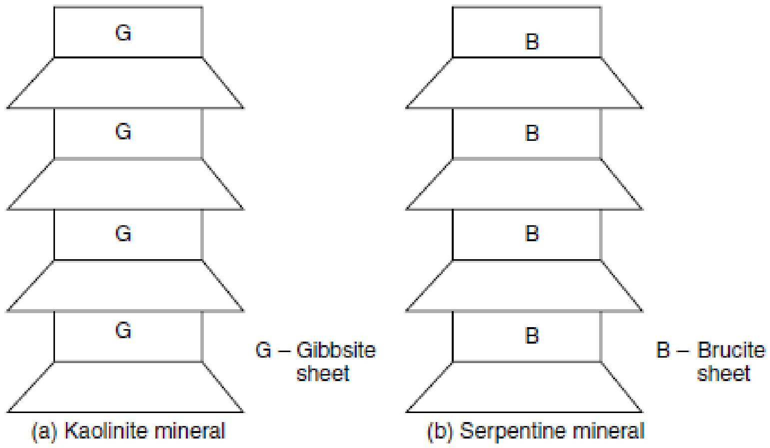


Figure 2.2. Symbolic Structures of Kaolinite and Serpentine (Raj, 2008)

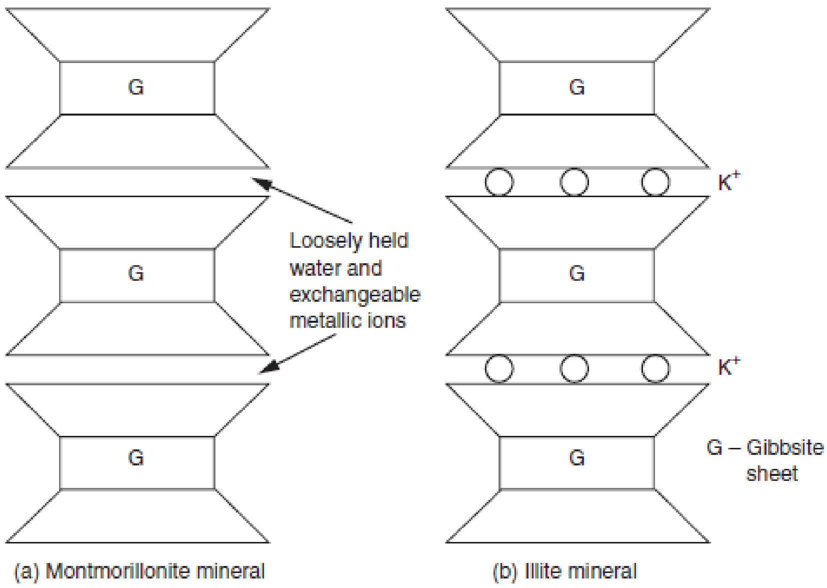


Figure 2.3. Symbolic Structures of Montmorillonite and Illite (Raj, 2008)

2.1.1.2. *Cation Exchange Capacity*

Clay minerals are able to adsorb anions and cations that are outside the structural unit and since they are negatively charged, cations are usually adsorbed. However one adsorbed cation can be replaced with another present cation. This replacement of excess cations is called cation exchange (Chen, 1975). Clay minerals' cation exchange capacity is the charge or electrical attraction of the cation per unit mass which is measured in milli-equivalent per 100 grams of soil.

Due to their structures, montmorillonites are 10 times more active than kaolinite on this matter. The most common exchangeable cations are Ca^{++}, Mg^{++}, H^{+}, K^{+}, NH_4^{+} and Na^{+}. The factors that determine cation exchange capacity of clays can be listed in three main titles (Grim, 1968).

1. *Isomorphous Substitution*, the replacement of the existing cation in the structure with a more active one, for example Al^{+3} for Si^{+4} in Silica sheet.
2. *Broken Bonds* around the particle edges and noncleavage surfaces (The major cause in Kaolinite).
3. *Replacement* of the hydrogen of an exposed *hydroxyl* (Mitchell, 1978).

The type of the exchangeable ion and the mineralogy of the particles affect the cation exchange capacity (Lambe and Whitman, 1969). Cation exchange capacity of kaolinite, illite and montmorillonite are given in Table 2.1.

Table 2.1. Cation Exchange Capacity of Some Clay Minerals (Terzaghi, Peck and Mesri, 1995)

Clay Minerals	CEC (mEq/100g)
Kaolinite	3-10
Illite	20-30
Montmorillonite	80-120

Cation exchange capacity has a definite relationship with Atterberg Limits. The relation between cation exchange capacity and methylene blue value is investigated in several studies (Birand and Çokça 1993, Yukselen and Kaya 2008). A part of Birand and Çokça's (1993) study was on determination of cation exchange capacity of clayey soils using methylene blue test. The samples obtained from Ankara was subjected to index, hydrometer, swell index and X-ray diffraction tests. Çokça and Birand (1993) determined the cation exchange capacity of clayey soils by methylene blue method and then compared results with the cation exchange capacity of the minerals obtained by X-ray diffraction tests. The authors concluded that methylene blue test is an easy and reliable method for determining the cation exchange capacity of clayey samples.

Yükselen and Kaya (2008) compared determination of specific surface area using methylene blue titration and N_2 adsorption methods, as well as prediction of cation exchange capacity by methylene blue spot test and NH_4-Na methods. The authors determined that the CEC values obtained by methylene blue test are lower than those obtained from NH_4-Na method but there is a linear relation between the two.

2.1.1.3. Diffuse Double Layer

Water molecules have an uneven charge distribution, dipolar character and are attracted to molecules in the solution. Clay minerals are negatively charged and they attract cations to balance the structure. When they are subjected to water, clay minerals attract cations, therefore the concentration of cation gets higher near the clay surface. Since concentration of cations increase near the clay surface, cations tend to diffuse, however electrostatic attraction attract them. The negative surface and the distributed charge in the adjacent phase are called the *diffuse double layer* (Mitchell, 1976).

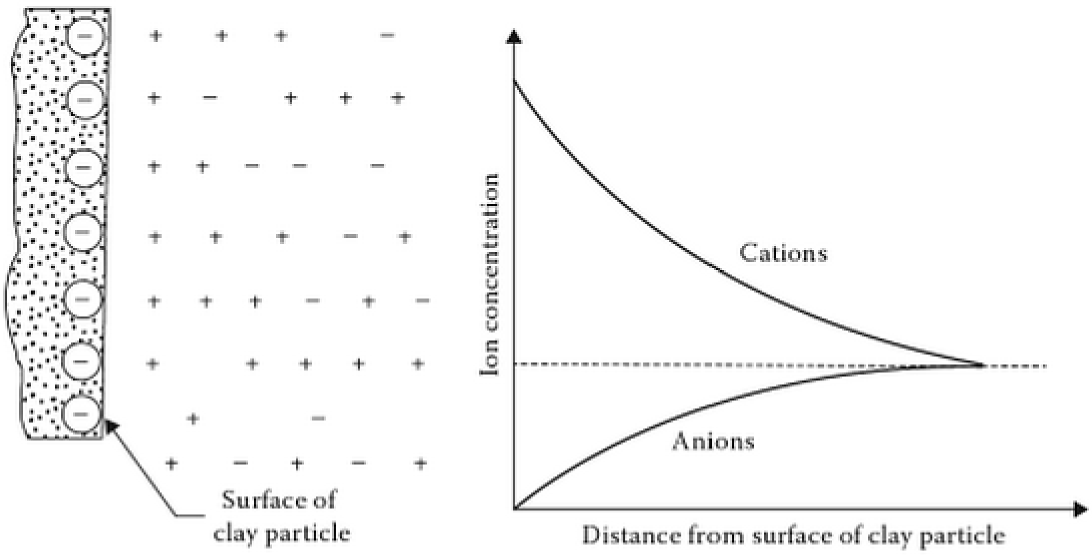


Figure 2.4. Diffuse Double Layer of Clays (Das, 2013)

The surface of the clay particle is more concentrated with positive charged particles and the anion concentration increases along the thickness of diffuse double layer. When two diffuse double layers overlap, the same charged particles create repulsive forces, therefore swelling potential of the clay minerals increase with increasing diffuse double layer thickness (Nelson and Miller, 1992).

There are several factors affecting diffuse double layer thickness. Diffuse double layer thickness depends on variations in surface charge density, surface potential, electrolyte concentration, cation valence, dielectric constant of the medium and temperature (Mitchell, 1976).

Particle spacing increases as the electrolyte concentration decreases, thus diffuse double layer thickness is increased (Mitchell, 1976). As cation valence decreases diffuse double layer thickness increases. Accordingly, for soils with identical mineralogy, the sample having exchangeable sodium cations (Na^{+}) will swell more than the sample possessing calcium cations (Ca^{+2}) (Nelson and Miller, 1992).

Double layer thickness increases as temperature increases, but increase in the temperature decreases dielectric constant of the medium. Therefore when both aspects are considered the effect is negligible (Mitchell, 1976).

2.1.2. Factors Affecting Swelling Characteristics of Clays

Swelling potential of clayey soils is affected by various factors that can be mainly divided into three groups. The internal structure of clays is an important factor that determines the swelling potential of the soil. Therefore one of these three main groups is *Properties of Soil.* The other important factors on swelling potential can be named as the *environmental conditions* and *the state of stress* (Nelson and Miller, 1992). The factors that influence swelling are presented in Table 2.2, Table 2.3 and Table 2.4.

Table 2.2. Soil Properties Influencing Volume Change (Nelson and Miller, 1992)

Factor	Description	References
Clay Mineralogy	Clay minerals which typically cause soil volume changes are montmorillonites, vermiculites, and some mixed layer minerals. Illites and kaolinites are frequently not expansive, however expansion can be expected when particle sizes are extremely fine.	Grim (1968); Mitchell (1973); Snethen et al. (1977)
Soil Water Chemistry	Swelling is repressed by increased cation concentration and increased cation valence. For example Mg+2 cations in the soil water would result in less swelling than Na+ cations.	Mitchell (1976)
Soil Suction	Soil suction is an independent effective stress variable, represented by the negative pore pressure in unsaturated soils. Soil suction is related to saturation, gravity, pore size and shape, surface tension and electrical and chemical characteristics of the soil particles and water.	Snethen (1980); Fredlund and Morgenstern (1977); Johnson (1973); Olsen and Langfelder (1965); Aitchison et al. (1965)
Plasticity	In general, soils that exhibit plastic behavior over wide ranges of moisture content and that have high liquid limits have greater potential for swelling and shrinking. Plasticity is an indicator of swell potential.	-
Soil Structure and Fabric	Flocculated clays tend to be more expansive than dispersed clays. Cemented particles reduce swell. Fabric and structure are altered by compaction at higher water content or remolding. Kneading compaction has been shown to create dispersed structures with lower swell potential than soils statically compacted at lower water contents.	Johnson and Snethen (1978); Seed et al. (1962a)
Dry Density	Higher densities usually indicate closer particle spacings, which may mean greater repulsive forces between particles and larger swelling potential.	Chen (1973); Komornik and David (1969); Uppal (1965)

Table 2.3. Environmental Conditions Influencing Volume Change (Nelson and Miller, 1992)

Factor	Description	References
1) Initial Moisture Condition	A desiccated expansive soil will have a higher affinity for water, or higher suction, than the same soil at higher water content. Conversely, a wet soil will lose water more readily than a relatively dry initial profile. The initial suction must be considered in conjunction with the expected range of final suction conditions.	
2) Moisture Variations	Changes in moisture in the active zone near the upper part of the profile primarily define heave. It is in those layers that the widest variation in moisture and volume change will occur.	Johnson(1969)
2.1) Climate	Amount and variation of precipitation and evapotranspiration greatly influence the moisture availability and depth of seasonal moisture fluctuation. Greatest seasonal heave occurs in semiarid climates that have pronounced, short wet periods.	Holland and Lawrence (1980)
2.2) Groundwater	Shallow water tables provide a source of moisture and fluctuating water tables contribute to moisture.	
2.3) Drainage and Manmade water sources	Surface drainage features, such as ponding around a poorly graded house foundation, provide sources of water at the surface; leaky plumbing can give the soil access to water at greater depth.	Krazynski (1980); Donaldson (1965)
2.4) Vegetation	Trees, shrubs and grasses deplete moisture from the soil through transpiration, and cause the soil to be differentially wetted in areas of varying vegetation.	Buckley(1974)
2.5) Permeability	Soils with higher permeabilities, particularly due to fissures and cracks in the field soil mass, allow faster migration of water and promote faster rates of swell.	Wise and Hudson (1971); De brujin (1965)
2.6) Temperature	Increasing temperatures cause moisture to diffuse to cooler areas beneath pavements and buildings.	Johnson and stroman (1976); Hamilton (1969)

Table 2.4. Stress Conditions Influencing Volume Change (Nelson and Miller, 1992)

Factor	Description	References
Stress History	An overconsolidated soil is more expansive than the normally consolidated same soil at the same void ratio. Swell pressures can increase on aging of compacted clays, but amount of swell under light loading has been shown to be unaffected by aging. Repeated wetting and drying tend to reduce swell in laboratory samples, but after a certain number of wetting-drying cycles, swell is unaffected.	Mitchell (1976); Kassif and Baker (1971)
In Situ Conditions	The initial stress state in a soil must be estimated in order to evaluate the probable consequences of loading the soil mass and/or altering the moisture environment therein. The initial effective stresses can be roughly determined through sampling and testing in a laboratory, or by making in situ measurements and observations.	
Loading	Magnitude of surcharge loads determines the amount of volume change that will occur for a given moisture content and density. An externally applied load acts to balance interparticle repulsive forces and reduces swell.	Holtz (1959)
Soil Profile	The thickness and location of potentially expansive layers in the profile considerably influence potential movement. Greatest movement will occur in profiles that have expansive clays extending from the surface to depths below the active zone. Less movement will occur if expansive soil is overlain by non-expansive material or overlies bedrock at a shallow depth.	Holand and Lawrence (1980)

2.1.3. Determining Swelling Potential

When determining swelling potential of clayey soils, two main subjects can be considered: (i) identification and classification of swelling characteristics, and (ii) heave prediction. There are numerous ways to identify and classify swelling soils.

As a matter of fact, even the definition of swelling potential lacks a standard. According to Holtz, 1959 "potential swell is the volume change of an air-dry undisturbed sample when saturated under 1 psi load." and according to Seed et al., 1962b, "swell potential is the volume change of a remolded sample at optimum moisture content and maximum dry density under 1 psi load." (cited in Nelson and Miller, 1992). In our study, definition of Seed et. al, 1962b is used. Experiments are conducted in the light of this definition and details of swell percentage test are presented in the following chapters.

2.1.3.1. Identification and Classification of Swelling Characteristics of Clayey Soils

Identification and classification of swelling characteristics of clayey soils are mostly conducted using soil index properties, mineralogical tests and cation exchange capacity. However there are also other tests used for classification, some of those tests are free swell test, potential volume change meter test (PVC Meter), and Expansion Index test.

2.1.3.1.1. Soil Index Properties

Classifications using soil index properties usually give a range of probable swelling, such as low or high swelling potential, range of swell pressure and/or swell percentage.

Soil index properties, mostly Atterberg limits of clayey soils, are used for identification and classification of clayey soils. Swelling potential of clayey soils are determined using various physical properties of clayey soils, such as liquid limit, plasticity index and etc.

Louisiana Department of Transportation uses liquid limit and plasticity index for classification of potential swell of clays, Kansas Highway Commission uses Plasticity Index, Raman Method uses plasticity index and I_s(LL-SL), Sowers Method uses plasticity index and shrinkage limit and etc. (Çokça, 1991).

Skempton, 1953 suggested the usage of activity of clays, combining Atterberg limits and clay content for determining swelling potential. Where activity is defined as;

$$A = \frac{PI}{CC} \qquad \text{Eqn. 2.1}$$

where,

A: Activity

PI: Plasticity Index (PI=LL-PL)

CC: Clay Content (% weight finer than 2µm)

Skempton suggested that, active clays are the most prone clays to swell. According to Skempton,1953 the activity of clays are ranged as;

<0.75	Inactive
0.75-1.25	Normal
>1.25	Active

Some example classifications are given below:

Table 2.5. Expansive Soil Classification (Holtz and Gibbs,1956)

Clay Content	Plasticity Index	Shrinkage Limit	Swell Potential (%)	Degree of expansion
>28	>35	<11	>30	Very High
20-31	25-41	7-12	20-30	High
13-23	15-28	10-16	10-20	Medium
<15	<18	>15	<10	Low

Table 2.6. Expansive Soil Classification (Chen,1988)

Plasticity Index	***Swell Potential***
>35	Very High
10-35	High
20-55	Medium
0-15	Low

Seed et. al. 1962b, also proposed a classification for compacted clays (Standard AASHTO compaction) at optimum water content and allowed to swell at 1 psi loading. Mitchell, 1993 developed the relationship using the data obtained from Seed et. al 1962b (Hergül, 2012) and determined the relation given below.

$$S = (3{,}6x10^{-5})xA^{2{,}44}xC^{3{,}44} \quad \text{Eqn. 2.2}$$

where,

S: Swell Potential

A:Activity

CC: Clay Content (% weight finer than 2µm)

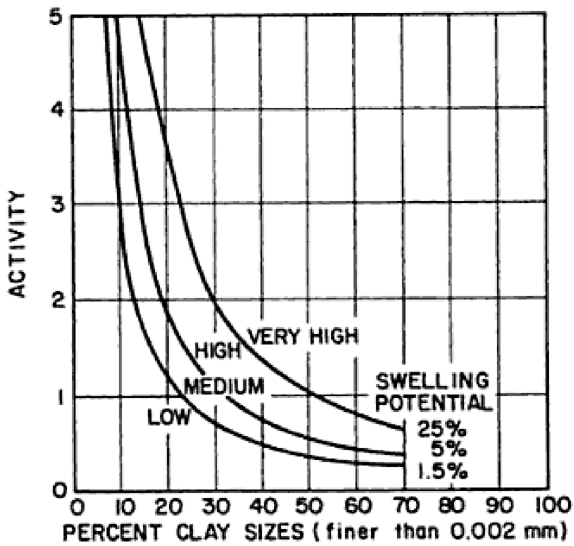


Figure 2.5. Swell potential classification (Seed et. al. 1962b)

2.1.3.1.2. Mineralogical Tests

Another way of determining swelling potential of clayey soils is by mineralogical tests. X-ray diffraction, differential thermal analysis and electron microscopy are used for determining mineralogy of clays (Nelson and Miller, 1992).

X-ray diffraction test is the most popular test among mineralogical tests. Basically, minerals are subjected to x-rays and according to the diffraction of x-rays, basal spacings of minerals are calculated. Basal spacing is characteristic for each mineral group and it is used for identification of the clay mineral.

2.1.3.1.3. Free Swell Test

Free Swell test is a test method to obtain volume change in a clayey sample, proposed by Holtz and Gibbs (1956). Clay sample is first oven dried and then sieved through No:40 sieve size. The known volume of the sample is then placed into a graduated cylinder and filled with water. The volume change in the clay sample is observed and percentage of swell is obtained.

$$FS = \frac{V - V_0}{V_0} \qquad \text{Eqn. 2.3}$$

where,

FS=Free Swell

V=Final volume of the sample

V_0=Initial volume of the soil

2.1.3.1.4. PVC Meter Test

PVC meter test was introduced by Lambe in 1960 to obtain a swell index. The samples obtained are compacted with modified proctor at its natural water content, after wetting the sample, it is allowed to change volume. Pressure on the ring is defined as Swell Index and a relation between Swell Index and PVC is determined by the Figure 2.6 below.

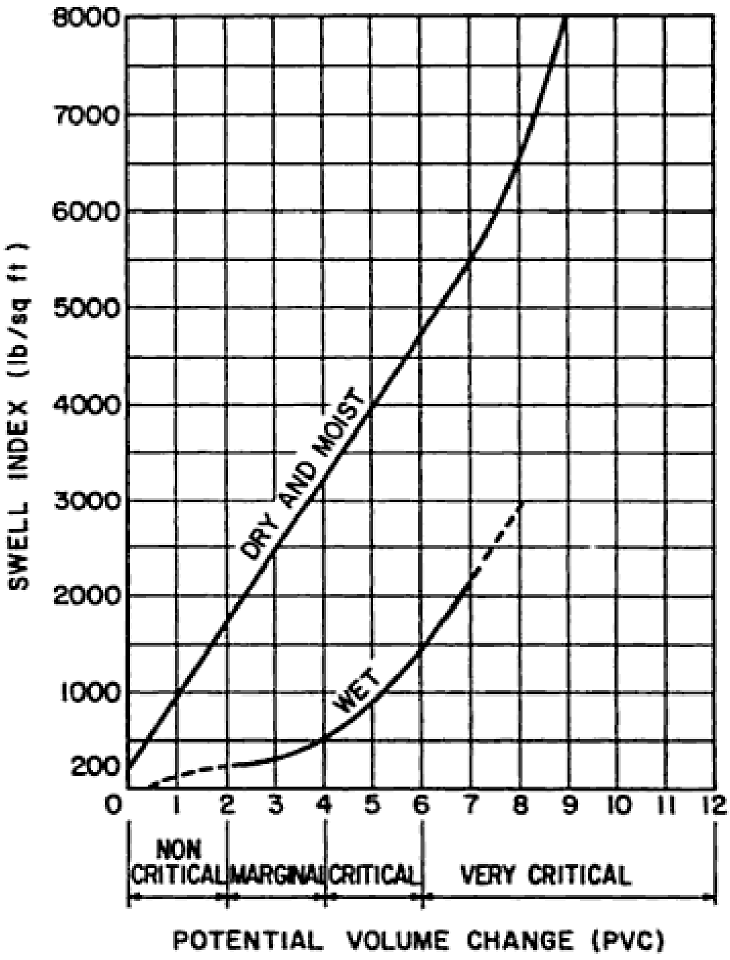


Figure 2.6. Swell Index vs PVC (Lambe, 1960)

2.1.3.1.5. Expansion Index Test

Expansion Index test was developed by the request of Uniform Building Code and California government. It is a very similar test to PVC with a few differences.

The samples are sieved through Sieve No:4 and water is added until the samples reach their optimum moisture content (According to ASTM D-1557-66T). Then the samples are cured for 6 to 30 hours and compacted into a standardized 4 in. diameter mold. The sample is adjusted to be at 50% moisture content and a 6.9 kPa load is applied. After the sample is wetted, the volume change is monitored for 24 hours and the expansion index is calculated as:

$$EI = 100\Delta h x F_r \qquad \text{Eqn. 2.4}$$

where,

Δh=percent Swell

F_r= Fraction passing No.4 sieve

Then the Table 2.7 is used for classification.

Table 2.7. Expansion Index-Expansion Potential

EI	Expansion Potential
0-20	Very Low
21-50	Low
51-90	Medium
91-130	High
>130	Very High

2.1.3.1.6. Cation Exchange Capacity

The studies show that cation exchange capacity of clays also play a great role on the swelling mechanism of clayey soils. Cation exchange capacity of clays can be determined by methylene blue test. This is one of the determination methods of swelling potential of clays using the physico-chemical properties of the soil. Methylene blue test is conducted on clay samples to find the cation exchange capacity of the soil. Basically the ion exchange capacity of clays is used to determine swelling potential of clayey soils.

2.1.3.2. Heave Prediction of Clayey Soils

Heave prediction of clayey soils is made according to two approaches (i) Soil suction measurements and (ii)oedometer tests, which are explained in the following sections.

2.1.3.2.1. Soil Suction

Soil materials above the ground water table is partially saturated where, particle and physico-chemical forces exerted by the soil creates a surface tension between the boundary of soil particle and water in the voids. This phenomena is termed as soil suction. It results with the arising of negative pore water pressure in the soil voids relative to the atmospheric pressure.

Matric suction and osmotic suction are two components of soil suction. Matric suction is the difference between the pore water pressure and pore air pressure. Osmotic suction can be described as the force applied on water molecules to even the concentration ratio of salt molecules present in the clay mineral.

Total suction is a function of matric suction and osmotic suction. However in geotechnical practice osmotic forces are constant, therefore the change in total suction is only due to matric suction (Krahn and Fredlund, 1972) (cited in Nelson and Miller, 1992).

The suction in a soil can be measured using many different methods, including; Tensiometer, filter paper, psychrometers, thermal matric potential sensors.

2.1.3.2.2. Oedometer Tests

The general purpose of oedometer test is to understand the consolidation settlement and drainage behaviour of a soil specimen. The test basically depends on measuring deformation characteristics of soil against different applied loads. By this way, actual site response and stress history of soil is tried to be estimated. However this test is also used for determining swelling characteristics of soils. Further information on the test method is presented in the chapter 4.

2.1.4. MB Test in Literature:

Various researchers studied dye adsorption by clay minerals for a long time in literature. Methylene blue test has various procedures and can be used for many different purposes. A slightly different version of methylene blue test is used for determination of clay amount in aggregates or rock types (Outhwaite and Morgan (1972), Hills and Pettifer (1985), Nikolides and Sarafidou (2007)). In addition, methylene blue test is also found to be a method of determination of cation exchange capacity and surface area of clayey soils.

One of the earliest papers on dye adsorption of clayey soils is "The Adsorption of Dyestuffs by Montmorillonite" (Emodi, 1949). She conducted a study on the chemical reaction that takes place when dye stuff is added to montmorillonite suspension. Base exchange and Van der Waals attraction is supposed to occur when the dye is mixed into the suspension. She concluded that simple base exchange was not observed and when methylene blue dye is considered, 46% of the calcium ions are trapped in the structure due to the placement of methylene blue molecules between the silica sheets.

Another study about methylene blue belongs to Robertson and Ward (1951) who described methylene blue dye preparation and adsorption. Cation exchange capacity deduction by methylene blue adsorption is studied and compared with barium ion exchange method. The authors pointed out that the obtaining chemically pure methylene blue is impossible and the impurity cannot be estimated. Also they stated that the quantity of water required for hydration is not constant. However they concluded that methylene blue adsorption is an easy and dependable test for cation exchange determination.

Fairbarn and Robertson (1956) studied the relation between liquid limit of clays and methylene blue adsorption. The study revealed a calibration between methylene blue value and Liquid Limit (LL) of clayey soils. However this calibration is only valid when clay samples are grouped according to their geological age.

Worrall (1958) studied adsorption of methylene blue dye by clayey soils. He concluded that "the adsorption of dyestuff is mainly by cation exchange capacity." He also stated that the adsorption of methylene blue is only reversible until a limit which is associated with surface area.

Brooks (1964) studied the mechanism of methylene blue adsorption and stated that there are other adsorption mechanisms rather than cation exchange. However MB test is decided to be a rapid and approximate way of determining cation exchange capacity.

Nevins and Weinritt (1967) studied determination of cation exchange capacity (cec) by methylene blue adsorption and compared the results with that obtained by ammonium accetate method. The study suggests that results are determined to be similar and MB test is a rapid and simple test. Robertson (1975) stated that the cation exchange capacity of clays can be found by methylene blue test. He also investigated the heat effect on clayey soils and determined that when clays are heated their methylene blue capacities are decreased (cited in Çokça, 1991).

Hang and Brindey (1970) studied methylene blue adsorption by Na saturated clays and determination of surface area and cation exchange capacities via methylene blue test. They used two methods to determine the exact methylene blue value. They stated that the adsorption of methylene blue by glass surface acts as an error source. They have conducted parallel tests to identify the amount of methylene blue. To identify surface area, Brauner, Emmett and Teller (BET) gas adsorption technique and for cation exchange capacity conventional titration technique was applied. Also X-ray diffraction measurements were taken on samples after adsorbing MB. They concluded that methylene blue adsorption method is an easy, simple and economical method for determining cation exchange capacity and surface area of clays.

The study of Brindley and Thompson (1970) is on the effect of initial cation saturation to methylene blue adsorption. Clay samples are saturated with Li^{+}, Na^{+}, K^{+}, Mg^{+2}, Ca^{+2}, Ba^{+2}, Fe^{+3}, Co^{+2}, Ni^{+2} cations. The researchers found different adsorption amounts and showed flocculation and particle size of clays as the reason for the differences.

Lan (1977) stated that specific surfaces of clay minerals can be determined using methylene blue test. The methylene blue adsorption capacity of clays indicates specific surface of clays. He also stated that MB value shows the activeness of the soil (cited in Çokça, 1991). Lan (1977) in another study proposed a correlation between plasticity index and MB value of a soil (cited in Çokça,1991).

Beaulieu (1979) focused on MB value and specific surface and proposed a relation for them (cited in Çokça,1991).

Locat, Lefebvre, Ballivy (1984) analyzed samples taken from eastern Canada to obtain a relation between index properties mineralogy and specific surface of clayey soils. Mineralogy of soils are determined by X-ray diffraction method and surface area of soils is determined using MB test. Authors suggest that " the specific surface area of a soil may be a single parameter that will correlate more significantly with the engineering index parameters."

Hills and Pettifer (1985) worked on the particle size effect. They compared the MB values of samples ground through 0,425 mm sieve and 0,075 mm sieve. They concluded that methylene blue value of coarser samples is about 60% of the methylene blue values of finer samples (cited in Çokça, 1991).

Lautrin (1987) used methylene blue value to obtain hazard coefficient (N). The author stated that the mineralogy and cation exchange are important when behavior of clays is considered. The hazard coefficient is defined as;

$$N = MB/CC \qquad \text{Eqn. 2.5}$$

where;

MB=Methylene blue value

CC=clay content

He determined that as CC increases MB value increases and suggested a classification using Hazard Coefficient Values. In another study, in the same year, Lautrin (1987) stated that MB test can also give an insight about the mineralogy of soils. Also, Lautrin (1989) criticized Activity (PI/CC) value of Skempton (1953) and suggested Hazard Coefficient is a more reliable method for determining swelling potential of soils (cited in Çokça, 1991).

Toureng and Lan (1989) studied using methylene blue test to detect clays in soils. They suggested a procedure and stated that MB test is an easy way of determining the properties of clay without having to separate the clay portion of the soil. They also suggested a relation between plasticity index and methylene blue values of clays (Cited in Çokça, 1991).

Magnan and Yousesefian (1989) studied classification of soils with the aid of grain size distribution curve based on colloidal activity using MB test (cited in Çokça, 1991).

Fourini Million-Devigne and Lan (1989) suggested that MB Value (Stain test results) is a good indicator of hazard potential of clays (cited in Çokça, 1991).

Çokça (1991), studied usage of methylene blue test for determination of swelling of Ankara Soils in his Phd thesis. He proposed a new swelling potential classification using methylene blue test and clay content of soils. The analyses were conducted on different samples from Ankara such as; Pliocene-Pleistocene Fluvial Lacustrine deposits (Terrace Deposits) and Recent Alluvial Deposits. 22 remolded samples from Terrace deposits, 6 remolded samples from alluvial deposits and 2 remolded samples from residual deposits were taken. Index tests, PVC meter tests, X-ray diffraction tests and methylene blue tests were performed on the samples. Also data from earlier studies were combined and searched for the same purpose. The proposed classification system was developed using the data obtained from the tests run on the samples obtained from Ankara. The distribution of methylene blue value versus clay content graphs of the data are given in Figure 2.7 and Figure 2.8 and the proposed classification system is given in Figure 2.9.

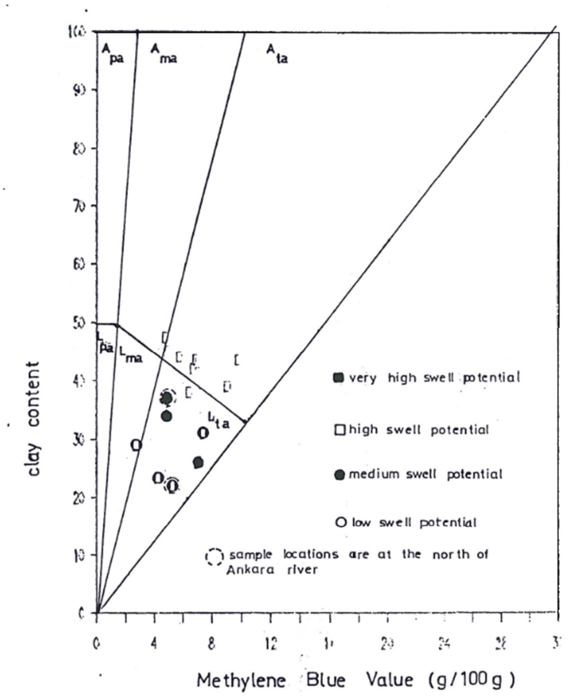


Figure 2.7. CC vs MBV Chart (Alluvial Soils) (Çokça, 1991)

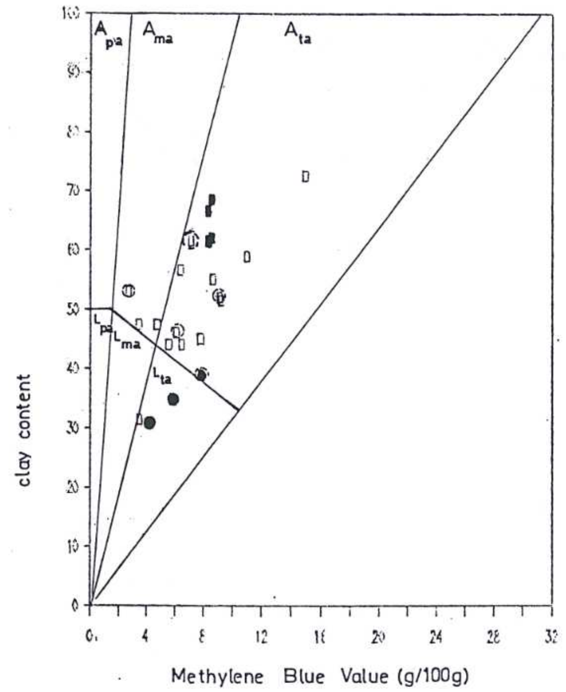


Figure 2.8. CC vs MBV Chart (Terrace Deposits) (Çokça, 1991)

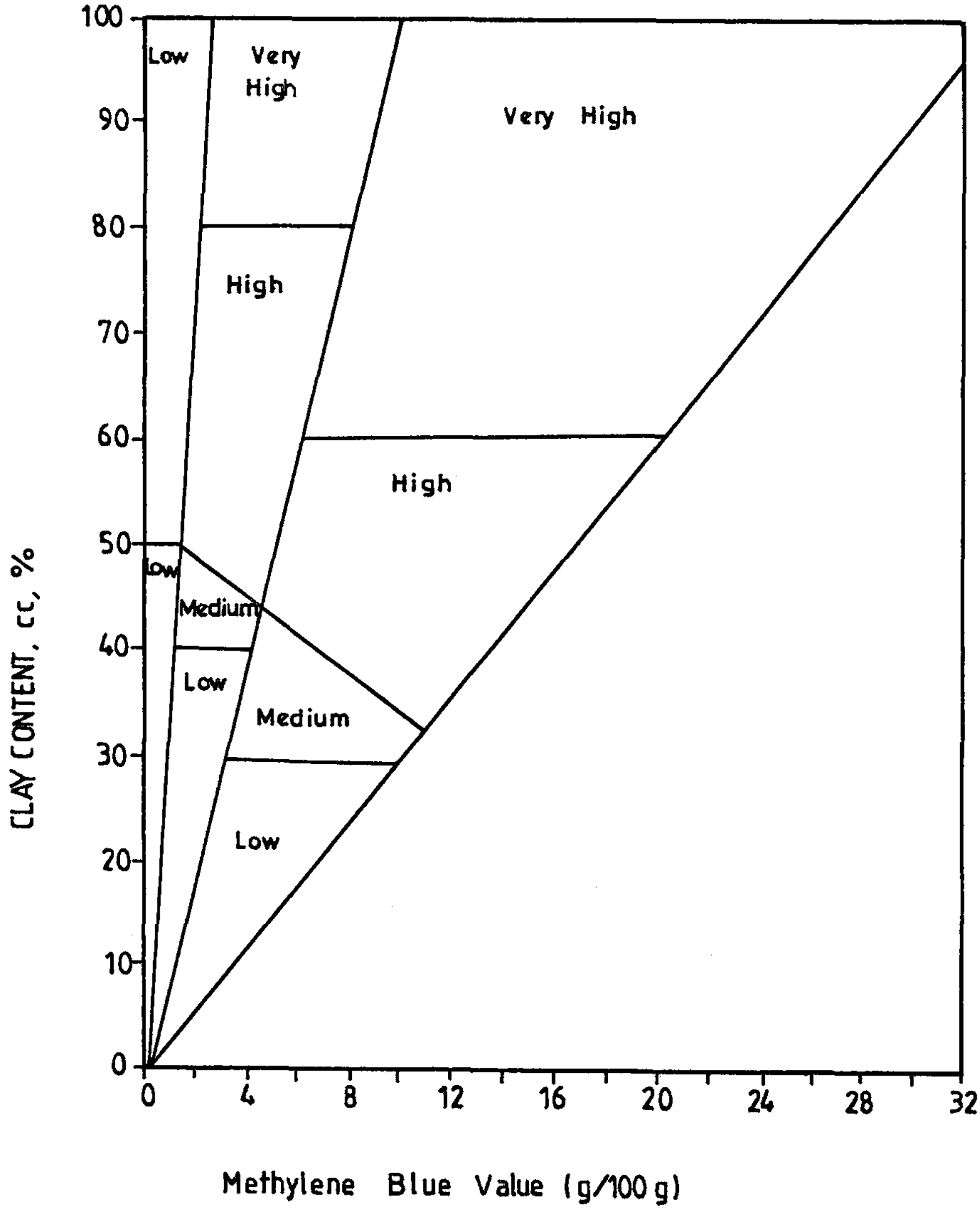


Figure 2.9. Proposed Classification method (Çokça, 1991)

Aringhieri et. al. (1992) studied determination of specific surface area of clays using MB test. They used three different methods to obtain and compare the results of the determined surface area. They used methylene blue adsorption method, N_2(BET) adsorption technique and water-vapor (H_2O) adsorption technique. They have stated that methylene blue cannot get in reaction with all of the surface of a bottleneck particle.

Çokça and Birand (1993) conducted studies proving that methylene blue test can be used for determining swelling potential of clayey soils as well as cation exchange capacity of clayey soils.

Karahan (1999) studied the relation between the swelling percentage, swell pressure and suction capacity, methylene blue value of expansive soils in his master's thesis. Karahan performed oedometer test, Atterberg limit tests, and suction measurements on artificial samples. 13 samples were created having different plasticity indices and suction was measured using Thermocouple Pyschrometer. He concluded that there is strong relation between swell percent, swell pressure and methylene blue value with Liquid Limit, plasticity limit, plasticity index and clay content. As a result of the study, Karahan stated that methylene blue value increases as swell percent, swell pressure and suction increases.

Çokça (2002) continued the study of Karahan (1999) and he suggested the following equation for determination of swell pressure in his paper;

$$Swell\ (\%) = -121.807 + (12.1969xMBV) + [27.6579xLog_{10}(ISS)]$$
Eqn. 2.6

Fityus, Smith and Jennar (2000), compared shrink swell test and methylene blue test on samples and determined a correlation between MBV and Shrink-swell Index. They have concluded that MB test is a fast and reliable test for determining surface area and cation exchange capacity of clayey soils. Although they have found a correlation between MBV and Shrink-Swell Index, the authors noted that the accuracy of the model is ±0.75 shrink-swell units which may be unacceptable for many applications.

Ergüler (2001) studied the effect of disturbance on swelling of Ankara Clay. He performed soil index tests (such as Atterberg limits, Unit weight, Specific Gravity, Sieve analysis) to obtain physical properties of soils. In order to obtain mineralogy of the samples, X-ray diffraction and major element analysis are performed. He also performed free swell test, modified free swell test, methylene blue test and oedometer tests in order to obtain swell characteristics of soils. In our study also an approach called $w_{max(24,72)}$ is proposed in order to obtain swelling characteristics of soils. He determined considerable relation between swelling characteristics of soils using especially w_{max24} MBV, smectite ratio and liquid limit and maximum dry unit weight of soils

Chiappome et. al (2004) studied the mineral characterization of clays using methylene blue test. In our study, the authors have compared two standards of methylene blue test by Française de Normalization (AFNOR) and by American Society of Testing and Materials (ASTM). They performed grain size and hydrometer analysis, Atterberg limit tests, X-ray diffraction tests and tests for determination of swelling pressure on the samples that were obtained from Piedomont, Italy. They have determined that there is possibility of obtaining a meaningful correlation between methylene blue value and swelling characteristics of clayey soils. They also suggest that MB test can be used for determination of character of clayey soils.

Claudia M., (2004) compared methylene blue test to other numerous swelling determination methods and determined a correlation. In order to use in the experiments, Claudia obtained 200 samples from Apennines-Italy. Dry density, water content, Atterberg limits, grain size analysis, sedimentometry, specific gravity, X-ray diffraction and methylene blue adsorption tests were performed on these samples. Swelling percentage and swelling pressure were determined using ASTM D4546-85 oedometer tests. In our study all regression models were analyzed and a new classification with regard to MBV is proposed. The classification proposed by Claudia (2000), is given in Table 2.8.

Table 2.8. Classification Based on Methylene Blue Value (Claudia, 2000)

MBV (g/100g of soil)	PI	SP (kPa)	Swelling/Shrinkage Potential
<2,5	<12	0-100	Low
2,5-4,5	12-35	100-300	Medium
4,5-9,0	35-45	300-500	High
>9	>45	>500	Very High

Türköz (2007) studied Harran Clay for determination of swelling potential in his doctoral thesis. He used direct methods to determine swelling potential of clayey soils as well as methylene blue test and compared the results. He determined Atterberg limits, water content, unit weight, specific gravity, dry unit weight and clay contents of the samples. In our study, methylene blue test, swell percent and swell pressure (ASTM D4546, Test Method B) were conducted on the samples. In our study swell percent is found using expansion index test and swell pressure is obtained by conducting PVC meter test. He studied the effect of disturbance of samples on swelling determination. He performed Neural network analysis and fuzzy logic analysis. He also used Geographical Information Systems in order to obtain the swelling potential map of the area.

Yükselen and Kaya (2008) published “Suitability of the methylene blue test for surface area, cation exchange capacity and swell potential determination of clayey soils". On 16 samples, Atterberg limit, specific gravity, hydrometer tests were conducted. X-ray powder diffraction patterns were obtained using diffractometer and CuKα radiation. The researchers performed also, N_2 adsorption method, methylene blue-spot test method and methylene blue- titration method on samples. They have concluded that MB test is a suitable method for determining specific surface area, cation exchange capacity and swelling behavior of soils.

2.2.Machine Learning Methods

The main work of this thesis consists of developing models to estimate (i) swell percent and (ii) Methylene Blue Value (MBV) of clays. In this chapter, two well-known machine learning algorithms (i) genetic expression programming (GEP) and (ii) artificial neural networks (ANN) are defined.

2.2.1. Genetic Expression Programming

2.2.1.1. Concept and Introduction

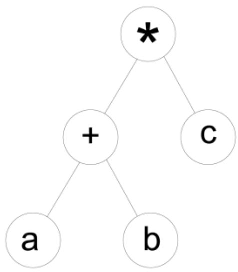
Genetic programming is a biologically inspired method, which employs genetic structures of living organisms into computer programming. A computer program which can create the required programs by itself is evolved using genetic operators such as; mutation, reproduction etc.

The first researcher to suggest using machine intelligence on this subject is Turing, (1948). Turing's objective was to develop computers which can solve problems on their own using artificial intelligence. Which is stated in his words as, "*The aim is to get machines to exhibit behavior, which if done by humans, would be assumed to involve the use intelligence, (Samuel,1983)*" (Cited in Koza and Poli, 2005).

One of the first improvements on genetic programming was by Samuel, (1959) who used machine learning so the computers could program themselves. This first progress in machine learning resulted in several researchers working on the subject, succeeding in creating new algorithms which were mostly concentrated on genetic algorithms This algorithm utilizes genetic operators such as mutation, reproduction, etc., and searches for the best solution in a specified database.

Genetic algorithm, genetic programming and genetic expression programming are similar methods where, the main difference of the three methods lies in chromosome definition. Genetic algorithms search for the best individual of a specified population using genetic operators and linear strings of fixed length named as chromosomes. On the other hand in genetic programming, rather than linear strings, "nonlinear entities of different sizes and shapes (parse trees)" are employed. Genetic expression programming combines the two different approach and puts linear strings(fixed length chromosomes) into use and defines the resulting programs as nonlinear entities of different sizes and shapes, namely, Expression trees (Ferreira, 2001).

The employment of tree structures in genetic programming was first introduced by Koza, (1992). Tree structures are a type of structure that represents the data/regression with organized branches and nodes where, branches are lines connecting literals and functions, which are called the nodes. A simple figure of a tree structure is given in Figure 2.10.



Equation: (a+b)xc

Figure 2.10. Basic Tree Structure

At the examples given in this chapter and at the analyses results for that manner '+' is used as the linking operator. The bold portions of the genes are tails and as can be seen from the figure tails contain noncoding regions which allow genes to be at the same length despite the program length. At the examples a, b, c are the literals and (*, +, /, -, Q) are mathematical operators, where Q operator represents the square root function.

The goal of genetic expression programming is to obtain a suitable computer program for the data pool given. Therefore initially the data are prepared for the analyses. Normalization described at this chapter is applied on the data. Then the data is divided into 2 groups, as training and testing data where, in our study, 10% of the initial dataset is used as testing data and the remaining 90% is used as training data. The flowchart representing the genetic expression algorithm is given in Figure 2.11.

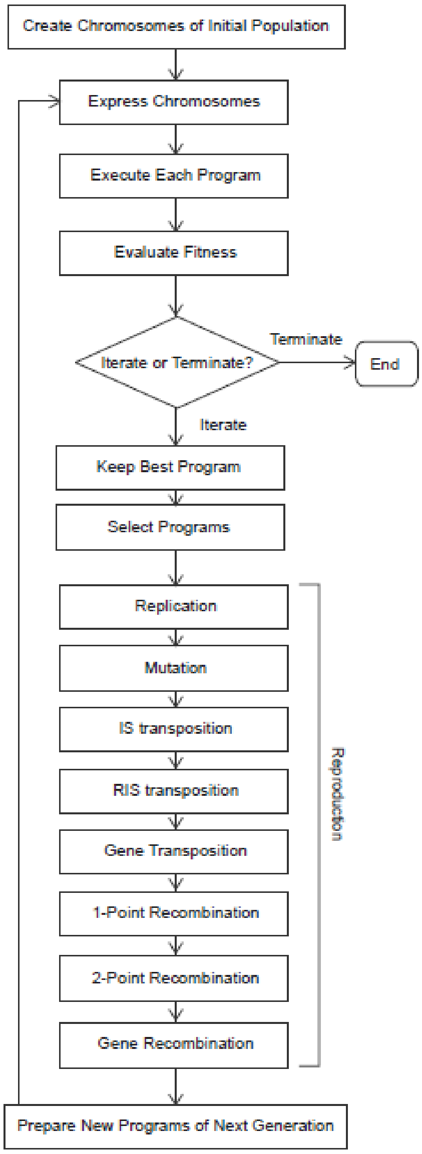


Figure 2.11. The flow chart of a genetic expression algorithm (Ferreira, 2001)

2.2.1.2. Components of Genetic Expression Programming

Genetic expression programming uses chromosomes composed of one or more genes and applies genetic operators such as, replication, inversion, mutation, transposition and recombination on them in order to create a valid program.

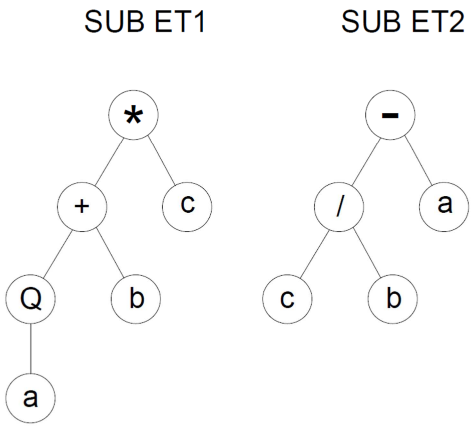
The genes creating the chromosomes consist of a head and tail where head is composed of both terminal elements and/or functions, unlike the tail which is solely composed of terminal elements. Genetic expression programming uses constant length chromosomes while creating the programs. However The program does not end where gene ends but where the last available program at the head of the gene is used. The part used at the coding is called "Open Reading Frames (ORFs). It is not the length of the gene or chromosome that determines the length of the program but the ORFs. Consequently a gene can posses noncoding areas which is used during the application of the genetic operators only.

Genetic expression programming uses karva language (K-expression) as its coding language, where K-expression is composed of the genes present in the chromosome including the noncoding region.

Each gene in a chromosome are decoded with separate expression trees and the resulting program is obtained by combining each expression tree with a linking function. Linking functions are simple mathematical operators (+, -, *, /) and the chosen operator is applied to get the resulting program.

In order to visualize ORFs and the resulting program of a genetic expression program, a chromosome consisting of 2 genes and both its expression tree and K-expression are given as an example in Figure 2.12.

0 1 2 3 4 5 6 7 8 0 1 2 3 4 5 6 7 8
* + c Q b a c b a - / a c b c a c c



Equation: $(\sqrt{a}+b)\mathrm{c}+\frac{c}{b}-a$

Figure 2.12. K-expression and Expression tree

There are five preparatory steps before running an analyses of genetic programming or for that matter genetic expression programming (Koza and Poli, 2005).

- ✓ Initially the terminals should be specified. Terminals can be defined as the input parameters although they may be independent parameters, zero argument functions or/and random constants that need to be defined prior to the analyses.
- ✓ Following definitions of the terminals, the function pool, that is going to be used at the analyses, is defined. This pool will contain the functions which are necessary for creating the head of the genes. These functions vary from simple mathematical functions such as adding, subtracting, multiplication and division to more complex mathematical or some logical functions, logistical terms.
- ✓ The third step is defining the fitness measure for analyses. The fitness of individuals created in the population can be computed using different

approaches but it is the difference between the known values-computed values of the individual in the population.

- ✓ In the fourth step parametric definitions, such as mutation, replication should be defined. According to Koza and Poli (2005) the most important parameter to select is the population size.
- ✓ The final preparatory step is to determine the termination criterion. Termination criterion should be chosen according to the program and can be defined as the maximum generation number, a defined maximum fitness value or maximum fitness.

When genetic programming is commenced, first a random initial population of programs are created using the defined functions. Until the termination criteria is met, genetic operators are applied on the constantly changing population.

The variation in the population is not achieved by replication alone, genetic operators mentioned above are also applied in order to get higher variety. These operators are randomly applied on chromosomes however each operator is allowed to modify a chromosome only once except for the mutation operator. On the other hand, chromosomes can be subjected to one or more genetic operations or may not be subjected to it at all. (Ferreira, 2001).

2.2.1.2.1. Genetic Operators

Genetic operators used at genetic expression programming are:

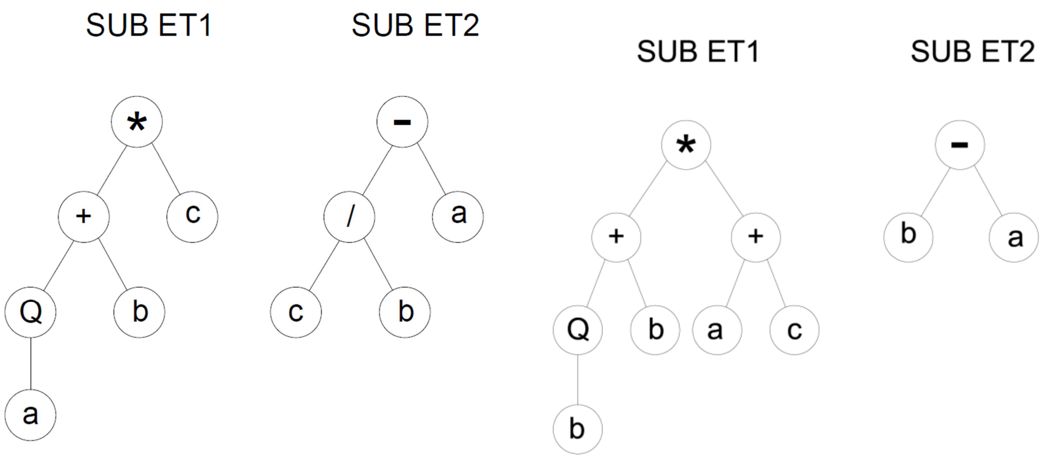
- ✓ Replication
- ✓ Inversion
- ✓ Mutation
- ✓ Transposition
- ✓ Recombination.

Replication is the genetic operator where the chromosomes are copied to the next generation according to their fitness.

Mutation can occur anywhere at the gene. However when mutation is in progress at the tail of the gene, terminals can be changed with terminals only, when it is occurring in the head there are no constraints. The reason behind this restriction is to ensure that the new individuals are structurally correct programs. An example of a mutation on K-expression is given in Figure 2.13.

0 1 2 3 4 5 6 7 8 0 1 2 3 4 5 6 7 8
* + c Q b a c b a - / a c b c a c c

0 1 2 3 4 5 6 7 8 0 1 2 3 4 5 6 7 8
* + + Q b a c b a -b a c b c a c c



Equation: $(\sqrt{a}+b)\mathrm{c}+\frac{c}{b}-a$

Equation: $(\sqrt{b}+b)*(a+c)+(b-a)$

(a) (b)

Figure 2.13. Example of a mutation (a) Sub-expression trees before mutation, (b) Sub-expression trees after mutation

As can be seen from the example the whole structure of the chromosome has changed. The resulting chromosomes represent a whole new program.

Transposition is basically copying and/or changing the location of a randomly chosen sequence in the chromosome. There are three types of transpositions, which are, Transposition of insertion sequence elements, root transposition and gene transposition.

- ✓ The transposition of insertion sequence elements is that a randomly chosen insertion sequence element is copied and placed into a randomly chosen location on the chromosome. To maintain the genes fixed length all the elements are shifted and the excess elements at the end of the gene are removed.
- ✓ Root transposition consists of copying and moving of a sequence starting with a function to the root of the gene (starting point of the gene).
- ✓ Gene transposition consists of selection of a random gene (except the first one) and moving it to the first position on the chromosome. Unlike other transposition methods the selected gene is removed from its original position.

Examples of transposition are given in Figure 2.14.

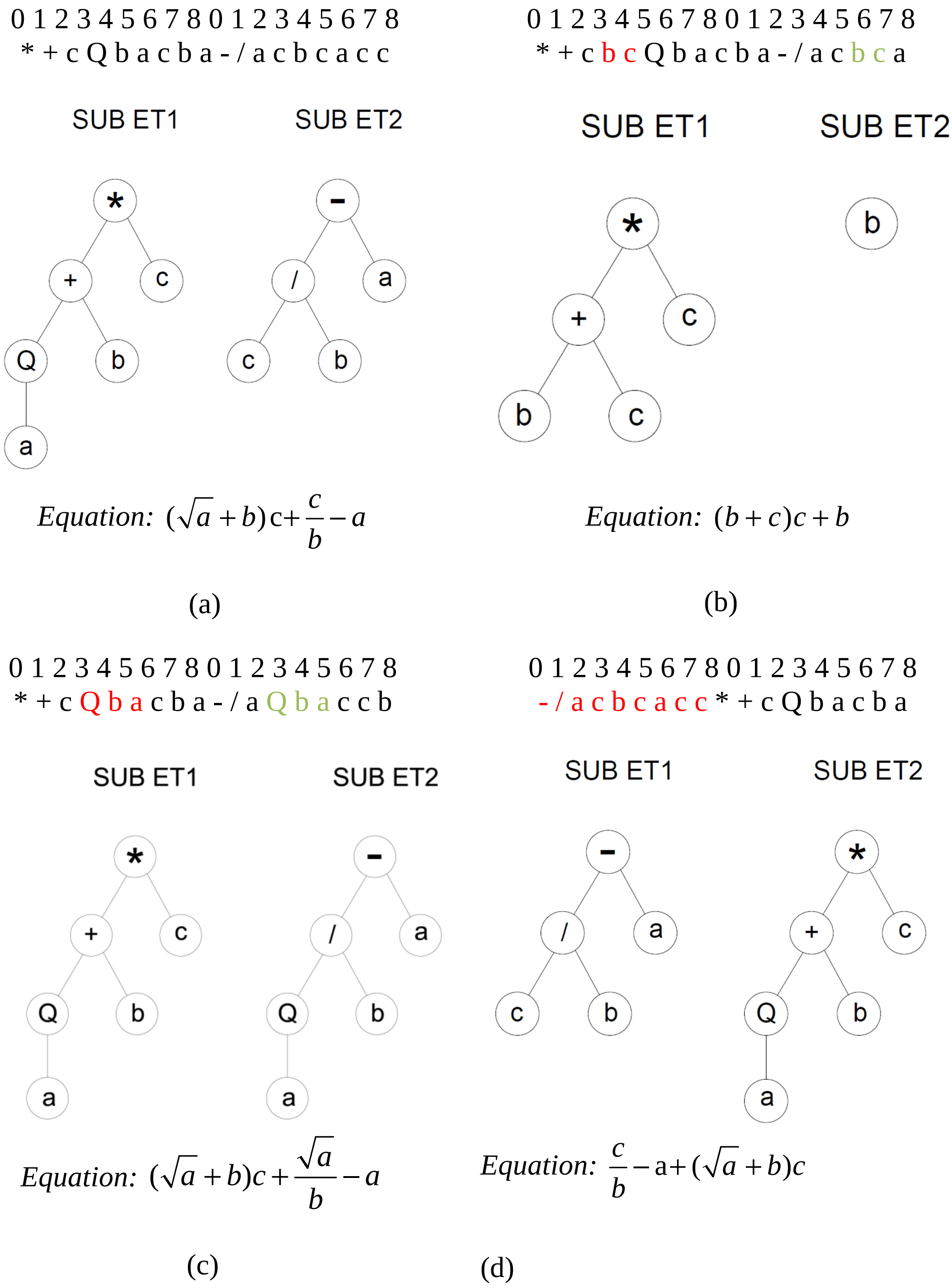


Figure 2.14. Examples of transposition –Subexpression trees (a) Before (b) After Insertion (c) After root transposition (d) After gene transposition

Recombination is the genetic operator which allows the exchange of elements, leading to the production of the daughter chromosomes. There are 3 types of recombination, which are, one point recombination, two point recombination and gene recombination.

- ✓ During one point recombination, a random point is selected at two different chromosomes and the elements of the chromosomes are exchanged with respect to this point.
- ✓ During two point recombination, two points are randomly selected on two different chromosomes and the elements between the chromosomes are exchanged.
- ✓ During gene recombination, randomly selected genes of two different chromosomes are exchanged.

Examples of Recombination are given in Figure 2.15, Figure 2.16 and Figure 2.17. The points marked with an arrow are the recombination points. The red and black portion of the chromosomes are swithched during the recombination process.

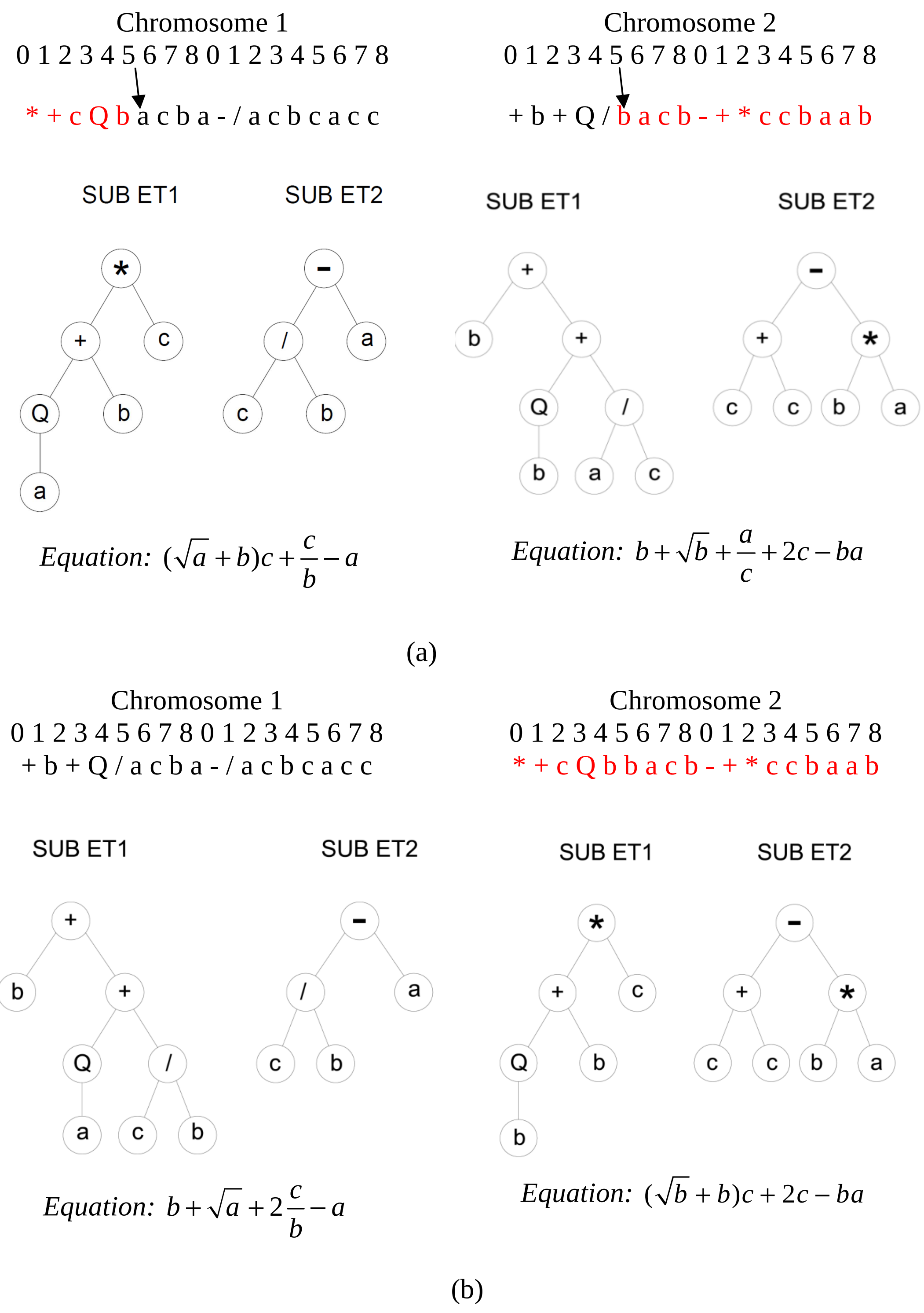


Figure 2.15. Example of one point recombination - Chromosomes (a) Before and (b) After one point recombination

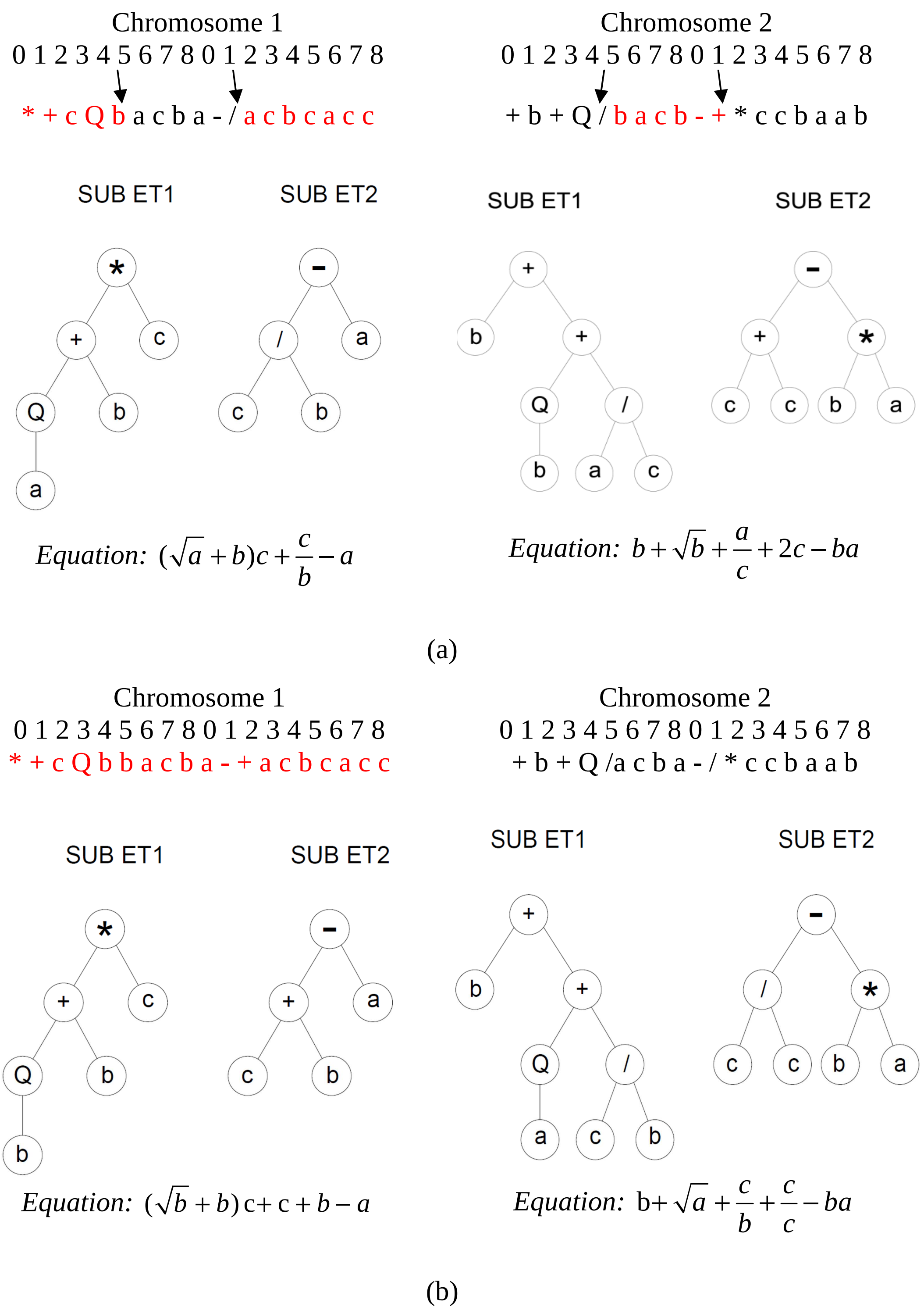


Figure 2.16. Example of two point recombination - Chromosomes (a) Before and (b) After two point recombination

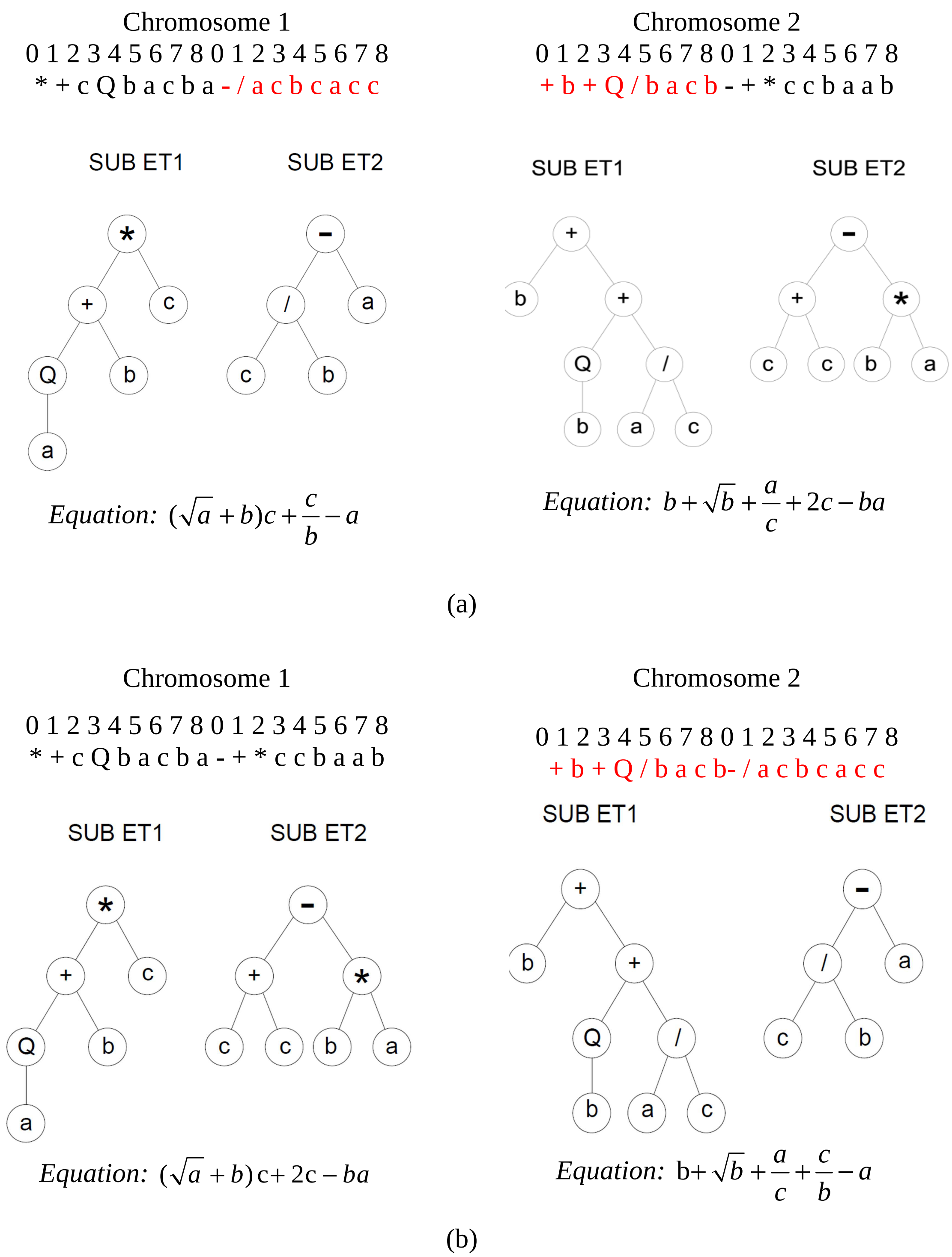
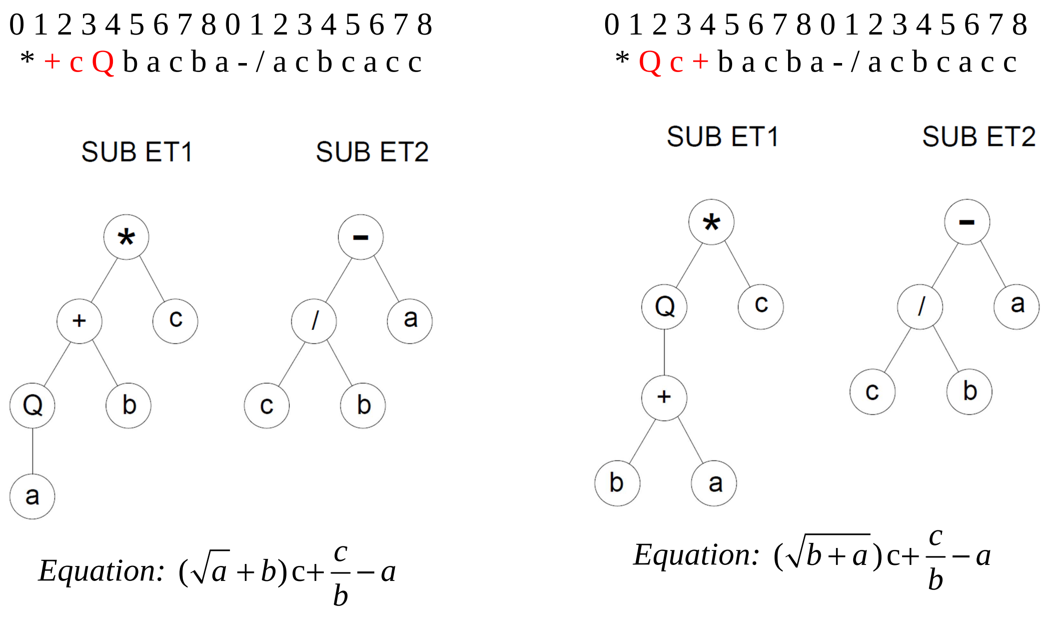


Figure 2.17. Example of gene recombination - Chromosomes (a) Before and (b) After gene recombination

Inversion operator randomly chooses the gene and reverses the alignment of the randomly chosen sequence. Example of an inversion is given at Figure 2.18.



(a) (b)

Figure 2.18. Example of an inversion Chromose (a)before and (b) after Inversion

Also *random numerical constants* are used at the analyses where, the number of numerical constants, the range and the data type are specified. Mutation, Inversion and transposition operators are also applicable to numerical constants. Numerical constants can be mutated in two different ways, either the numerical constant can be mutated directly (RNC mutation) such as from -1.600 to 0.525 or the number representing the constant variable can be changed (DC mutation) such as from c1 to c5 (the values of c1 and c5 are kept the same but the place of the constant in the gene is changed). The inversion and transposition operator are applied to numerical constants in the same manner as described above.

2.2.2. Neural Networks

Neural networks are biologically inspired systems that provide a solution using well interconnected processing elements called 'neurons'. The problem can be a classification problem or as in this case target fitting problem. Supervised neural networks are trained using accumulated knowledge. Neural networks do not provide a function in the known sense. It creates a system that predicts the expected solution according to the input fed using the prior knowledge it obtained through training.

This section comprises of detailed introduction to neural networks and presents the results of regression analyses.

Neural networks are computational networks based on simply the nervous system of a human body; the performance of human brain; *"a highly complex nonlinear and parallel compute*r" *(Haykin, 2009)* inspired neural network methodology.

Nervous system of a human body uses quite a number of simple structures-neurons to solve complex problems as visual recognition etc since neural networks are inspired by the human brain, they are also massively interconnected. This structure of neural networks and the ability to learn is achieved by back propagation algorithm, which allows networks to generalize. Thus neural networks are successful at solving non-linear problems and can be trained easily and are error tolerant.

McCulloch and Pitts first introduced the idea of such a computational program in 1943 (McCulloch and Pitts, 1943). The authors created a network based on threshold function (all or none model), mimicking the nervous system of a human body, able to compute any computable function. Soon after the authors have published “Recognition of spacial patterns by neural networks” where they used neural networks to model a practical problem-spacial patterns. In 1949 Hebb published *"The Organization of Behaviour"* which includes, now known as, Hebbian rule, suggesting that if one neuron activates another repeatedly, synaptic strength (their effectiveness) increases thus education of neurons can be possible.

Rosenblatt proposed a new approach to neural networks, perceptron convergence theorem (Rosenblatt, 1960b) which was formulated in 1961. Perceptron converge procedure is an iterative procedure where, initially weights and threshold values are appointed, then at the second step a new input vector and the desired output are presented to the system and outputs are calculated. After the calculations, weights are adapted in according to the residuals and weights are adapted until the error is within the specified margin. The author proved that by perceptron convergence procedure, two data groups can be classified correctly.

Werbos (1974), suggested backpropagation algorithm in his Phd thesis. However it was 1986 when backpropagation algorithm using hidden layers was developed by Rumelhart, Hinton and Williams. (Rummelhart, et. al., 1986).

In our study multilayer feedforward neural networks with backpropagation algorithm is used. The network has single input layer and a single output layer, with 1 hidden layers. A fully connected neural network with two hidden layers is presented in Figure 2.19 in order to visualize a neural network architecture.

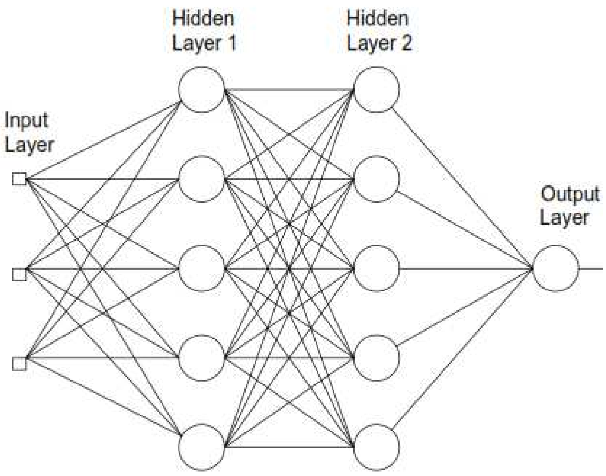


Figure 2.19. Neural Network with two hidden layers

The generalized weight update rule for a multilayer networks is;

$$w_{j(L-1)h_L}(t+1) = w_{j(L-1)h_L}(t) + \eta \delta_{h_L}^{k} x_{j(L-1)}$$

Eqn. 2.7

Where L denotes hidden layers, h represents the element number and j represents the input node.

A database containing input and target data are fed into the ANN architecture. First, output of the system is computed using randomly attended weights. After obtaining the first outputs with those weights, they are adjusted according to the residuals calculated based on the desired outcomes. This loop, which is called iterations or epochs, is continued until a stopping criterion is satisfied.

The ANN architecture consists of an input layer, hidden layer(s) and an output layer. Certain number of neurons are present in the hidden layers which leads to the division of a complex problem into simpler tasks. Neurons are simple processors which consists of functions. Each input is fed into each neuron in hidden layers and the weights are computed using a transfer functions present in the neurons. The computed weights are combined in the last stage of a network, output layer which also consists of a function, used when calculating the desired outputs.

The data for the training of ANNs is divided into three sections in order to determine the best fitting model for the problem. These are training data, cross validation data and testing data. Training data are the input-target pairs used in the training process and validation data are used to evaluate the systems fitness. Both training and validation data's target values are compulsory. The trained system is first tested with the cross validation data. Testing data on the other hand is reserved and is not used in the training. These data are presented to system for the first time after the network is completely created. The aim of testing data is to evaluate the performance of the network when a completely stranger input set is presented to the system.

One of the main problems encountered while conducting neural network analyses is memorization. When a network is trained more than necessary, the system tends to memorize the input - target data pairs, leading to very high performance values when data used in training is presented. However when validation data is delivered, the performance of neural network drops significantly. Performance graphics of training and cross validation data are observed in order to avoid memorization. Stopping criteria or architecture of the network can be modified in order to avoid memorization.

CHAPTER 3

EXPERIMENTAL STUDIES AND SWELLING DATABASE

3.1.General

The main idea of this thesis is to provide a robust and accurate model for the assessment of swelling potential of clayey soils. With this idea in mind, the results of methylene blue (MB) tests and swell percent (SP) tests are used to form a database, where the soil characteristics will be the inputs and MB and SP are desired parameters to be estimated. Using this database, relationship among those parameters are investigated to develop models to understand the swelling behavior of clayey soils. In order for these models to be as general, i.e., universal, as possible, the database should be enriched by collecting data from various resources in the literature from all over the world.

This chapter mainly explains the experimental study held on within the scope of our study, details of the compiled database such as how the data are compiled, its limitations and our contributions. First our own laboratory experiments, which are MB and SP tests on the samples collected from Turkey, are given. The statistical characteristics of the data obtained from laboratory results are also given in this chapter.

3.2.Laboratory Work

The laboratory experiments are conducted on 32 samples, which are obtained from different regions in Turkey. In the laboratory, Atterberg limit tests, specific gravity test, sieve analysis, hydrometer test and MB test are performed on all samples. However, swell percent and swell pressure tests are performed only on 20 of these 32 samples because of the lack of amount of soil samples as they are collected from soil mechanics laboratories and have already been used for other purposes. The maximum dry density and the optimum water content of these samples are determined in the lab through Harvard Miniature Compaction test. The Atterberg limits, sieve size - hydrometer analysis, and specific gravity tests are conducted according to ASTM D4318, ASTM D422 and ASTM D854 accordingly. Test procedures of methylene blue test and oedometer test are presented in the following sections.

3.2.1. Methylene Blue (MB) test

Methylene blue (MB) test is conducted using AFNOR 80181 P18-592. The MB test setup is given in Figure 3.1.

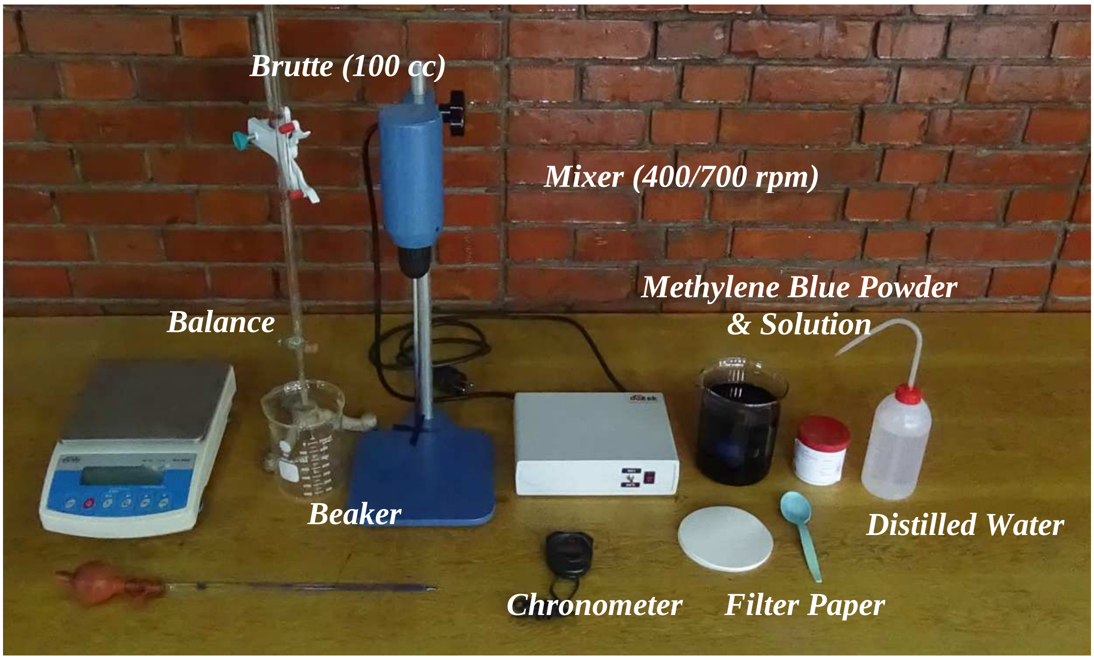


Figure 3.1. Methylene Blue Test Set-up

3.2.1.1. Test Procedure

Methylene blue used in this test procedure is a chemical that has the following formula, $C_{16}H_{18}N_3SCl$ (Çokça, 1991) (Figure 3.2). When methylene blue is dissolved in the water, methlyene blue cations and chloride anions are formed. Methylene blue cations replace with cations existing in the clay particles. When clay particles reach their cation exchange capacity indicates that the existing amount of methylene blue cations is reacted with clay particles.

Methylene blue dye is prepared using 10 g ± 0.1 g of methylene blue powder and 1 lt. of distilled water. Distilled water and methylene blue are very well mixed and the mixture is contained in a beaker.

Soil sample is oven dried and sieved through No.40 sieve. A total of 7.5 g of soil sample is placed in a beaker and 50cc of distilled water is added. The suspension is mixed at 700 rpm for 5 minutes.

Soil suspension is continued to be mixed at 400 rpm. With 1 min intervals 5 cc of methylene blue dye is dropped into clay solution. After mixing the suspension for 1 minute, a drop of the solution is dropped on the filter paper and occurrence of light blue ring is checked. Any excess methylene blue cations (simply obtained by continuing to add methylene blue dye) remain in the solution creating the light blue ring around the solution dropped on the filter paper (Çokça,1991).

When light blue ring is detected, the methylene blue dye addition is stopped and the solution is mixed for 5 minutes. At the end of each minute a drop of the solution is dropped on the filter paper and the existence of light blue ring is checked. If the ring disappears, then 2 cc of methylene blue dye is continued to be added to the solution with 1 min intervals (Figure 3.2).

This whole procedure is conducted for each clay specimen, several times to obtain better results. In order to get more accurate results the clay samples and the mixtures are prepared carefully and each step of the procedure is performed meticulously.

The procedure of MB test is outlines in Figure 3.3.

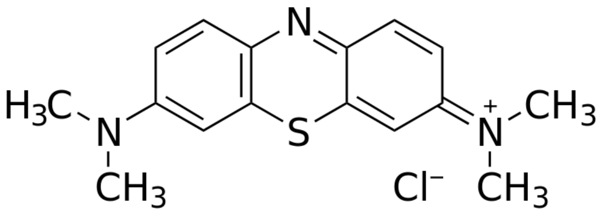


Figure 3.2. Methylene Blue Formula

The methylene blue value is calculated using equation 2.5 for 100 g of the soil.

$$\boldsymbol{MBV(g/100g)} = \frac{\boldsymbol{V_{CC}}}{\boldsymbol{f'}} \qquad \text{Eqn. 3.1}$$

Where,

$V_{CC} = Volume\ of\ methylene\ blue\ solution\ injected\ to\ the\ soil\ solution\ (mL)$

$f' = Dry\ weight\ of\ the\ specimen\ (g)$

Example MB test performed on four different soil samples are given in Figure 3.4 to Figure 3.7. In these pictures, the dark blue drop and the light blue ring around the dark blue core can be observed clearly.

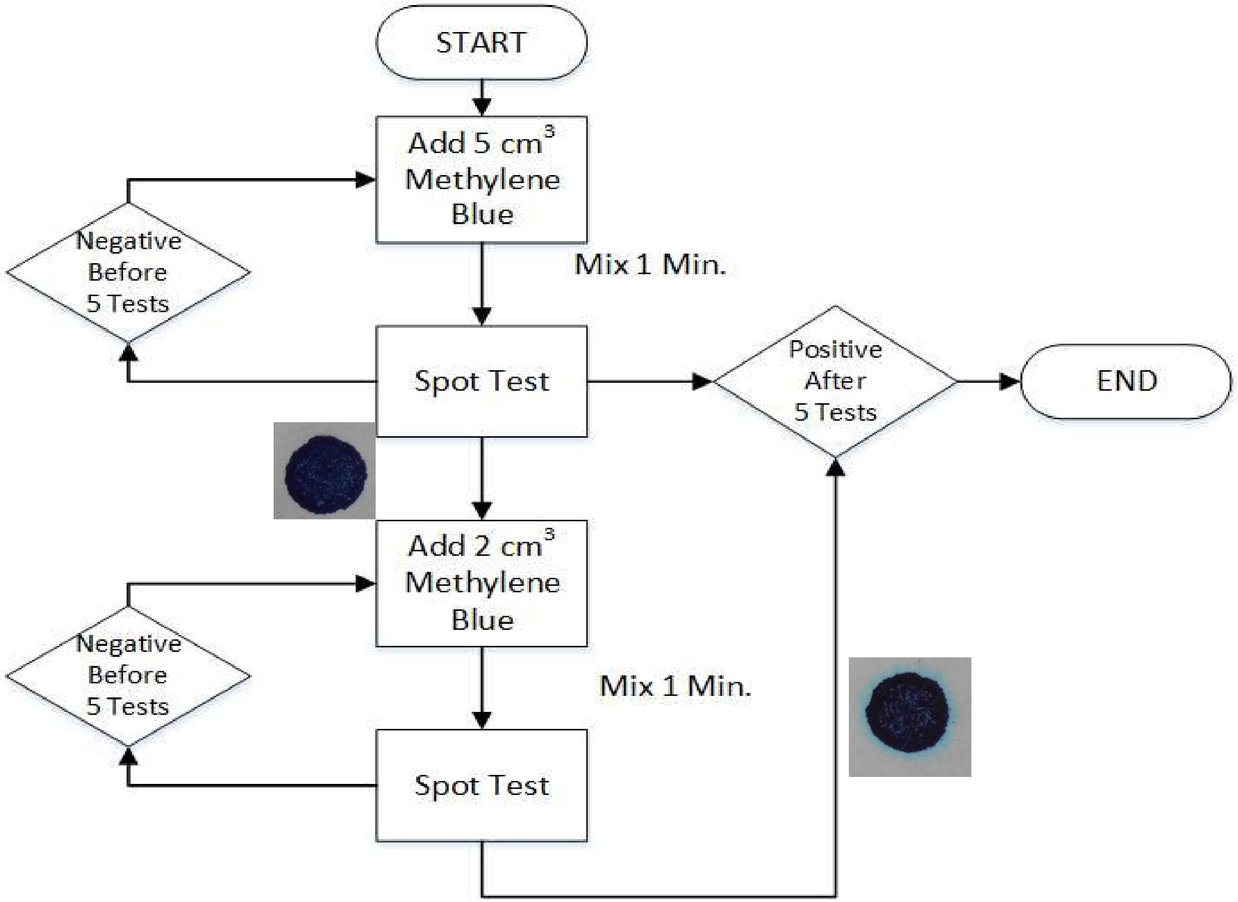


Figure 3.3. General Overview of Methylene Blue Test Procedure

Figure 3.4. Methylene Blue Test Result (Sample No:22)

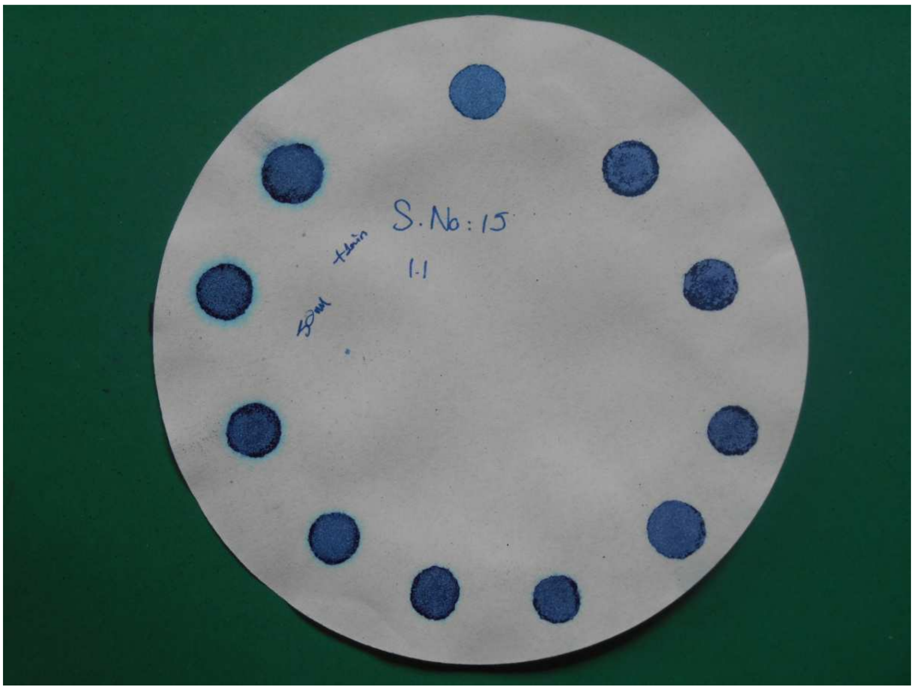


Figure 3.5. Methylene Blue Test Result (Sample No:15-1)

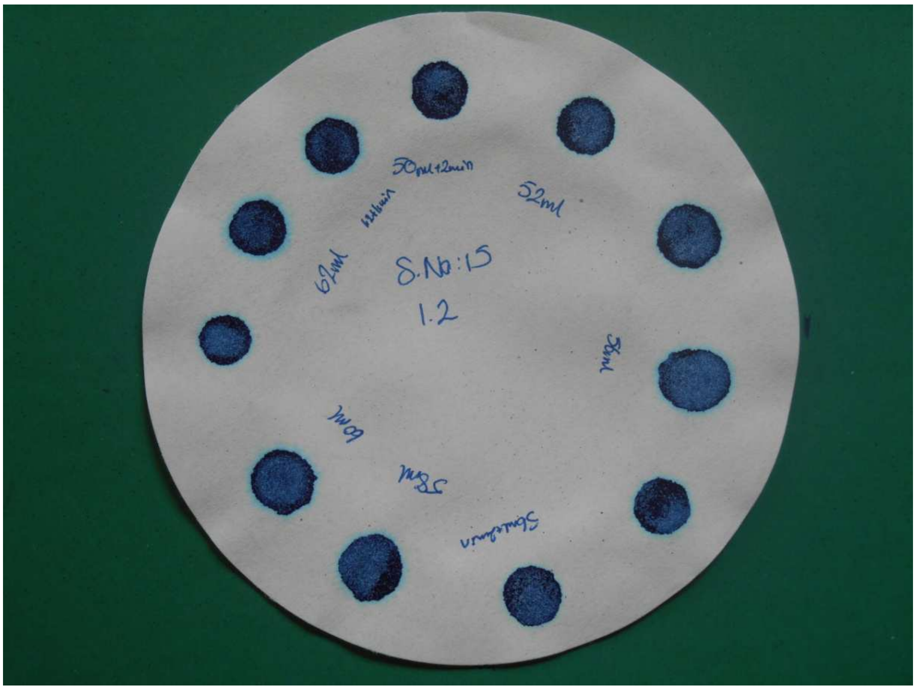


Figure 3.6. Methylene Blue Test Result (Sample No:15-2)

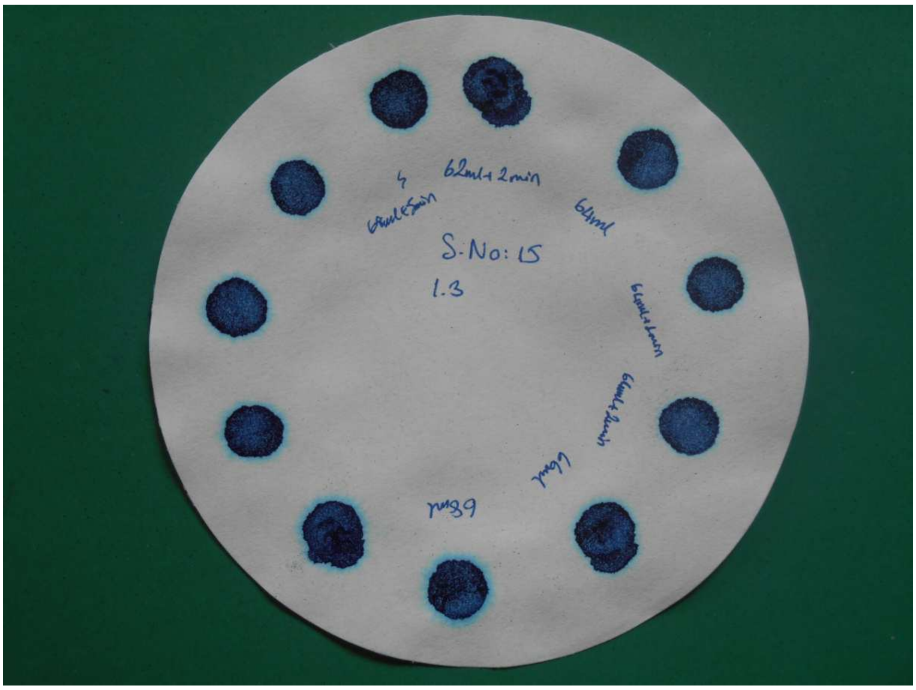


Figure 3.7. Methylene Blue Test Result (Sample No:15-3)

3.2.2. Oedometer Test

In this study oedometer test is used for determining swelling characteristics of clayey soils, which are swelling percent and swelling pressure. The test procedures for determining these parameters are named respectively as free swell test and constant volume (swell pressure) test. There are a quite variety of test procedures, however ASTM D4546-14 suggests 3 test methods for swell characteristics determinations. The specimen height and specimen diameter are specified at ASTM D4546-14.

In our study, for the swell percent test, ASTM D4546-14 Test method A is applied on the samples. During the oedometer test, the samples having optimum water content and maximum dry density are placed to rings. 7 kPa of load is placed on the samples and after the samples are inundated, the free swell is observed. In addition, for the determination of swell pressure, constant volume swell test is used. The samples having maximum dry density and optimum water content are placed in oedometer test apparatus and after the samples are inundated, swelling is avoided by placing the additional loads.

The test set up is given at the Figures 2.7 and 2.8 below. The apparatus used in the test are listed below.

3.2.2.1. Apparatus:

-Oedometer Cell

-Specimen Ring

-Porous Disks

-Deformation Indicator

(a)

(b)

Figure 3.8. (a) Oedometer Test Apparatus, (b) Oedometer Test Set-up

3.2.2.2. Test Procedures

a) Test Method A

Test Method A uses reconstituted samples simulating field conditions or compacted fills. This test method can determine free swell, swell pressure and swell under a certain load. 1 kPa load is applied on the sample placed into oedometer apparatus, for half an hour. The test is continued by applying an initial vertical stress on the sample for another half an hour. After that the load is removed and applied again for a second consecutive half-hours. Then, the next step is to flood the sample with water and to take at 0.1, 0.2, 0.5, 1.0, 2.0, 4.0, 8.0, 15.0, 30.0 minutes and 1, 2, 4, 8, 24, 48, 72 hours. After the primary swell is accomplished the sample is loaded again until the initial void ratio or height is obtained (Swell Pressure).

b) Test Method B

Test Method B uses intact samples taken from natural deposit or an existing compacted fill. This test method can determine free swell, swell pressure and swell under a certain load. The sample is placed into consolidometer apparatus and loaded with the weight chosen equal to the vertical in-situ stress corresponding to the sampling depth. The sample having the natural moisture content is then inundated. The procedure continues same as Test method A.

c) Test Method C

Either intact sample or reconstituted samples can be used for Test Method C. Intact samples can be taken from natural deposit or an existing compacted fill. This test method can determine swell pressure. Initially, after the sample is placed, 1 kPa is applied for half an hour. Then the load is increased to obtain in-situ loading conditions. The specimen is then inundated and vertical stress is adjusted to maintain the initial volume.

3.2.3. Laboratory Test Results

The results of all laboratory test results are provided in Table 3.1. This table also provides the locations of soils where the samples are collected. (Particle size distributions and Harvard miniature compaction curves are presented in Appendix A.) Figure 3.1 illustrates the distribution of these samples on Plasticity Index vs. Liquid Limit chart. Similarly, Figure 3.2 shows our study's data in clay content vs. methylene blue Value chart. Lastly, Figure 3.3 displays the same data on Activity vs. Percent Clay Sizes chart.

Methylene blue test is applied twice on each samples and the average of the results are presented.

Table 3.1. The Results of Laboratory Experiments on Samples from Turkey

S. No:	Location	G_s	No 4 Retaining	No200 Passing	LL	PL	PI	Clay	Silt	Sand	A	MBV (g/100g)	Swell Pressure (kPa)	Swell Percent (%)	ρ_{dry} (g/cm^3)	w_{opt}
1	Unknown-46095	2,70	18,9	57,1	88	30	59	29	45	22	2,07	11,8	269,2	20,3	1,5	27,5
2	Unknown-13-B	2,71	0,0	77,5	65	28	37	46	39	15	0,81	10,0	106,1	6,4	1,6	28,5
3	Unknown-NN-1	2,59	2,2	77,3	67	23	44	54	24	20	0,81	12,7	-	-	-	-
4	Unknown-NN-2	2,62	0,0	91,6	75	24	51	60	25	15	0,85	14,5	-	-	-	-
5	Adana-41 A	2,70	6,3	64,2	35	16	19	17	33	30	1,14	4,1	10,1	5,8	1,75	18,5
6	Adana-N-37	2,71	1,1	83,0	52	22	30	33	33	29	0,91	7,6	20,3	4,5	1,72	20,5
7	Adana-N-38A	2,70	0,4	92,6	34	16	19	22	38	30	0,86	5	11,5	1,2	1,71	17
8	Adana-N-51	2,70	0,0	95,5	45	21	24	31	37	28	0,76	6,3	12,6	1,9	1,67	21,6
9	Adana-N-90	2,69	0,0	97,9	54	25	29	36	29	32	0,81	6,8	13,9	2,5	1,51	25,3
10	Ankara-Alacaatlı	2,67	1,2	32,7	69	27	42	50	16	27	0,85	9,9	55,6	4,8	1,47	25,3
11	Ankara-Batıkent Kreş	2,64	0,4	55,9	60	27	33	49	40	11	0,68	11,1	-	-	-	-
12	Ankara-Elmadağ SK-2	2,72	0,2	43,6	46	21	25	36	27	35	0,70	6,3	-	-	-	-
13	Ankara-Etimesgut 46137-2	2,68	1,2	87,7	97	30	67	48	40	10	1,40	11,9	-	-	-	-
14	Ankara-Etimesgut- SK-2	2,72	3,5	67,6	63	35	28	10	43	33	2,95	5,9	-	-	-	-
15	Ankara-Etimesgut-1	2,71	0,0	89,7	64	26	38	44	35	18	0,87	9,0	-	-	-	-
16	Ankara-Polatlı	2,59	1,7	88,9	48	22	26	35	46	19	0,74	5,7	21,9	4,5	1,65	21
17	Ankara -Sincan 190-11	2,65	0,6	85,8	109	30	79	51	38	11	1,55	12,0	-	-	-	-
18	Ankara-Sincan 699-11	2,66	0,0	92,1	106	29	77	45	43	12	1,72	12,8	-	-	-	-

Table 3.1. The Results of Laboratory Experiments on Samples from Turkey (Cont'd)

S. No:	Location	G_s	No 4 Retainin	No200 Passing	LL	PL	PI	Clay	Silt	Sand	A	MBV (g/100g)	Swell Pressure	Swell Percent	ρ_{dry} (g/cm^3)	w_{opt}
19	Ankara-Temelli	2,62	1,1	72,2	62	25	36	34	27	35	1,05	10,4	35,4	5,8	1,50	24,5
20	Bayramhacı Sulaması, Kocaeli aç-35	2,61	1,1	60,8	41	19	21	17	15	61	1,26	6,1	12,7	1,7	1,83	17
21	Bayramhacı Sulaması, Kocaeli aç-6	2,54	1,9	89,8	46	27	18	11	31	53	1,68	4,6	11,49	1,05	1,41	28,2
22	Bilecik Bozeyik	2,74	0,4	96,2	29	15	14	25	36	34	0,56	2,4	18,95	6,05	1,89	16
23	Bursa-Orhangazi	2,65	7,9	64,6	51	22	29	37	43	19	0,79	8,9	69,73	5,58	1,67	22,5
24	Çorum-Karahacip	2,60	0,1	96,7	55	26	29	40	41	16	0,73	7,7	-	-	-	-
25	Çubuk SK-1	2,70	0,2	79,2	73	31	42	49	44	5	0,85	11,9	-	-	-	-
26	Eskişehir-Bakırköy Göleti-ESK	2,65	0,0	87,9	64	31	33	42	44	11	0,79	9,7	60,63	3,37	1,47	29
27	İnebeyli sulaması aç-101	2,73	0,4	92,3	50	21	29	44	42	11	0,66	9,5	30,92	3,73	1,666	19,2
28	İnebeyli sulaması aç-106	2,74	0,0	94,6	47	23	24	31	55	13	0,76	9,1	17,68	2,74	1,64	20,5
29	Karabük Mer. SK-1	2,69	0,0	99,4	65	31	34	50	48	2	0,68	8,1				
30	Kırıkkale-Keskin	2,61	2,0	89,4	48	21	27	27	28	42	0,99	9,2	24,00	2,76	1,61	22
31	Yumrudere Sulaması, Kırklareli ı aç-321	2,83	0,0	91,6	38	26	12	7	33	47	1,80	3,1	7,58	1,17	1,67	21
32	Yumrudere-Sarpdere Kırklareli	2,75	1,4	74,8	49	26	23	29	36	29	0,80	4,8	28,40	3,55	1,7	23,2

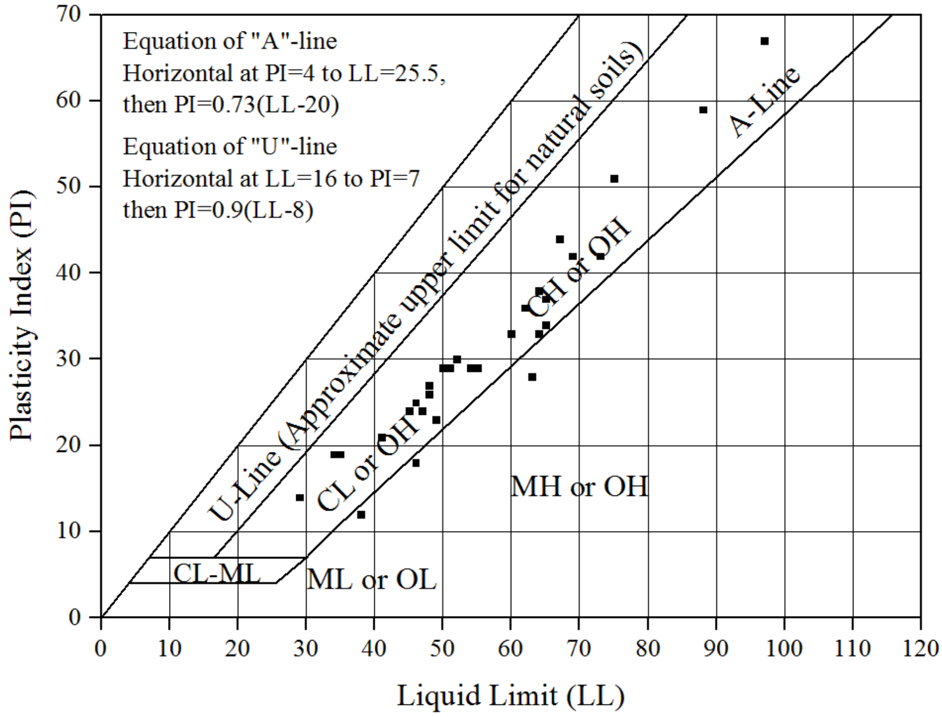


Figure 3.9. Casagrande chart

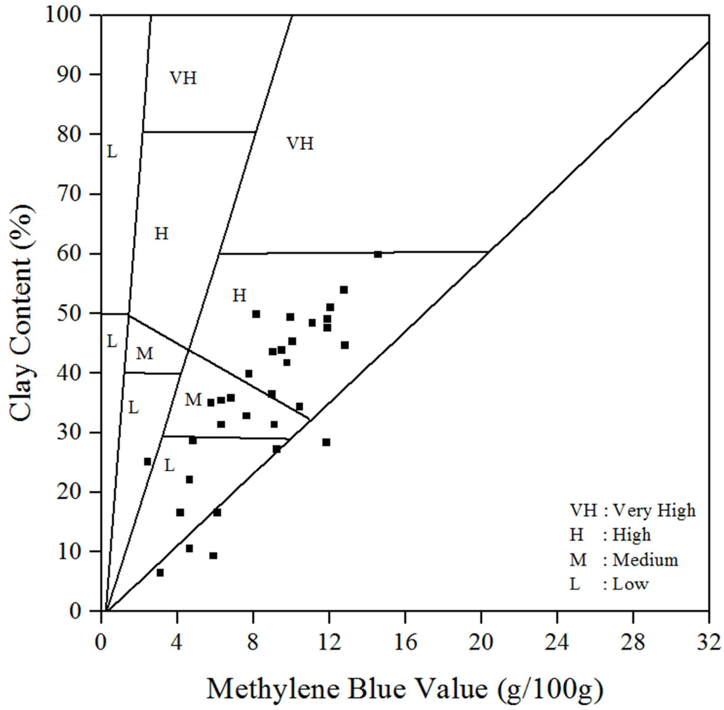


Figure 3.10. Swelling Potential Classification Chart (Çokça,1991)

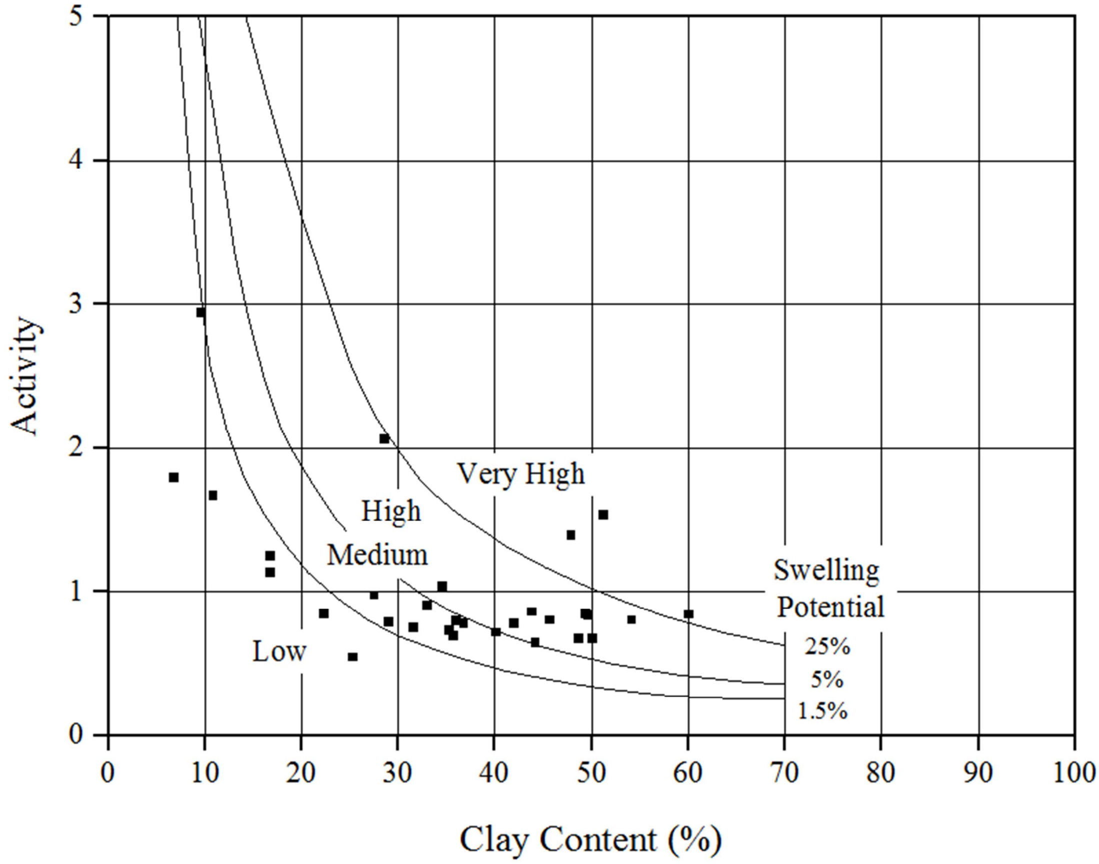


Figure 3.11. Swelling Potential classification chart (Seed et. al., 1963)

3.3. Swelling Database for Clayey Soils

In order for the mathematical or machine learning based regression techniques to work, the data obtained from the laboratory work should cover various ranges for the variables included in the database. Therefore, our laboratory data is extended using the data available in the literature.

3.3.1. Data Source

The main source of the literature data is the PhD thesis of Çokça (1991). Çokça (1991) created a data pool by combining his own data and using the laboratory test results obtained from the studies of Lautrin (1989), Magnan (1989). In addition, the data available in the works of Erguler (2001) and Türköz (2007) are also used when forming the swelling database for clayey soils. Although other studies provide data containing methylene blue value, either due to the difference in methylene blue test procedure or unavailability of datasets they are not included in our study.

The study of Ergüler (2001) is based on determination of swelling potential of clayey soils using MB test on Ankara clay. He obtained statistical relations using regression analyses. He determined Atterberg limits, water content, natural unit weight, specific gravity, dry unit weight and clay contents of the samples. In Erguler's study, free swell index, modified free swell index, oedometer tests (swell percent and swell pressure) and MB test were conducted on the samples. Optimum water contents of the samples are not determined for the samples, however w_{max24} and w_{max72} values, the maximum water contents that samples can get after 24 hours and 72 hours, are determined.

Türköz (2007) also studied about determination of swelling potential of clays and conducted some experiments on Harran clay. He used direct methods to determine swelling potential of clayey soils as well as MB test and compared the results of these tests. He determined Atterberg limits, water content, unit weight, specific gravity, dry unit weight and clay contents of the samples. In addition, MB test, swell percent and swell pressure tests were conducted on the samples. In Türköz's study, swell percent is found using expansion index test and swell pressure is obtained through conducting PVC meter test. In Ergüler (2001), Türköz (2007), Çokça (1991) and our study MB tests are conducted according to AFNOR.

3.3.2. Statistical Data Characteristics

In the swelling database for clayey soils, there are total of 332 samples obtained through literature survey and 32 samples from our laboratory experiments. For all samples, the results of Atterberg limits, sieve analysis, hydrometer analysis and specific gravity tests are obtained. However, due to limitations of literature and availability of sample quantities, some lab experiments are not complete. Therefore the total data pool is divided into 4 different data sets and for each data set linear regression and machine learning analyses are performed.

Data Set I: This data set consists of the results of 32 samples obtained from different regions in Turkey. Atterberg limit tests, specific gravity test, sieve analysis, hydrometer test, MB Test are performed on all of the samples. Swell percent (ASTM D4546-14, Method A) and swell pressure tests (Constant Volume Test) are performed on only 20 of the samples. Data set I is used for both methylene blue value (MBV) prediction and swell percent prediction analysis.

Data Set II: It consists of total of 75 data samples, combination of data from the literature, Ergüler (2001) and Türköz (2007), as well the ones from our own results. This data set is used for both MBV prediction and Swell percent prediction analysis. Both of the literature studies conducted basic tests and methlyene blue tests similarly. However swell percents of the samples are determined slightly differently such that, our study and Ergüler (2001) used oedometer test (Method A and Method B) for swell percent determination while Türköz (2007) used expansion index.

The differences between the test methodologies are listed as;

- ✓ In our laboratory experiments, samples having maximum dry density and optimum water content are used for oedometer test.
- ✓ Ergüler (2001) tested both undisturbed and disturbed samples. Disturbed sample results are used at this data set. Ergüler (2001) determined in-situ density and water content from undisturbed samples and the disturbed samples are placed to oedometer test apparatus having natural density and water content.
- ✓ Türköz (2007) performed Expansion Index test to obtain swell percent. The height of the mold is 10.2 cm. 50% water content of the samples are used at expansion index test.

Data Set III: This data set consists of 125 data in total combined from this thesis, Çokça (1991), Ergüler (2001) and Türköz (2007). The data set is only used for MBV prediction.

Data Set IV: This data set consists of 365 data in total combined from our study, Lautrin (1989), Magnan (1989), Çokça (1991), Ergüler (2001) and Türköz (2007). This data set is only used for MBV prediction too. For this data set, the outliers that stand outside the limits are omitted during both linear regression machine learning analysis.

The summary of the contents of the dataset are summarized in Table 3.2.

The statistical measures for individual data from different resources are given in Table 3.3 and Table 3.4. In addition, all the data are plotted in Figure 3.12 to Figure 3.16.

Table 3.2. Summary of datasets

	Dataset I		Dataset II		Dataset III	Dataset IV
Predicted Parameters	MBV	Swell Percent	MBV	Swell Percent	MBV	MBV
Number of Data	32	20	73	73	125	343
Studies Composed of	Our Study		Ergüler (2001) Türköz (2007) Our Study		Çokça (1991) Ergüler (2001) Türköz (2007) Our Study	Magnan (1989) Lautrin (1989) Çokça (1991) Ergüler (2001) Türköz (2007) Our Study

Table 3.3. Data Characteristics Data Set I & Data Set II

	LL	PL	PI	CC	A	MBV	ρ_{dry*}	w_{opt*}	SP*	Swell Pressure*
DATA SET I										
Mean	59.2	24.9	34.3	35.6	1.1	8.4	41.9	4.5	1.6	22.4
Median	54.5	25.5	29.0	36.0	0.9	9.0	21.1	3.6	1.7	21.8
Variance	383.1	23.3	276.1	186.4	0.3	9.7	3487.5	16.9	0.0	15.7
Std Dev.	19.6	4.8	16.6	13.7	0.5	3.1	59.1	4.1	0.1	4.0
Std. Error	3.5	0.9	2.9	2.4	0.1	0.6	13.2	0.9	0.0	0.9
Skewness	1.0	-0.2	1.4	-0.4	2.1	-0.1	3.3	3.2	0.1	0.2
Kurtosis	0.9	-0.3	1.6	-0.5	4.9	-0.8	12.5	12.4	-0.3	-0.9
Range	80.0	20.0	67.0	53.0	2.4	12.1	261.6	19.2	0.5	13.0
Max	109.0	35.0	79.0	60.0	3.0	14.5	269.2	20.3	1.9	29.0
Min	29.0	15.0	12.0	7.0	0.6	2.4	7.6	1.1	1.4	16.0
DATA SET II										
Mean	58.8	28.7	30.2	39.2	0.9	7.4	1.6	37.5	40.8	6.7
Median	60.0	28.0	30.0	41.0	0.8	7.2	1.6	25.0	30.3	3.6
Variance	151.3	31.2	74.7	200.1	0.1	4.6	0.0	666.6	152	64.5
Std Dev.	12.3	5.6	8.6	14.1	0.3	2.1	0.1	25.8	39.1	8.0
Std. Error	1.4	0.7	1.0	1.7	0.0	0.3	0.0	3.0	4.6	0.9
Skewness	0.5	0.2	1.2	-0.3	2.0	0.5	1.0	1.7	3.5	2.6
Kurtosis	1.8	1.3	4.8	-0.1	4.0	1.0	0.9	1.9	16.5	7.0
Range	73.8	30.0	55.2	68.3	1.6	12.4	0.5	109.0	267.	42.2
Max	103.0	45.0	67.0	75.0	2.1	14.8	1.9	125.0	269.	43.0
Min	29.2	15.0	11.8	6.7	0.5	2.4	1.4	16.0	1.4	0.8

**For 20 samples, 12 missing test results out of 32 Test Results (For Data Set I)*

Table 3.4. Data Characteristics Data Set III & Data Set IV

	LL	PL	PI	CC	A	MBV
DATA SET III						
Mean	60.7	27.0	33.8	41.7	0.9	7.4
Median	60.0	27.0	32.0	43.7	0.8	7.1
Variance	251.4	33.3	185.2	190.1	0.1	6.2
Std Dev.	15.9	5.8	13.6	13.8	0.3	2.5
Std. Error	1.4	0.5	1.2	1.2	0.0	0.2
Skewness	1.7	0.3	2.6	-0.3	3.0	0.6
Kurtosis	6.9	0.6	11.4	0.1	11.8	0.6
Range	117.0	31.0	104.0	68.3	2.5	12.5
Max.	146.0	45.0	116.0	75.0	2.9	14.9
Min.	29.0	14.0	12.0	6.7	0.5	2.4
Sum	7592,00	3377,00	4223,00	5207,48	108,02	927,68
DATA SET IV						
Mean	60.0	31.8	52.4	49.0	0.7	5.2
Median	59.0	31.0	48.1	45.0	0.7	4.9
Variance	391.5	188.9	394.3	371.7	0.1	10.2
Std Dev.	19.8	13.7	19.9	19.3	0.3	3.2
Std. Error	1.1	0.7	1.1	1.0	0.0	0.2
Skewness	0.6	0.6	0.4	0.5	2.7	0.5
Kurtosis	-0.1	0.2	-0.7	-0.3	12.8	-0.8
Range	90.0	73.2	85.8	87.3	2.8	13.0
Max.	114.0	79.2	94.0	94.0	3.0	13.2
Min.	24.0	6.0	8.2	6.7	0.2	0.2
Sum	20530,01	10863,43	17915,89	16768,78	235,49	1770,30

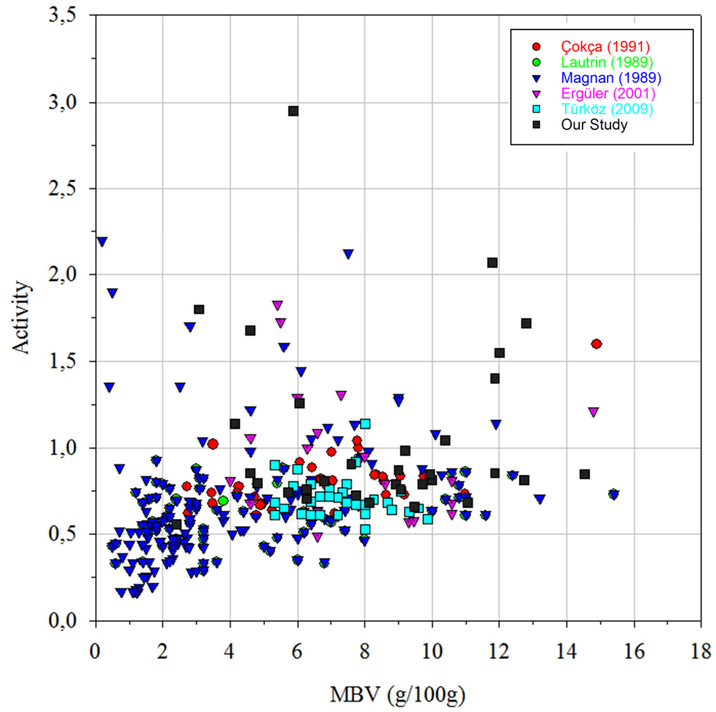


Figure 3.12. Methylene Blue Values vs Activity

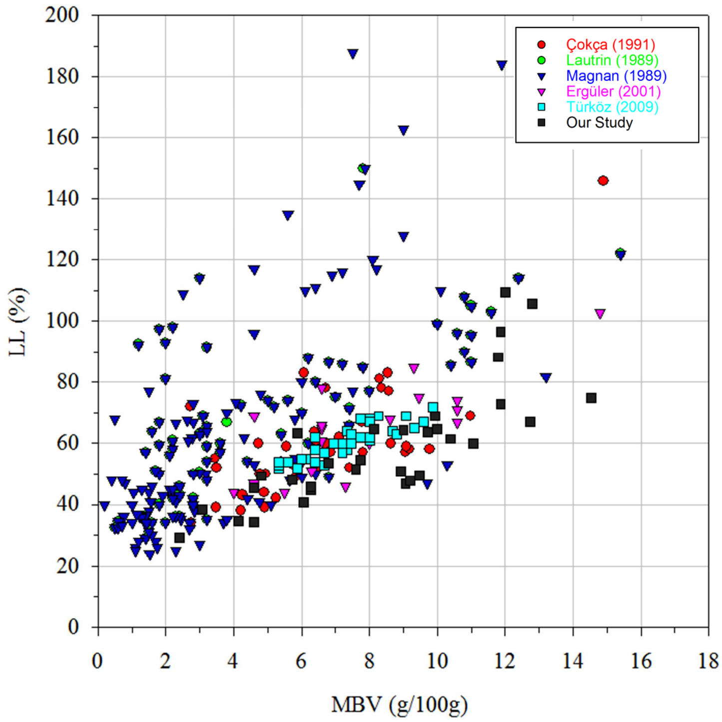


Figure 3.13. Methylene Blue Values vs Liquid Limit

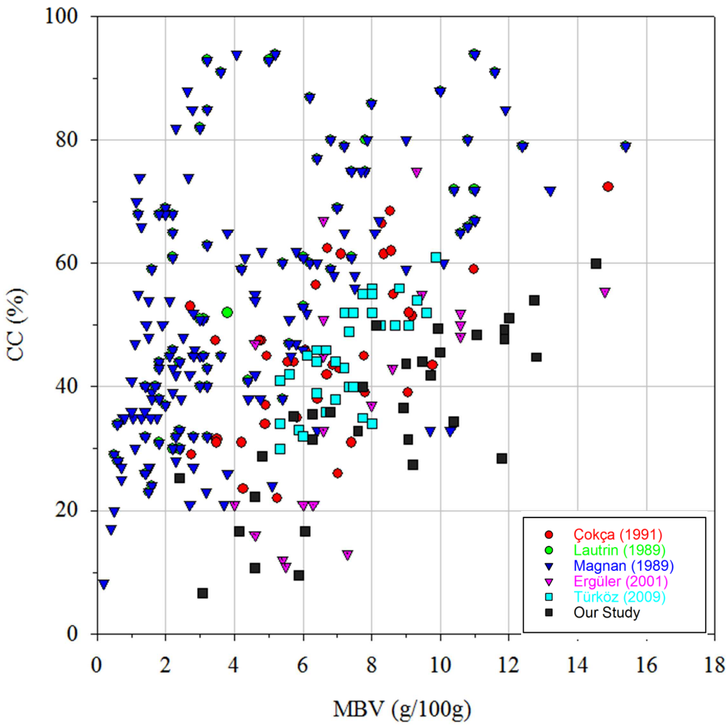


Figure 3.14. Methylene Blue Values vs Clay Content

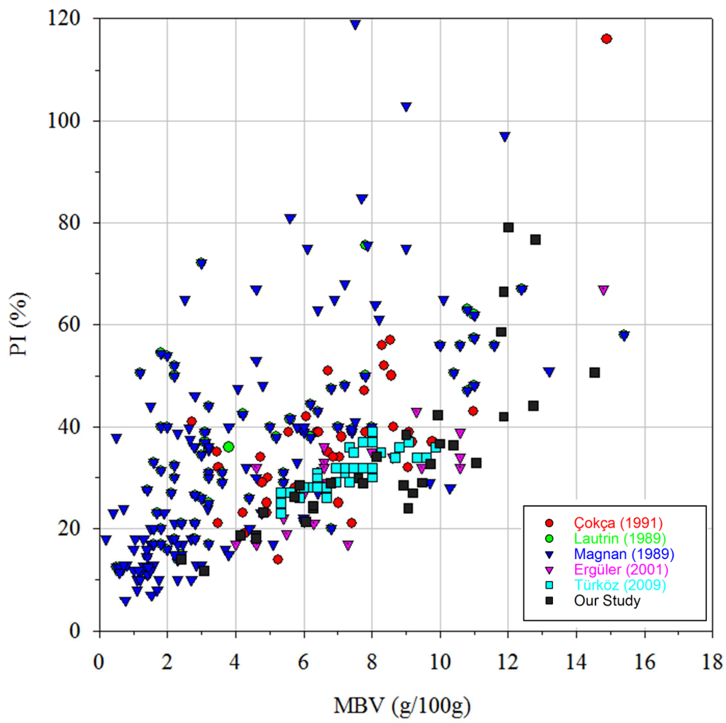


Figure 3.15. Methylene Blue Values vs Plasticity Index

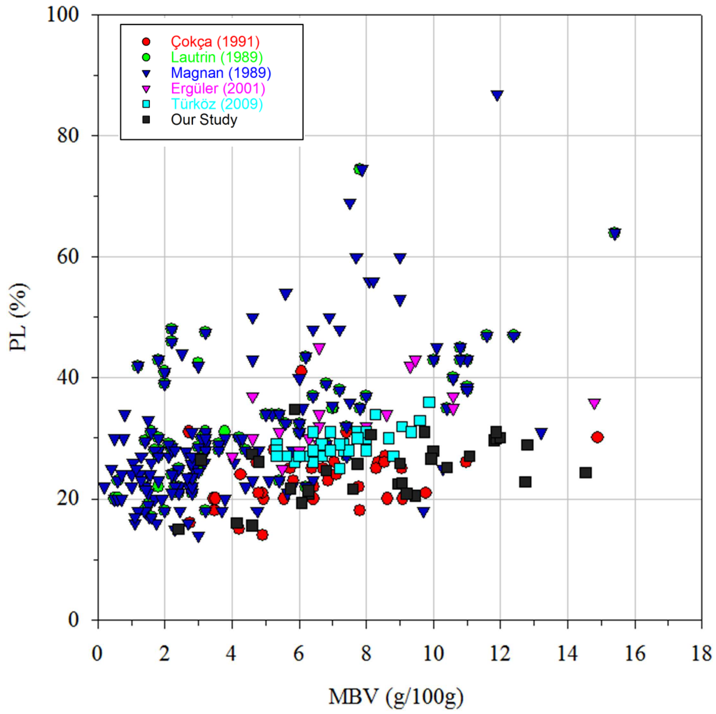


Figure 3.16. Methylene Blue Values vs Plastic Limit

CHAPTER 4

PREDICTIVE MODELS

4.1. General

The main work of this thesis consists of developing models to estimate (i) swell percent and (ii) methylene blue value (MBV) of clays. This chapter consists of the details of models developed using (i) linear regression, (ii) genetic expression programming and (iii) artificial neural networks (ANN).

4.2. Linear Regression Models

In any regression type of analyses, the first thing to attempt is to use linear regression analyses. In this section, the predictive models using multivariate linear regression analyses on the datasets constructed from literature as well as experimental work results are mentioned. Datasets I, II, III, and IV are used at MBV prediction analyses. Also predictive models for swelling percent are constructed for Datasets I and II.

Robust multivariate linear regression analyses are conducted on datasets. The main reason behind using robust multivariate linear regression analyses is to prevent having non-uniform distribution of residuals when the model is constructed, the phenomenon known as heteroscedasticity. Total of 153 linear regression analysis are conducted, where the performance of the analyses are evaluated according to the mean absolute percentage error (MAPE) and the coefficient of determination (R^2). In addition, heteroscedasticity of each model is verified within 85% confidence interval. According to these criteria, the best performing models are presented in Table 4.1. For each model, the statistical measures such as the average Variance Influence Factors (VIFs), the results of student t-tests (F) and each variable confidence interval results are given in Table 4.1. The summary of all linear regression analyses are provided in Appendix B. In addition, equations 4.1 to 4.5 provide the equations for these best predictive models. The performances of these linear predictive models are also given in the form of “Target vs. Output” graphs and “Residuals vs. Output” graphs in Figure 4.1 to Figure 4.10. Lastly, the linear regression results of Data Set IV is given in Figure 4.11 and Figure 4.12. As seen on these plots, this model should be used carefully because it shows a trend for heteroscedasticity. Therefore it can be said that no significant relation can be obtained using Data Set IV.

For the best performing linear predictive models for MBV prediction, the average MAPE values change between 17.4 % and 19.8 %, which is not really good considering the applicability of these models to more comprehensive datasets. In other words, as there are more data to be added, the performance will have the tendency to get worse. Still compared to the performance of predictive models to estimate the swell percent, MBV estimation models work better. The MAPE values of swell percent models change between 44.6% and 57.9%, which shows that either the database is not constructed carefully considering all the variability of laboratory experiments or the linear models cannot be used successfully.

Table 4.1. The Summary of Linear Regression Analyses

Data Set	LR No	Input Data	Output Data	MAPE (%)	R^2	Mean VIF	F
Data Set I	LR1	PI, CC	MBV	17.4	0.80	1.67	58.25
Data Set II	LR2	LL, PL, w_{opt}	MBV	14.0	0.68	2.60	65.06
Data Set III	LR3	PI, CC	MBV	19.8	0.53	1.69	47.06
Data Set I	LR5	A, LL, PL, ρ_{dry}, MBV	Swell Percent	44.6	0.84	4.24	16.38
Data Set II	LR6	LL, ρ_{dry}, w_{opt}, MBV	Swell Percent	57.9	0.79	1.93	24.35

The performance of preliminary models prove to be insufficient for various datasets. It seems that there is a need to better model the relations to understand the swelling behavior using simple laboratory experiments. When the trends in a database cannot be modeled using regression analysis, then, instead of simple statistical approaches, machine learning tools can be used for modeling. In the following sections of this chapter, two well-known machine learning algorithms (i) genetic expression programming (GEP) and (ii) artificial neural networks (ANN) are used for predictive modelling.

Data Set I, MBV Prediction:

$$MBV = 0.096PI + 0.108CC + 1.254 \qquad \text{Eqn. 4.1}$$

Data Set II, MBV Prediction:

$$MBV = 0.200LL - 0.229PL + 0.012w_{opt} + 1.697 \qquad \text{Eqn. 4.2}$$

Data Set III, MBV Prediction:

$$MBV = 0.114PI + 0.023\mathrm{CC} + 2.620 \qquad \text{Eqn. 4.3}$$

Data Set I, Swell Percent Prediction:

$$Swell\ Percentage = 2.771A + 0.488LL - 0.522PL + 12.810\rho_{dry} - 0.559MBV - 27.794 \qquad \text{Eqn. 4.4}$$

Data Set II, Swell Percent Prediction:

$$Swell\ Percentage = 0.108LL + 8.300\rho_{dry} + 0.219w_{opt} + 0.673MBV - 25.875 \qquad \text{Eqn. 4.5}$$

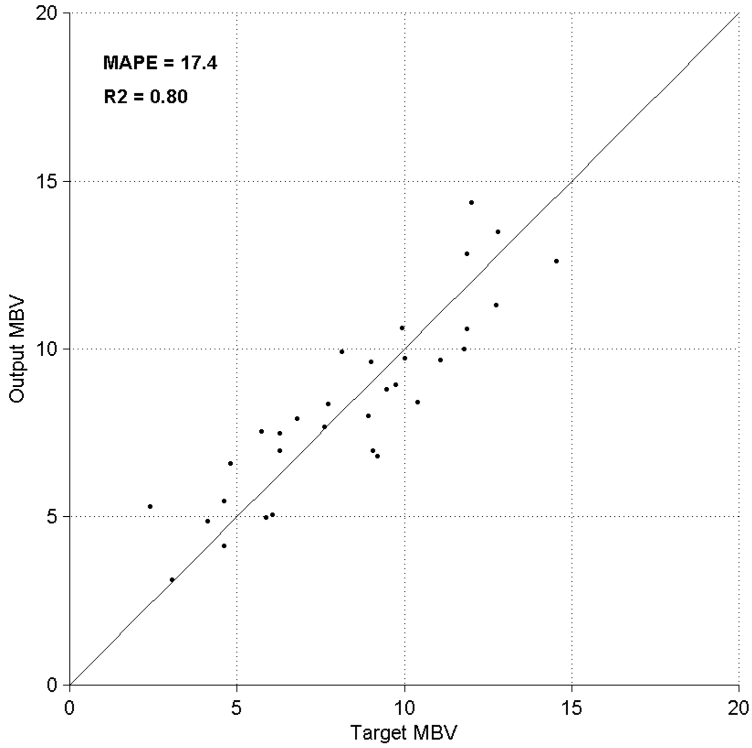


Figure 4.1. Data Set I, MBV Prediction, Linear Regression Graph

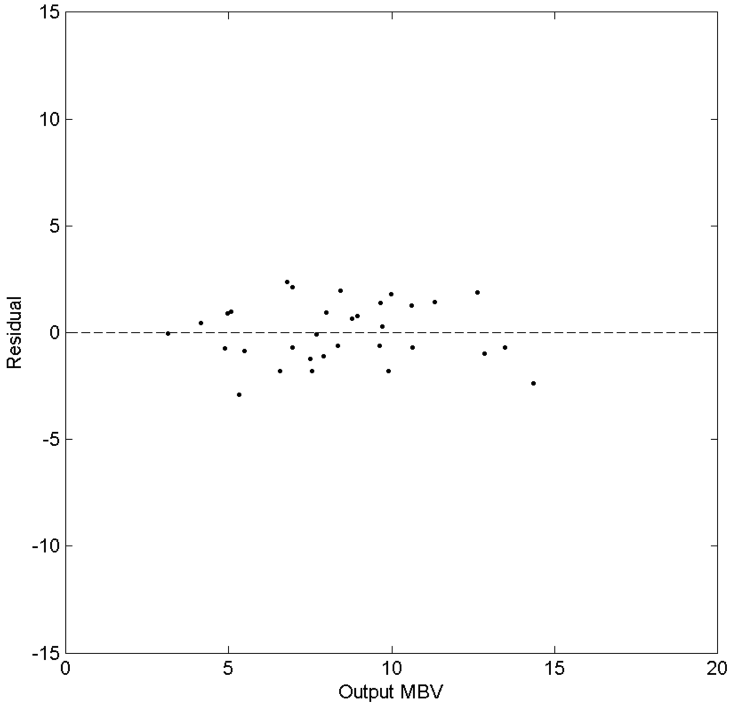


Figure 4.2. Data Set I, MBV Prediction, Residuals vs. Output Graph

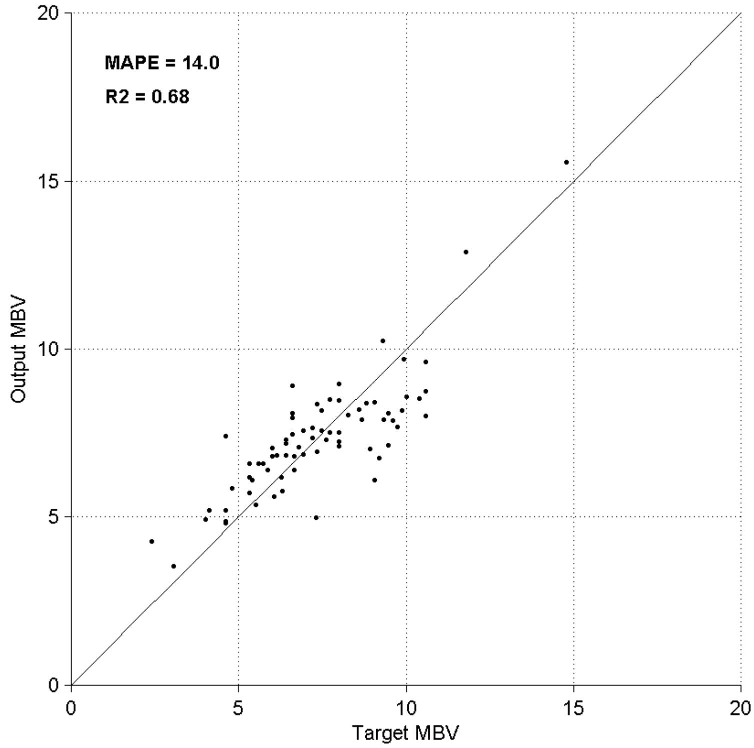


Figure 4.3. Data Set II, MBV Prediction, Linear Regression Graph

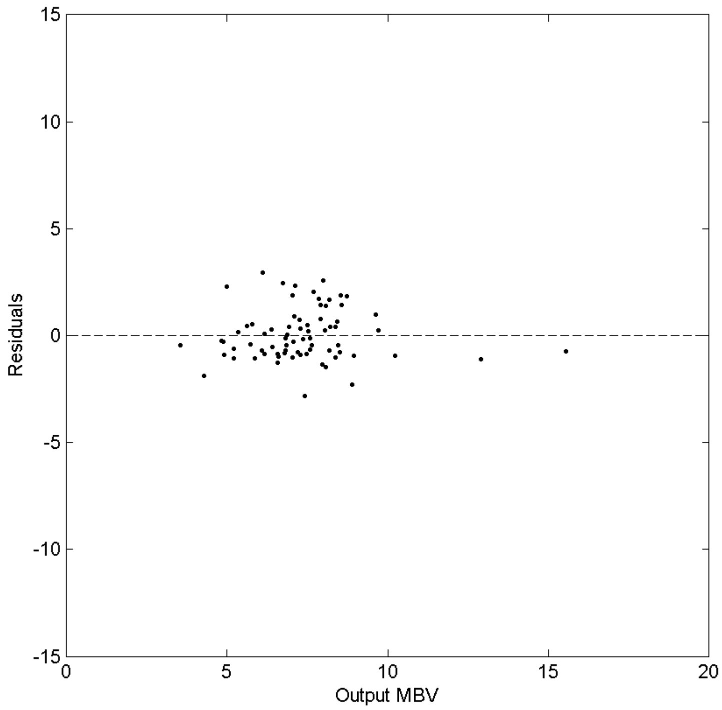


Figure 4.4. Data Set II, MBV Prediction, Residuals vs. Output Graph

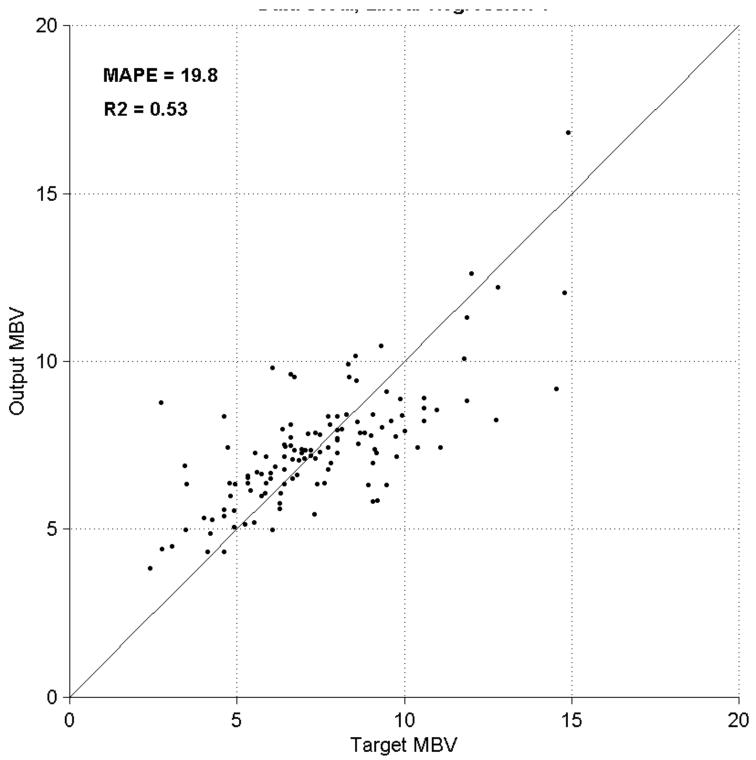


Figure 4.5. Data Set III, MBV Prediction, Linear Regression Graph

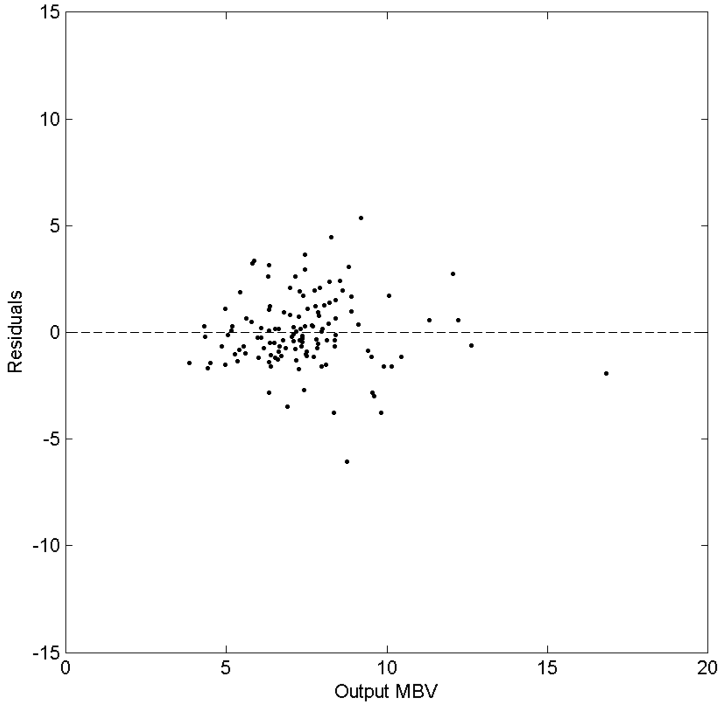


Figure 4.6. Data Set III, MBV Prediction, Residuals vs. Output Graph

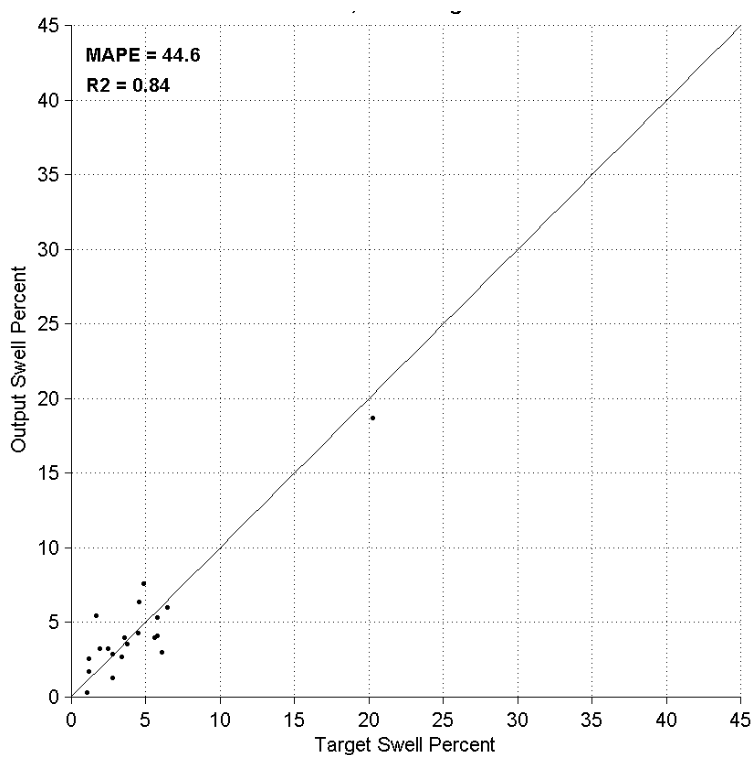


Figure 4.7. Data Set I, Swell Percent Prediction, Linear Regression Graph

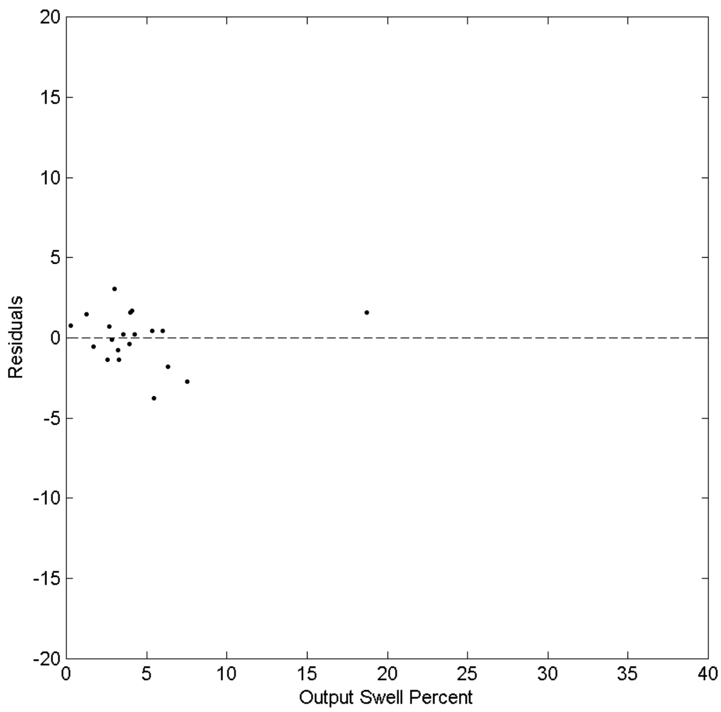


Figure 4.8. Data Set I, Swell Percent Prediction, Residuals vs. Output Graph

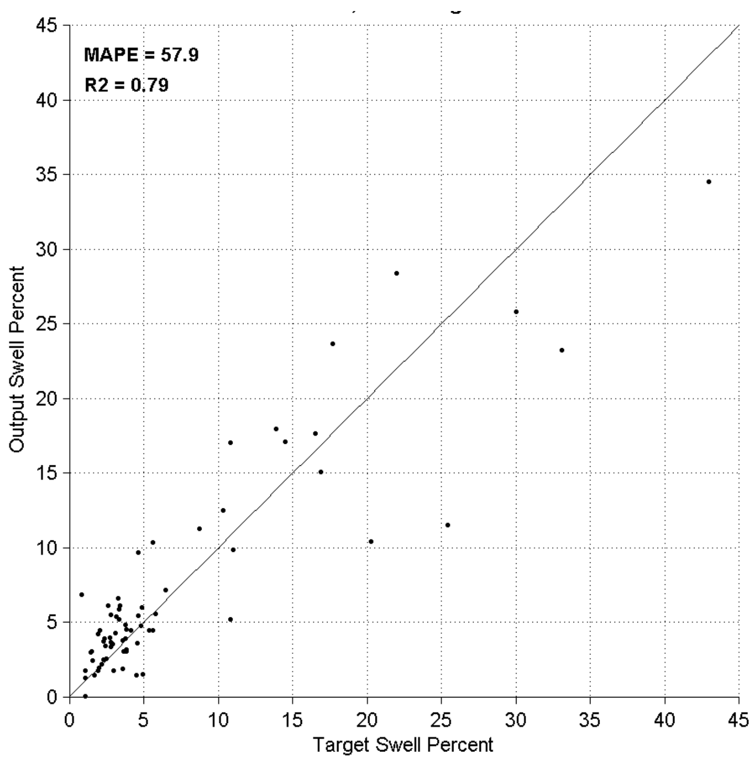


Figure 4.9. Data Set II, Swell Percent Prediction, Linear Regression Graph

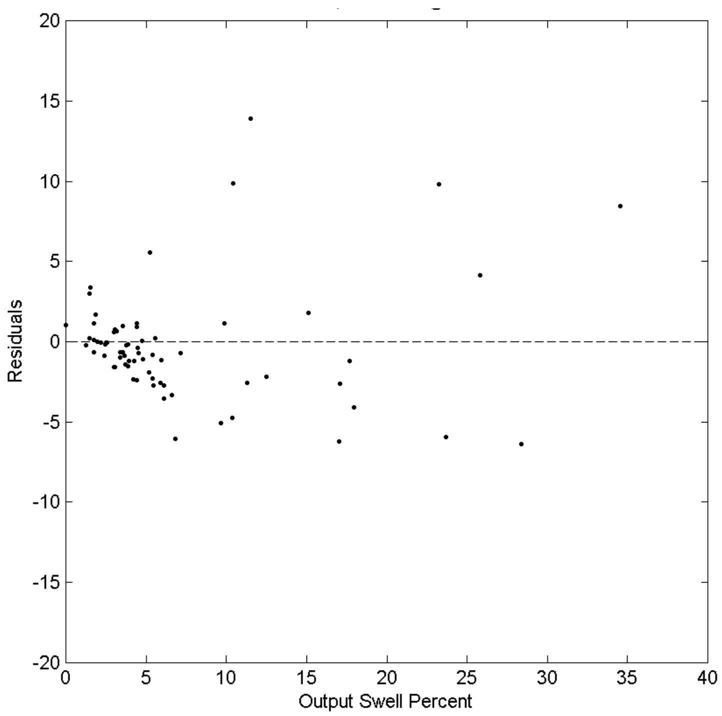


Figure 4.10. Data Set II, Swell Percent Prediction, Residuals vs. Output Graph

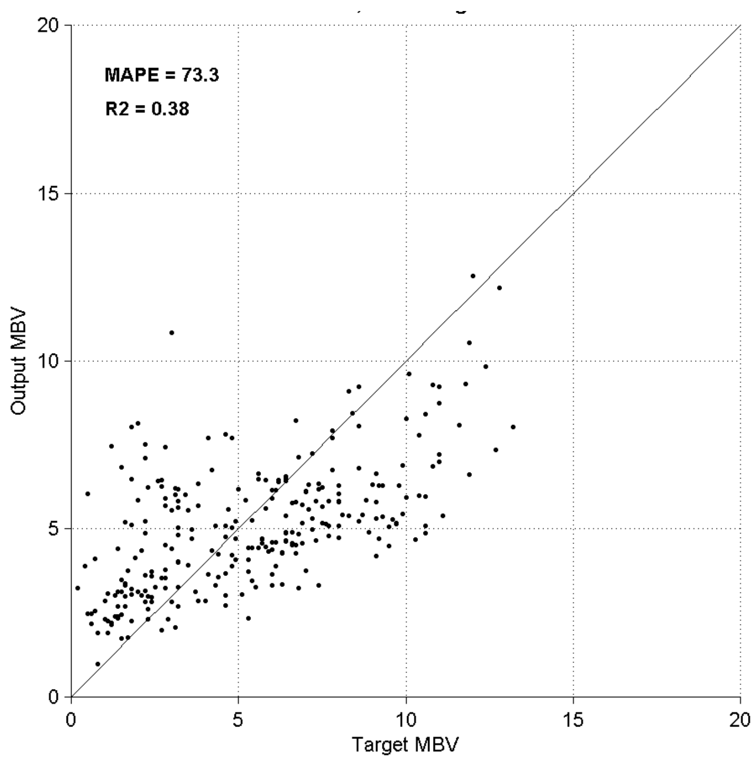


Figure 4.11. Data Set IV, Swell Percent Prediction, Linear Regression Graph

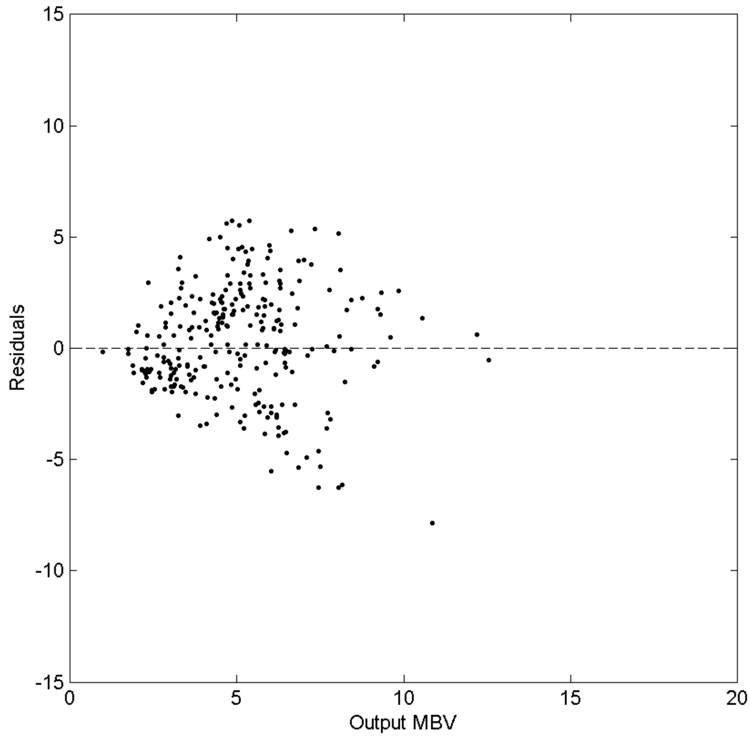


Figure 4.12. Data Set IV, Swell Percent Prediction, Residuals vs. Output Graph

4.3. GEP Models

Genex-Pro 4.3 program is used in our study to implement Genetic Expression Programming. As previously mentioned; Dataset I, Dataset II, Dataset III and Dataset IV are used at MBV prediction analyses. Also predictive models for swelling percent are constructed for Dataset I (subset data) and Dataset II.

Different number of analyses are run for each dataset, in the search of best input set. GEP analysis and results part of the chapter composes of definition of GEP structure in use and analyses results.

4.3.1. Performance Evaluation Criteria

The analyses are evaluated according to their Mean Absolute Percentage Error and fitness values. Mean Absolute Percentage Error (MAPE) is calculated with the function below

$$MAPE = \sum \frac{\left|(T_i - O_{i)}\right|}{T_i} x \frac{100}{n} \qquad \text{Eqn. 4.6}$$

where

T_i: The target value

O_i: Output value

n: Number of data points

Mean Absolute Error (MAE) is used as the fitness function and the equation used for fitness calculation is given below:

$$f_i = 1000 \frac{1}{1 + E_i} \qquad \text{Eqn. 4.7}$$

where E_i : Mean absolute error of an individual program computed by:

$$E_i = \frac{1}{n}\sum_{j=1}^{n}\left|P_{ij} - T_i\right| \qquad \text{Eqn. 4.8}$$

where,

P_{ij}: The value predicted by the individual program i for fitness case j

T_i: The target value for fitness case j

When f_i=1000, the perfect fitting program is obtained.

For both GEP and ANN models, normalized datasets are utilized. The data are normalized by using standard deviation-mean normalization before they are fed into the machine learning algorithms. Standard deviation -mean normalization is performed according to equation 4.9.

$$A_n(i) = \frac{A(i) - \bar{A}}{\sigma_{\mathrm{A}}} \qquad \text{Eqn. 4.9}$$

4.3.2. Genetic Operator Parameters

All of the genetic operators described at the previous section of this chapter are used during the analysis. The list of the operators and their parameters are presented at Table 4.2. Sensitivity analysis are also conducted on these parameters and it is observed that the default parameters yield the optimum solution.

Table 4.2. Genetic Operator Parameters

Genetic Operators	*Parameters*
Mutation	0.0051
Inversion	0.1
IS Transposition	0.1
RIS Transposition	0.1
Gene Transposition	0.3
One-Point Recombination	0.3
Two-Point Recombination	0.1
Gene Recombination	0.1
Constants per gene	10
Lower Bound	-10
Upper Bound	10
RNC Mutation	0.0051
Dc Mutation	0.0051
Dc Inversion	0.1
Dc IS Transposition	0.2

The remaining parameters, which are number of generations, gene number, head size and chromosome number are defined through sensitivity analyses. However, in order to define the best input set, all analyses are run using following operator parameters before conducting sensitivity analyses:

- ✓ Generation number:2000
- ✓ Chromosome number:30
- ✓ Gene Number: 10
- ✓ Head Size:9

In total 65 analyses are run in order to determine the best input set. The results of these analyses are given in Table 4.3, Table 4.4, Table 4.5, Table 4.6, Table 4.7 and Table 4.8. The best models are selected according to the smallest MAPE value and fitness . Sensitivity analyses in order to determine generation number, chromosome number, gene number and head size are applied on the best models.

The best fitting analyses are selected for sensitivity analyses. The program designations of these analyses are:

- ✓ For MBV Prediction;
  - o Dataset I: GEP 9
  - o Dataset II: GEP 1
  - o Dataset III:GEP 8
  - o Dataset IV:GEP 7
- ✓ For Swell Percent Prediction:
  - o Dataset I: GEP 4
  - o Dataset II: GEP 7

Table 4.3. Dataset I, Genetic Expression Programming MBV Prediction

Program No:	Inputs			Output	Dataset I					
					R^2		Fitness		MAPE (%)	
					Training	Testing	Training	Testing	Training	Testing
GEP1	LL	PL		MBV	0,92	0,95	816,0	791,6	3,1	4,0
GEP2	LL	CC		MBV	0,83	0,97	768,3	747,9	4,2	4,7
GEP3	PI	CC		MBV	0,89	0,83	807,0	785,3	3,0	4,1
GEP4	PL	CC		MBV	0,84	0,43	771,4	532,4	4,2	12,4
GEP5	LL	A		MBV	0,84	0,90	765,5	874,9	4,0	2,0
GEP6	LL	PL	CC	MBV	0,79	1,00	749,7	662,5	4,1	5,8
GEP7	LL	PI	CC	MBV	0,90	1,00	796,7	771,6	3,5	4,4
GEP8	PL	PI	CC	MBV	0,76	0,87	731,5	832,9	5,4	2,5
GEP9	A	LL	PL	MBV	0,91	1,00	820,1	832,6	2,7	3,0

Table 4.4. Dataset III, Genetic Expression Programming MBV Prediction

Program No:	Inputs			Output	Dataset III					
					R^2		Fitness		MAPE (%)	
					Training	Testing	Training	Testing	Training	Testing
GEP1	LL	PL		MBV	0,59	0,60	676,5	716,2	7,8	6,5
GEP2	LL	CC		MBV	0,57	0,76	668,5	726,6	7,2	6,4
GEP3	PI	CC		MBV	0,54	0,54	665,2	766,4	8,0	4,5
GEP4	PL	CC		MBV	0,32	0,53	613,2	725,9	9,3	6,6
GEP5	LL	A		MBV	0,57	0,70	678,3	734,2	7,3	6,0
GEP6	LL	PL	CC	MBV	0,53	0,83	662,0	766,5	7,8	5,1
GEP7	LL	PI	CC	MBV	0,51	0,73	662,7	776,2	8,7	4,6
GEP8	PL	PI	CC	MBV	0,56	0,88	691,7	801,5	6,8	3,9
GEP9	A	LL	PL	MBV	0,52	0,81	657,9	762,4	7,9	5,3

Table 4.5. Dataset VI, Genetic Expression Programming MBV Prediction

Program No:	Inputs			Output	Dataset IV					
					R^2		Fitness		MAPE (%)	
					Training	Testing	Training	Testing	Training	Testing
GEP1	LL	PL		MBV	0,46	0,41	633,5	656,8	14,8	10,1
GEP2	LL	CC		MBV	0,39	0,41	619,9	673,1	19,3	8,9
GEP3	PI	CC		MBV	0,45	0,47	639,6	693,8	18,3	8,6
GEP4	PL	CC		MBV	0,10	0,11	556,0	581,5	22,5	11,9
GEP5	LL	A		MBV	0,42	0,34	628,0	649,0	17,2	11,3
GEP6	LL	PL	CC	MBV	0,45	0,54	632,9	696,8	17,6	8,8
GEP7	LL	PI	CC	MBV	0,40	0,53	639,1	695,0	17,7	7,8
GEP8	PL	PI	CC	MBV	0,40	0,45	628,1	672,7	19,1	9,7
GEP9	A	LL	PL	MBV	0,45	0,45	631,2	671,9	16,4	9,7

Table 4.6. Dataset II, Genetic Expression Programming MBV Prediction

P. No:	Dataset II											
	Inputs					Output	R2		Fitness		MAPE (%)	
							Train.	Test.	Train.	Test.	Train.	Test.
GEP1	LL	PL	CC	ρ_{dry}	w_{opt}	MBV	0,81	0,83	764,6	817,4	4,0	3,2
GEP2	LL	PI	CC	ρ_{dry}	w_{opt}	MBV	0,70	0,90	704,2	815,2	5,6	3,3
GEP3	PL	PI	CC	ρ_{dry}	w_{opt}	MBV	0,80	0,93	738,0	547,4	4,9	2,7
GEP4	A	LL	PL	ρ_{dry}	w_{opt}	MBV	0,64	0,89	708,8	787,1	5,9	3,9
GEP5	LL	PL	CC	ρ_{dry}		MBV	0,68	0,98	704,6	877,2	6,0	2,3
GEP6	LL	PI	CC	ρ_{dry}		MBV	0,71	0,75	716,9	817,6	5,6	3,1
GEP7	PL	PI	CC	ρ_{dry}		MBV	0,63	0,91	697,7	865,6	5,7	2,4
GEP8	A	LL	PL	ρ_{dry}		MBV	0,56	0,81	670,5	822,3	7,8	3,4
GEP9	LL	PL	CC	w_{opt}		MBV	0,73	0,89	719,2	864,8	5,8	2,3
GEP10	LL	PI	CC	w_{opt}		MBV	0,73	0,88	724,4	823,9	5,2	3,0
GEP11	PL	PI	CC	w_{opt}		MBV	0,76	0,86	727,6	799,3	5,3	3,7
GEP12	A	LL	PL	w_{opt}		MBV	0,78	0,77	739,4	828,6	4,9	3,1

Table 4.7. Dataset I, Genetic Expression Programming Swell Percent Prediction

Program No:	Inputs						Outputs	Dataset I					
								R^2		Fitness		MAPE (%)	
								Training	Testing	Training	Testing	Training	Testing
GEP1	LL	PL	CC	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,93	1,00	839,2	707,5	8,8	18,2
GEP2	LL	PI	CC	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,87	1,00	804,0	564,4	8,9	31,0
GEP3	PL	PI	CC	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,82	1,00	769,0	398,7	9,7	70,6
GEP4	A	LL	PL	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,96	1,00	867,0	814,6	4,9	6,0
GEP5	LL	PL	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,92	1,00	836,6	591,3	5,3	39,5
GEP6	LL	CC	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,89	1,00	813,9	622,8	8,8	19,0
GEP7	PI	CC	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,85	1,00	769,7	670,0	8,5	18,4
GEP8	PL	CC	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,83	1,00	796,2	46,9	12,2	386,3
GEP9	LL	A	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,94	1,00	858,6	464,5	5,1	50,7
GEP10	LL	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,79	1,00	763,4	849,3	13,1	8,5
GEP11	PL	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,79	1,00	789,0	113,1	8,0	190,9
GEP12	PI	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,88	1,00	782,5	269,0	8,7	136,3
GEP13	CC	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,92	1,00	783,9	633,3	7,5	29,8

Table 4.8. Dataset II, Genetic Expression Programming Swell Percent Prediction

Program No:	Inputs						Outputs	Dataset II					
								R^2		Fitness		MAPE (%)	
								Training	Testing	Training	Testing	Training	Testing
GEP1	LL	PL	CC	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,84	0,90	765,2	806,0	26,1	21,1
GEP2	LL	PI	CC	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,87	0,67	805,8	802,3	20,7	24,4
GEP3	PL	PI	CC	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,82	0,96	813,1	821,5	23,5	18,5
GEP4	A	LL	PL	ρ_{dry}	w_{opt}	MBV	Swell Percent	0,79	0,98	770,2	884,9	22,2	9,0
GEP5	LL	PL	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,92	0,71	838,3	837,9	1730	14,6
GEP6	LL	CC	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,86	0,75	796,8	846,9	22,0	19,1
GEP7	PI	CC	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,83	0,97	825,2	860,3	18,1	14,8
GEP8	PL	CC	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,83	0,82	804,9	818,4	15,6	23,9
GEP9	LL	A	ρ_{dry}	w_{opt}	MBV		Swell Percent	0,94	0,71	842,7	803,6	17,8	22,8
GEP10	LL	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,81	0,27	746,9	661,0	30,0	38,5
GEP11	PL	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,83	0,46	792,0	830,0	23,4	10,7
GEP12	PI	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,81	0,91	784,0	895,3	21,6	12,3
GEP13	CC	ρ_{dry}	w_{opt}	MBV			Swell Percent	0,86	0,66	808,0	802,1	20,2	27,2

Sensitivity analyses concerning the aforementioned parameters for MBV and Swell percent predictions are given in Figure 4.13 to Figure 4.18.

The final parameters of the analyses are selected according to sensitivity analyses. As can be seen from the figures. Parameter selection has a slight affect on the fitness values. The selected parameters are given at Table 4.9.

Table 4.9. Selected Parameters from Sensitivity Analyses

		# of Generations	Chromosome #	Gene #	Head Size
MBV	Dataset I	2500	30	11	10
	Dataset II	2500	30	5	7
	Dataset III	2000	90	8	7
	Dataset IV	100000	30	7	5
Swell Percent	Dataset I	2500	110	4	3
	Dataset II	5000	30	5	6

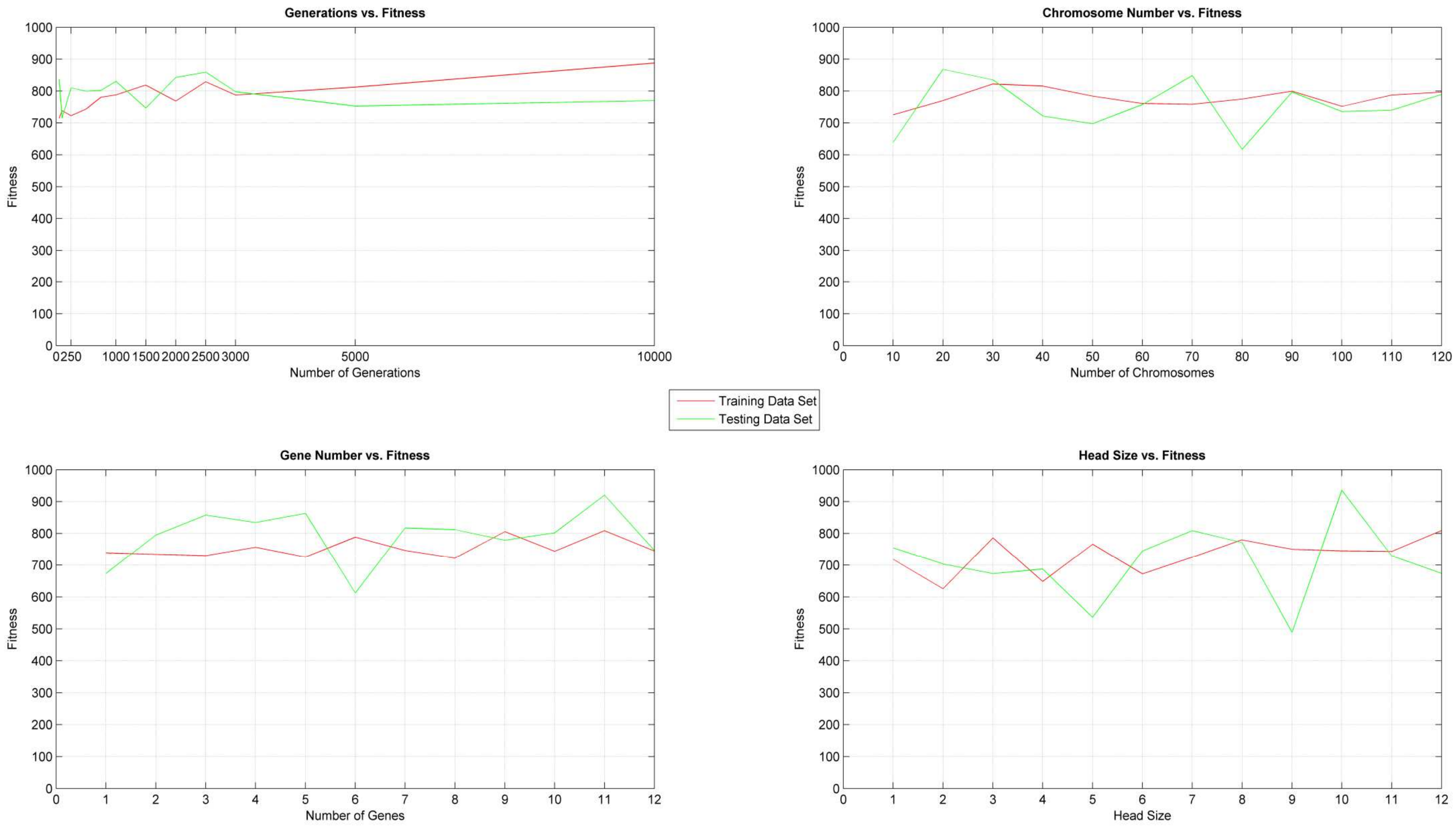


Figure 4.13. Dataset I , MBV Prediction Sensitivity Analyses

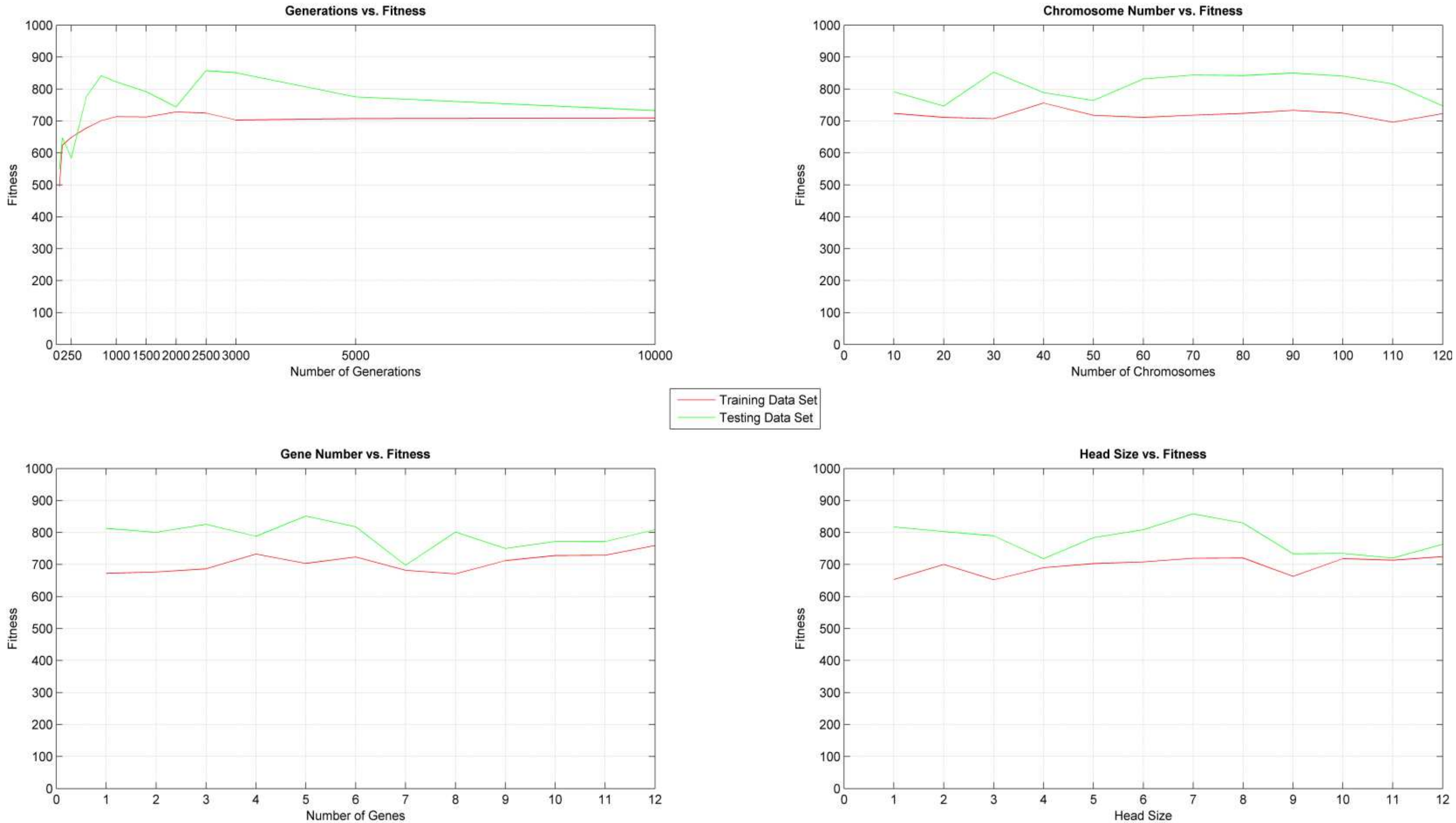


Figure 4.14. Dataset II , MBV Prediction Sensitivity Analyses

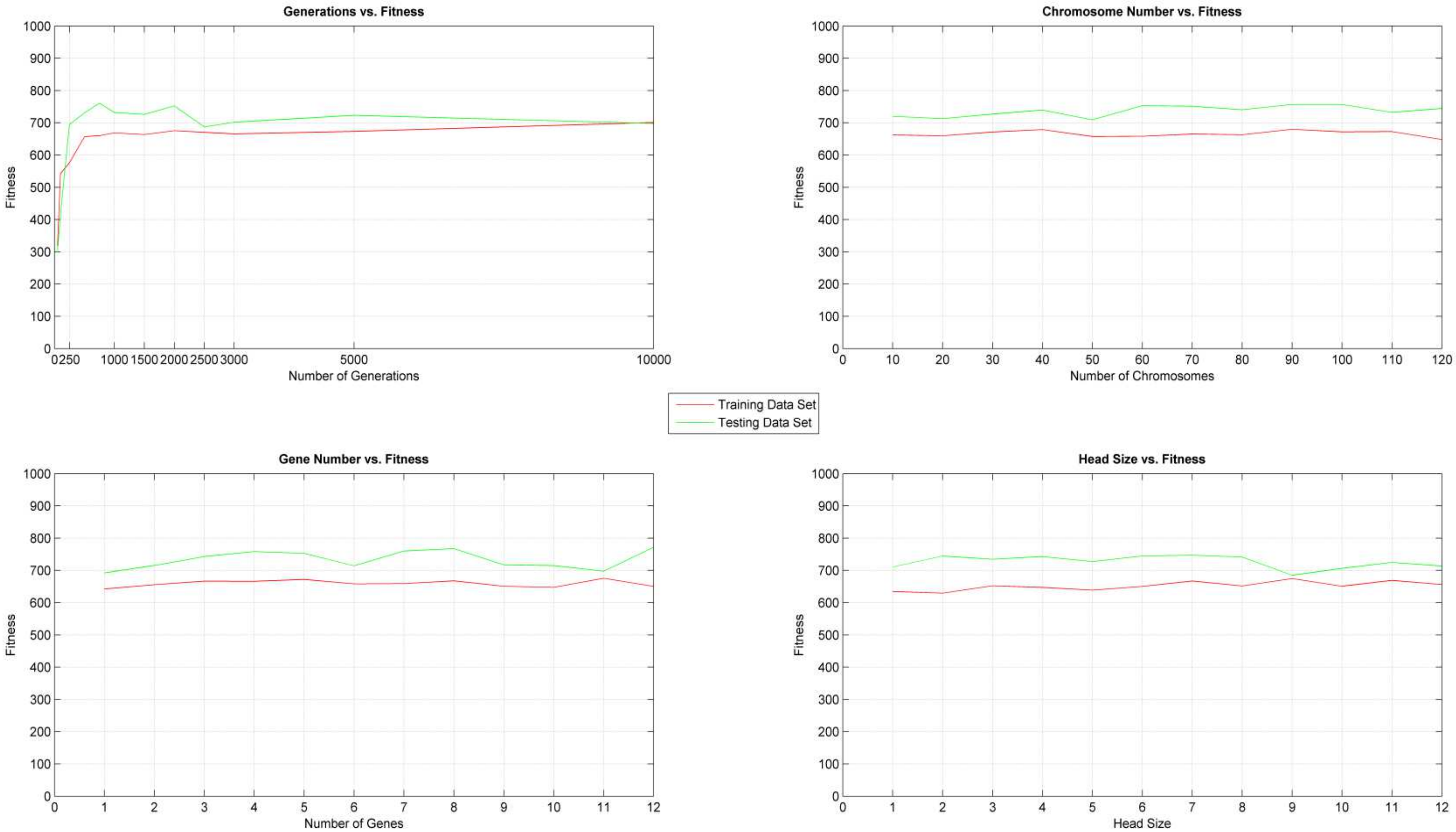


Figure 4.15. Dataset III , MBV Prediction Sensitivity Analyses

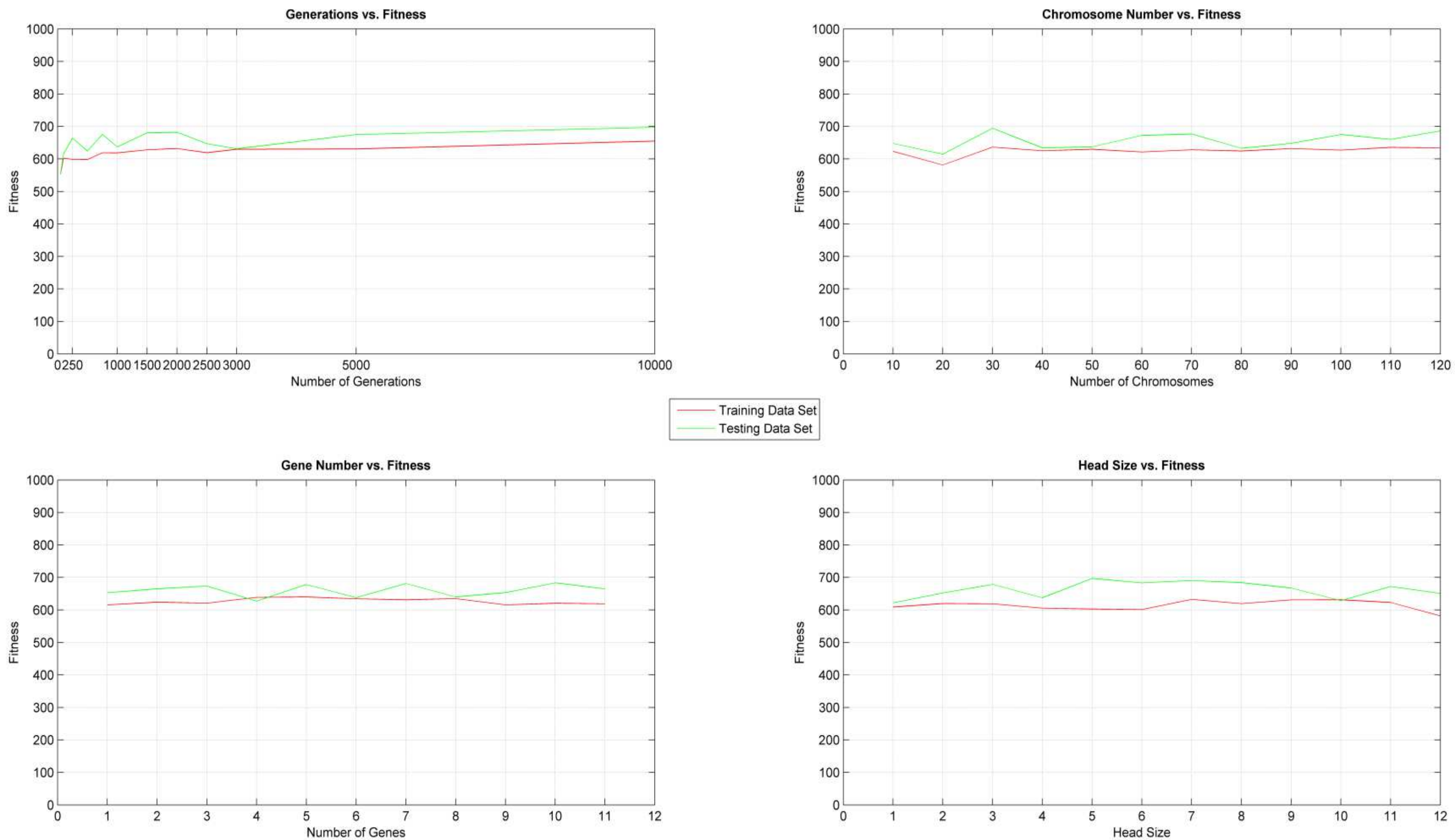


Figure 4.16. Dataset IV , MBV Prediction Sensitivity Analyses

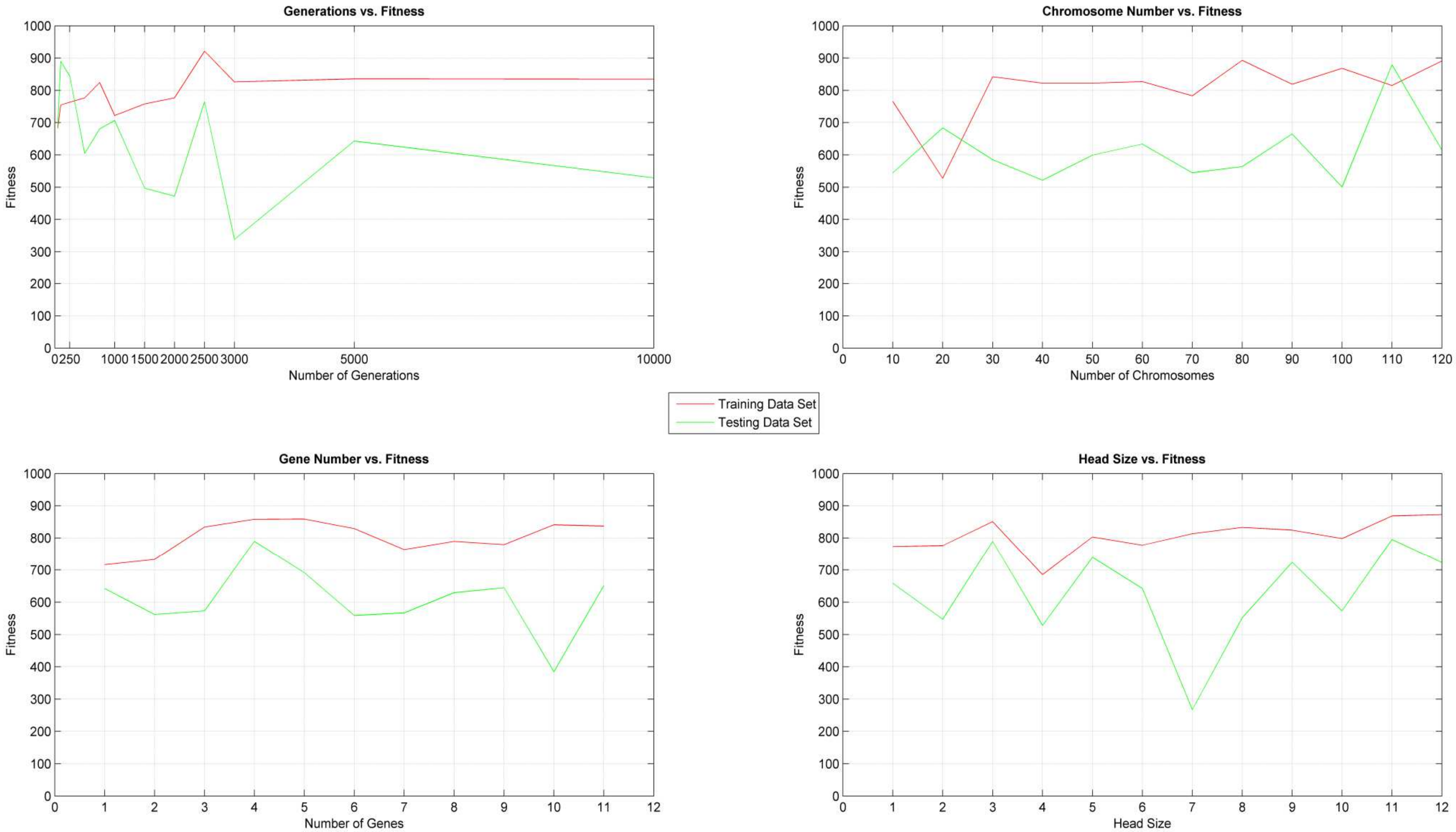


Figure 4.17. Dataset I , Swell Percent Prediction Sensitivity Analyses

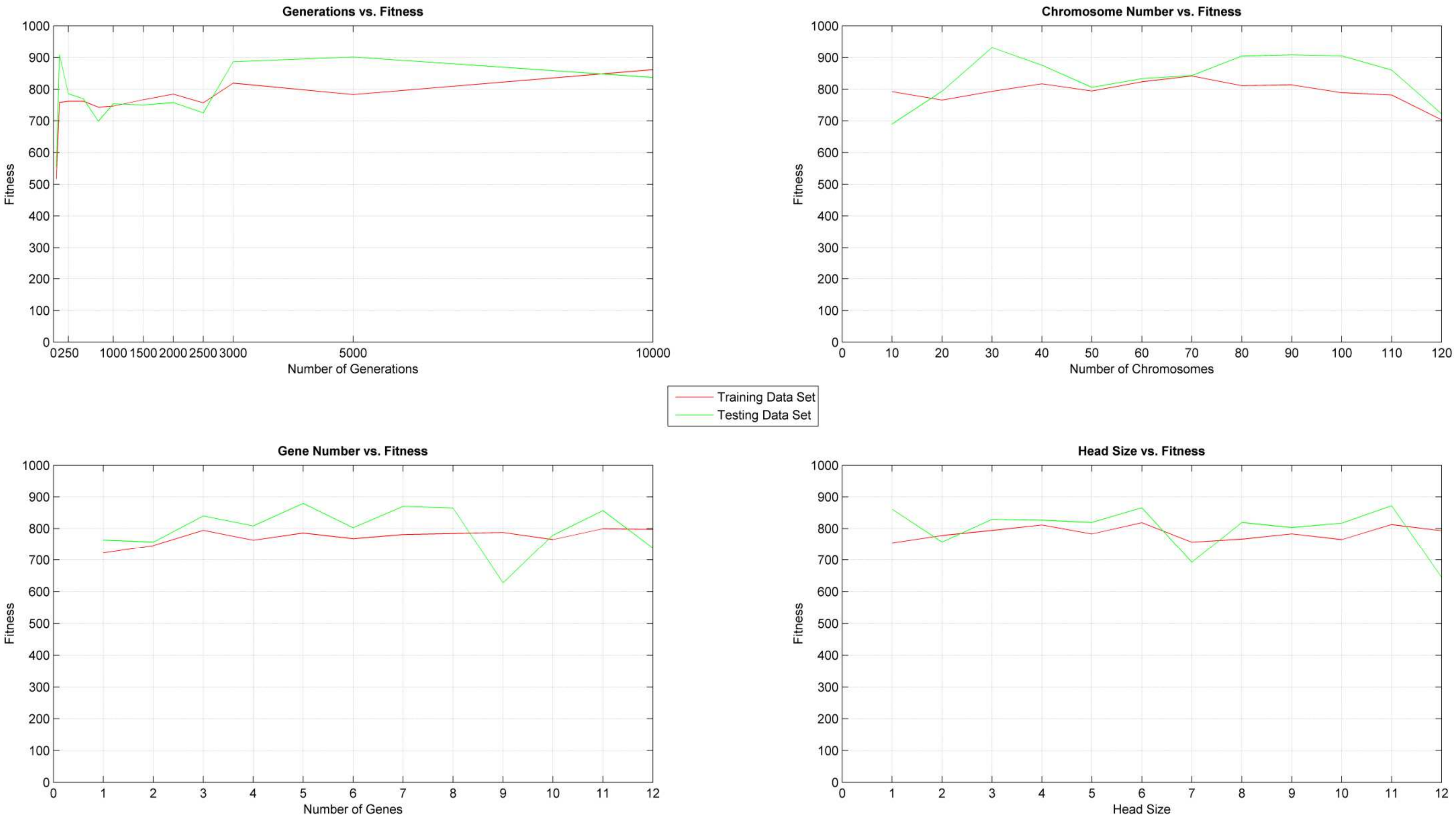


Figure 4.18. Dataset II , Swell Percent Prediction Sensitivity Analyses

4.3.3. Analyses Results

The analyses are repeated using the selected parameters and 5-fold cross validation is applied. The results of the 5-fold cross validation results and the expression tree of the best analyses are given at appendices. The best analyses are selected according to testing MAPE values and the summary of these analyses results are given at Table 4.10.

Table 4.10. Genetic Expression Programming Results

Dataset	NN No	Input Data	Output Data	Testing	
				R^2	MAPE (%)
Dataset I	GEP 9	A, LL, PL	MBV	0,98	8,3
Dataset II	GEP 1	LL, PL, CC, γ_{dry},w_{opt}	MBV	0,75	14,5
Dataset III	GEP 3	PL, PI, CC	MBV	0,58	13,2
Dataset IV	GEP 7	LL, PI, CC	MBV	0,45	46,3
Dataset I	GEP 4	A, LL, PL, CC, γ_{dry}, w_{opt}, MBV	Swell Percent	1,00	11,5
Dataset II	GEP 7	PI, CC, γ_{dry},w_{opt}, MBV	Swell Percent	0,78	22,1

"Targets vs. Outputs" graphs are given in Figure 4.19 to Figure 4.24.

For the best performing genetic expression programming models for MBV prediction, the average MAPE values change between 13,2 % and 14.5 %, which are moderately good considering the applicability of these models to more comprehensive datasets. The MAPE values of swell percent models change between 11,5% and 22,1%, which provide much better models than linear regression analysis. Even though the performance of the predicted models are better than linear regression, again as the dataset gets larger, models tend to perform worse.

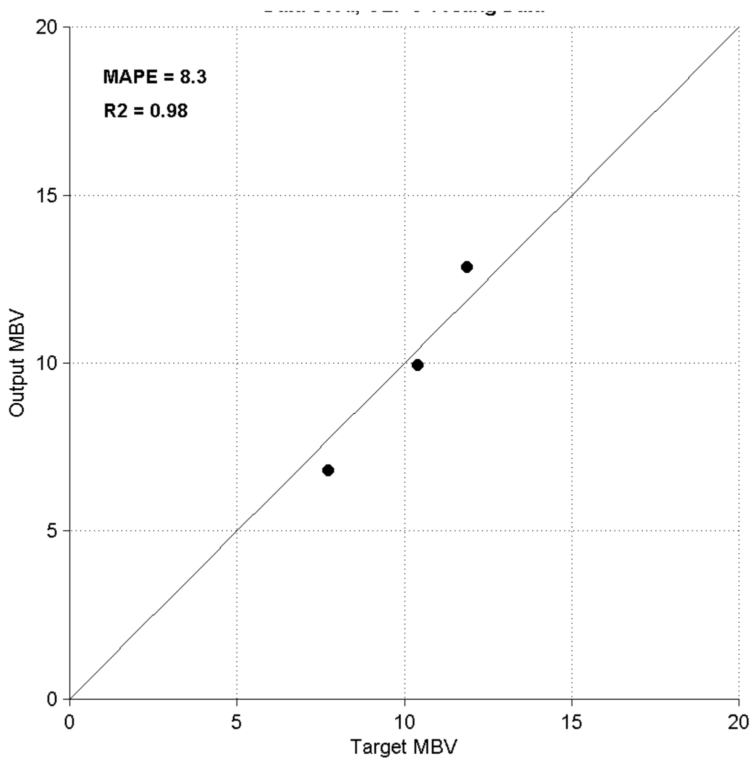


Figure 4.19. Dataset I, MBV Prediction, Target vs. Output Graph of GEP 9 Testing Data

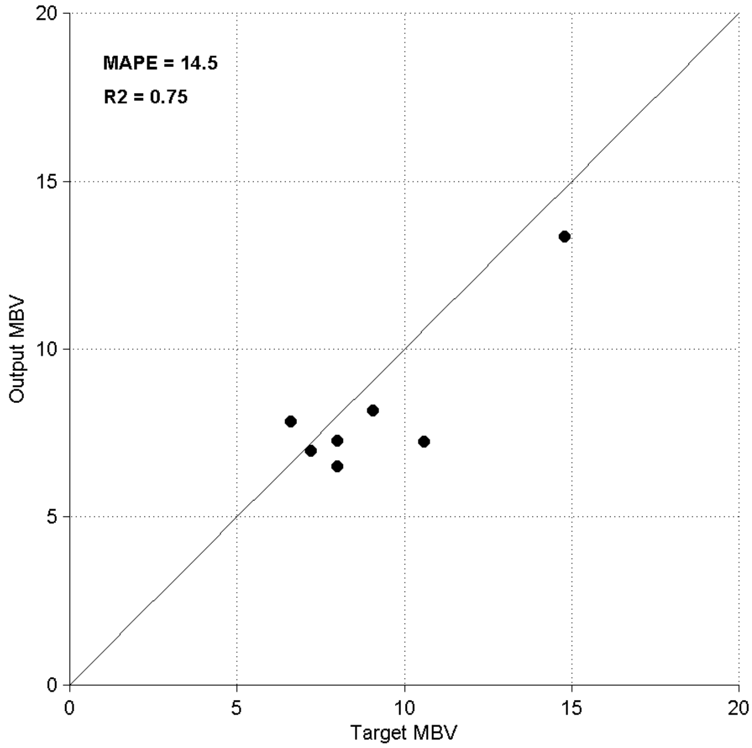


Figure 4.20. Dataset II, MBV Prediction, Target vs. Output Graph of GEP 1 Testing Data

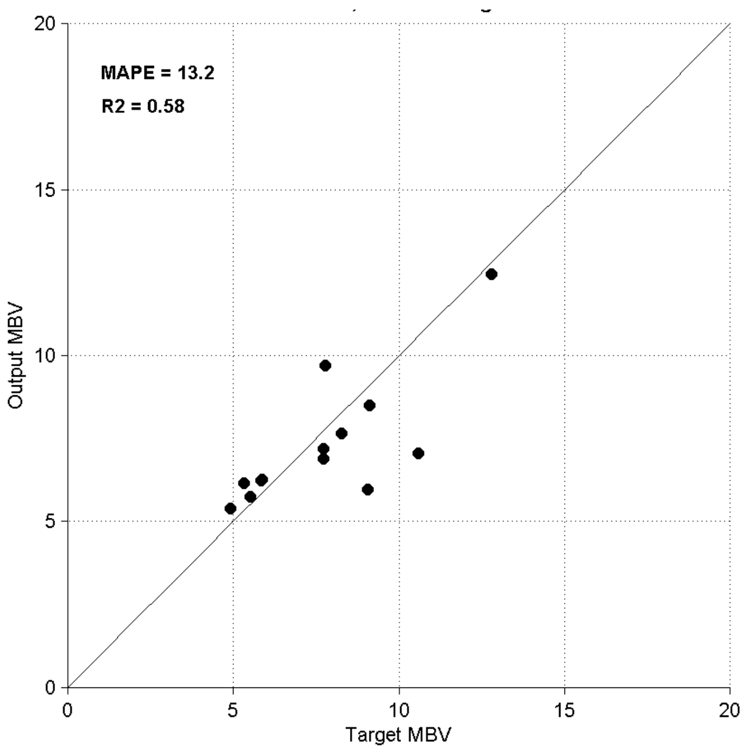


Figure 4.21. Dataset III, MBV Prediction, Target vs. Output Graph of GEP 8 Testing Data

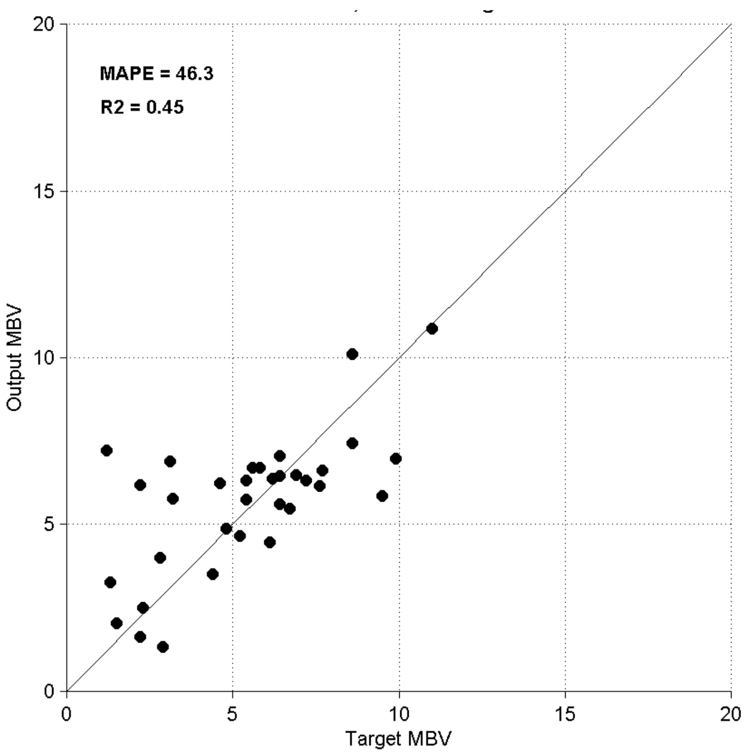


Figure 4.22. Dataset IV, MBV Prediction, Target vs. Output Graph of GEP 7 Testing Data

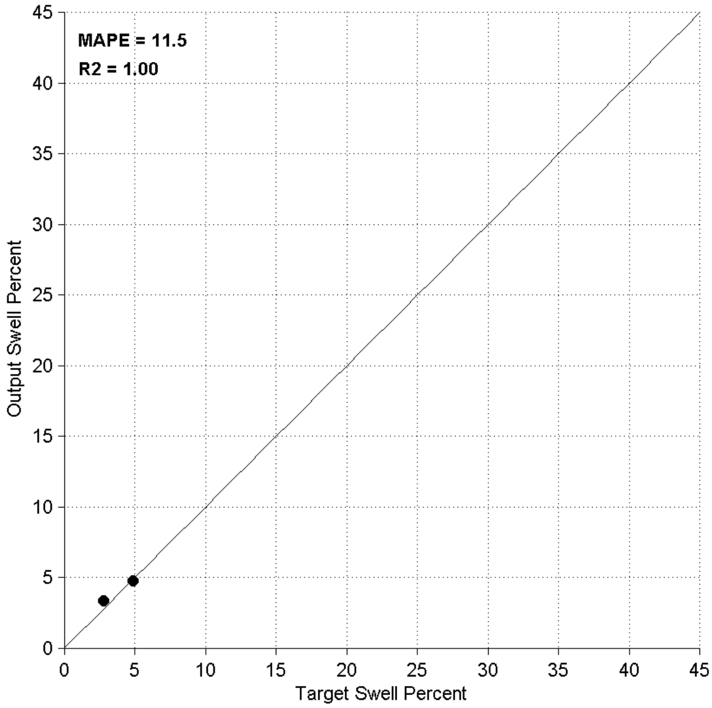


Figure 4.23. Dataset I, Swell Percent Prediction, Target vs. Output Graph of GEP 4 Testing Data

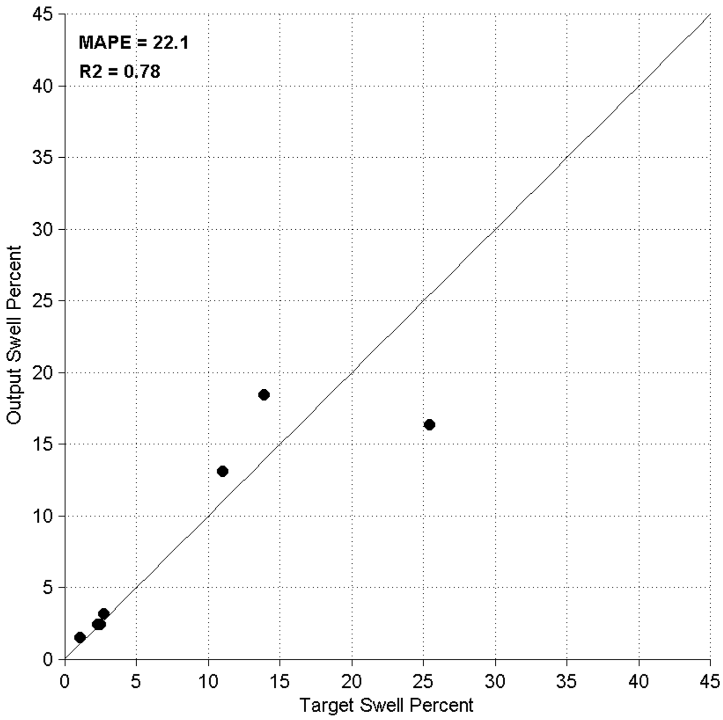


Figure 4.24. Dataset II, Swell Percent Prediction, Target vs. Output Graph of GEP 7 Testing Data

4.4.Neural Network Models

In our study, neural networks with 1 hidden layer are used in analyses. Number of neurons are determined using sensitivity analyses with different neuron numbers, such as; 3, 5, 10 and 20. The analyses giving the best fitness are presented below.

- ✓ For MBV Prediction;
  - Dataset I: 3 neurons
  - Dataset II: 3 neurons
  - Dataset III: 10 neurons
  - Dataset IV: 5 neurons
- ✓ For Swell Percent Prediction:
  - Dataset I: 3 neurons
  - Dataset II: 3 neurons

The dataset is divided into training, cross validation and testing data using following proportions, 70% for training data, 10% for cross validation and 10% for testing data. Training data is the dataset used in training process, validation data is used for the performance checks of the network. Testing data on the other hand is used when the network is desired to be compared with different models performances.

Levenberg marquardt algorithm (Haykin, 2009) is used while training the neural network and mean square error is used as the performance measure. Tan-Sigmoid function (Haykin, 2009) is used as transfer function. Stopping when the increase in performance does not exceed 1e-5 for 6 validation calculations is used as the stopping criteria.

5-fold cross validation is used for selecting the network system. Cross validation and testing data are randomly chosen for 5 times and the best performing network is presented.

Neural network analyses are conducted for MBV and Swell Percent predictions. All datasets are used at MBV prediction and Dataset I (subset of Data I) and Dataset II are used for swell percent determinations. In order to obtain the best input set, 65 different analyses are run. The input set are selected according to the testing sets MAPE values. The best performing networks' input data, hidden layer, neuron numbers and neural network properties are presented at Table 4.11.

Table 4.11. Neural network properties

NN No	Dataset	Input Data	Hidden Layer #	# of neurons	Training Function	Transfer Function
MBV Prediction						
NN1	Dataset I	PI, CC	1	3	Levenberg Marquardt	Tan-Sigmoid
NN2	Dataset II	LL,PL,CC, ρ_{dry},w_{opt}	1	3	Levenberg Marquardt	Tan-Sigmoid
NN3	Dataset III	PL, LL	1	10	Levenberg Marquardt	Tan-Sigmoid
NN4	Dataset IV	LL, A	1	5	Levenberg Marquardt	Tan-Sigmoid
Swell Percent Prediction						
NN5	Dataset I	A, LL, ρ_{dry}, w_{opt}, MBV	1	3	Levenberg Marquardt	Tan-Sigmoid
NN6	Dataset II	A, LL, PL, ρ_{dry}, w_{opt}, MBV	1	3	Levenberg Marquardt	Tan-Sigmoid

The analysis results (Mean absolute performance error and coefficient of determination values) are presented in Table 4.12.

Table 4.12. Neural Network Analyses Results

Dataset	NN No	Input Data	Output Data	Testing	
				R^2	MAPE (%)
Dataset I	NN1	PI, CC	MBV	1,00	4,2
Dataset II	NN2	LL, PL, CC, ρ_{dry}, w_{opt}	MBV	0,88	5,0
Dataset III	NN3	LL, PL	MBV	0,28	11,5
Dataset IV	NN4	LL, A	MBV	0,81	30,6
Dataset I	NN5	LL, A, ρ_{dry}, w_{opt}, MBV	SP	1,00	1,8
Dataset II	NN6	A, LL, PL, ρ_{dry}, w_{opt}, MBV	SP	0,99	20,7

For the best performing neural network models for MBV prediction, the average MAPE values changs between 5.0 % and 11.5 %, which are quite good considering the applicability of these models to more comprehensive datasets. Although as the dataset gets larger the models work worse in general, the positive effect of introducing γ_{dry} and w_{opt} into the input data can be seen. The MAPE values of swell percent models change between 1,8% and 20,7%, eventhough neural network analyses provide much better models, as the dataset gets larger the models work worse. That is an indicator of database not being constructed carefully considering all the variability of laboratory experiments.

Training performances and "Targets vs Outputs" graphs of Dataset I, MBV prediction (NN1) are presented in Figure 4.25 and Figure 4.26. Training performances and "Targets vs Outputs" graphs of Dataset II, MBV prediction (NN2) are presented in Figure 4.27 and Figure 4.28. Training performances and "Targets vs Outputs" graphs of Dataset III, MBV prediction (NN3) are presented in Figure 4.29 and Figure 4.30. Training performances and "Targets vs Outputs" graphs of Dataset IV, MBV prediction (NN4) are presented in Figure 4.31 and Figure 4.32. Training performances and "Targets vs Outputs" graphs of Dataset I, Swell Percent prediction (NN5) are presented in Figure 4.33 and Figure 4.34. Training performances and "Targets vs Outputs" graphs of Dataset II, Swell Percent prediction (NN6) are presented in Figure 4.35 and Figure 4.36.

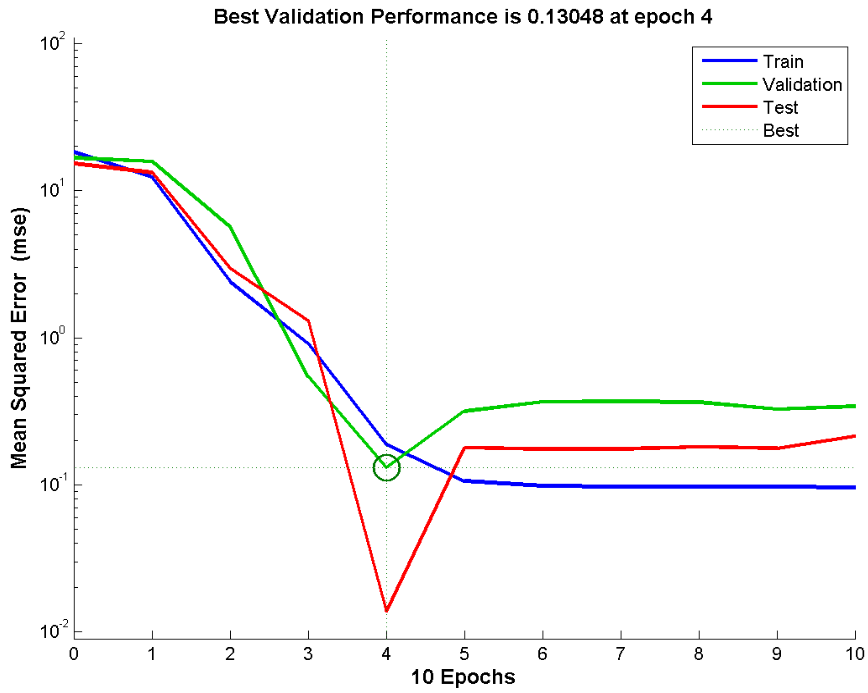


Figure 4.25. Dataset I, MBV Prediction, Training performance of NN1

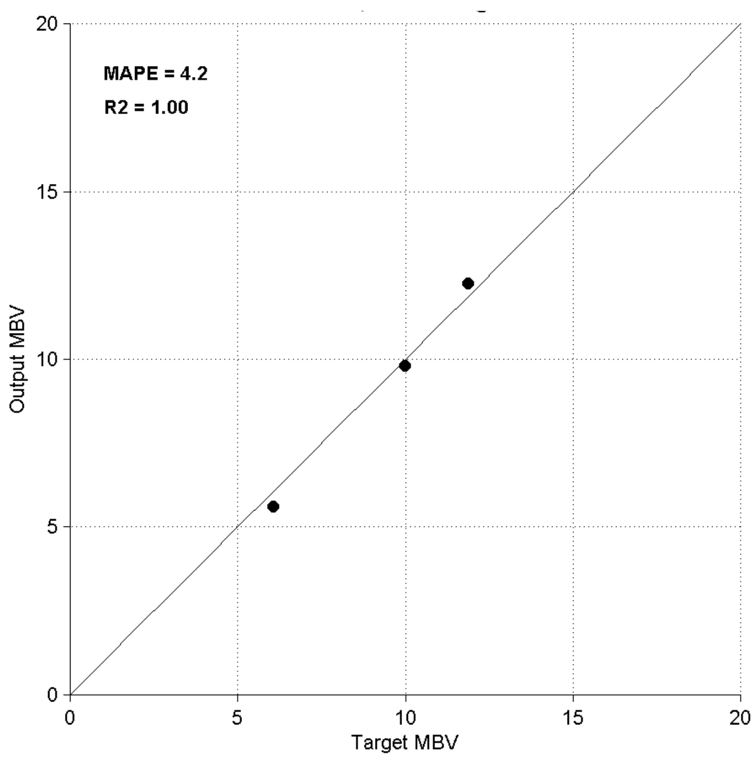


Figure 4.26. Dataset I, MBV Prediction, Target vs. Output Graph of NN1 Testing Data

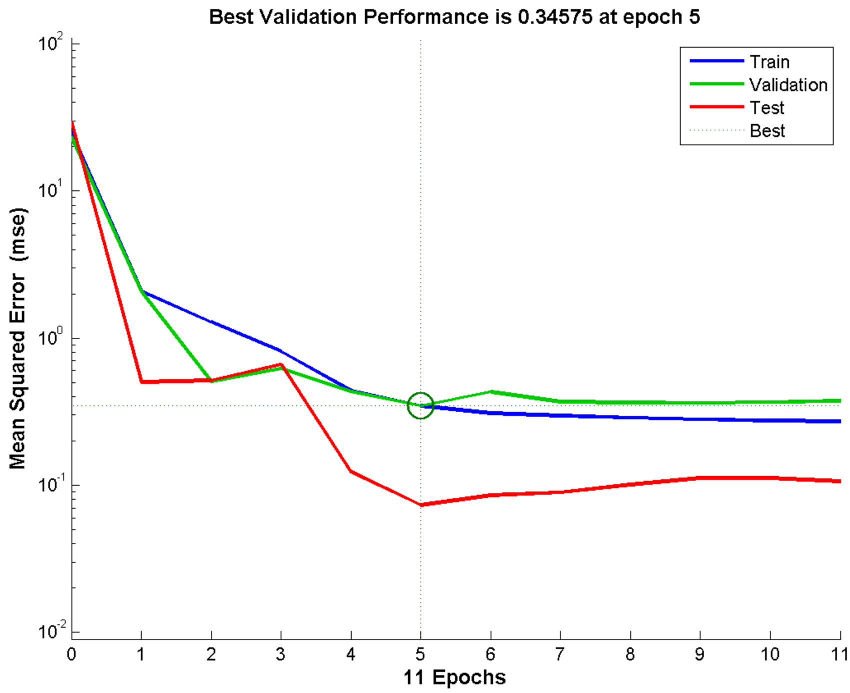


Figure 4.27. Dataset II, MBV Prediction, Training performance of NN2

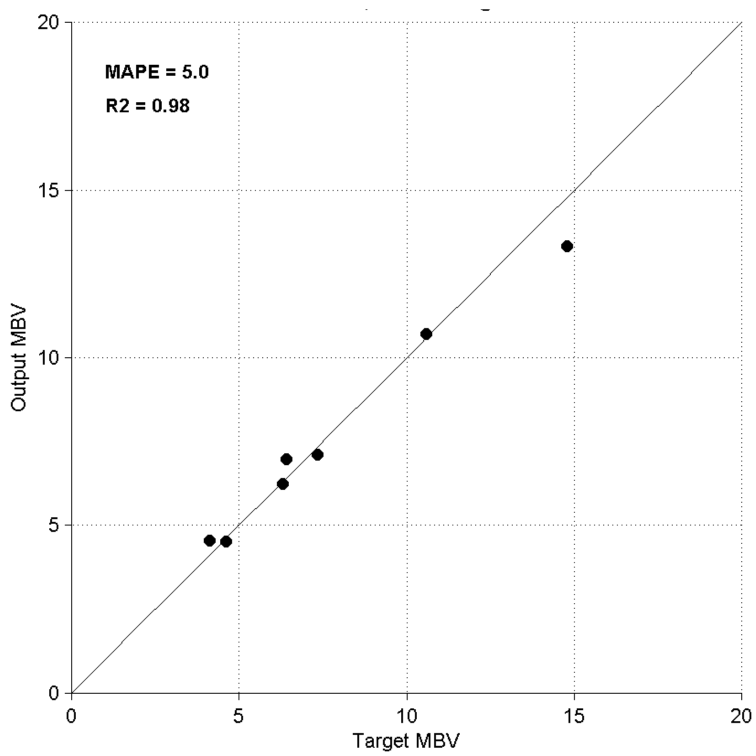


Figure 4.28. Dataset II, MBV Prediction, Target vs. Output Graph of NN2 Testing Data

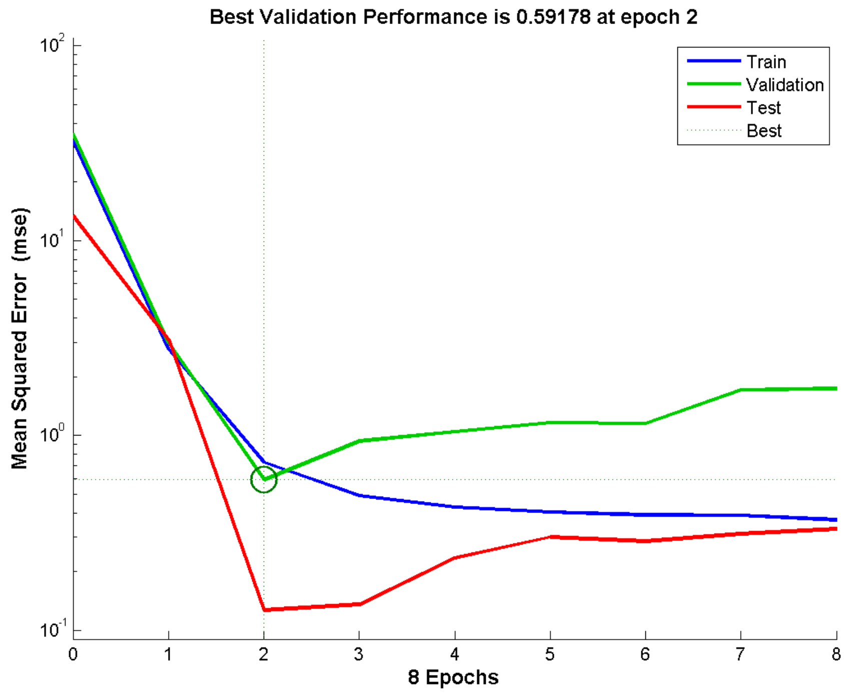


Figure 4.29. Dataset III, MBV Prediction, Training performance of NN3

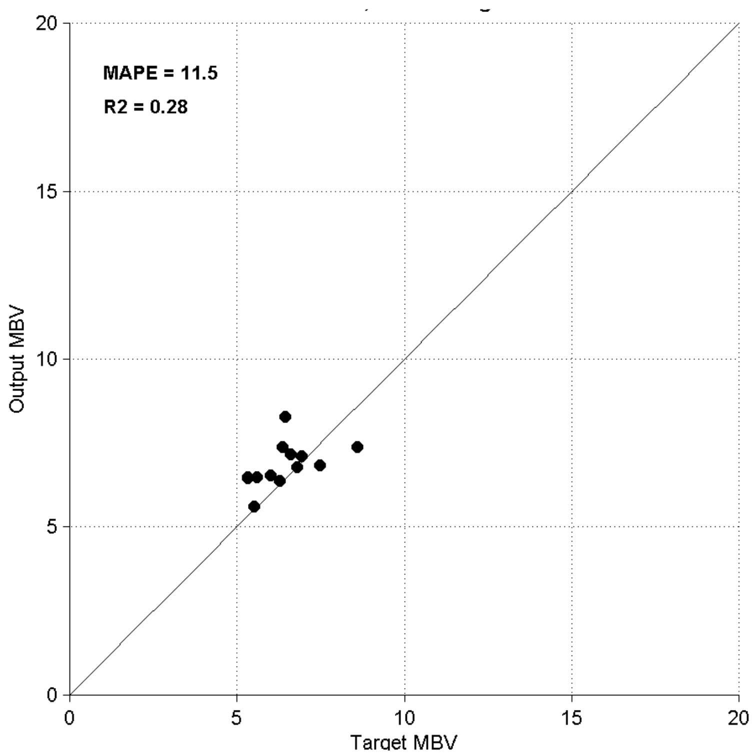


Figure 4.30. Dataset III, MBV Prediction, Target vs. Output Graph of NN3 Testing Data

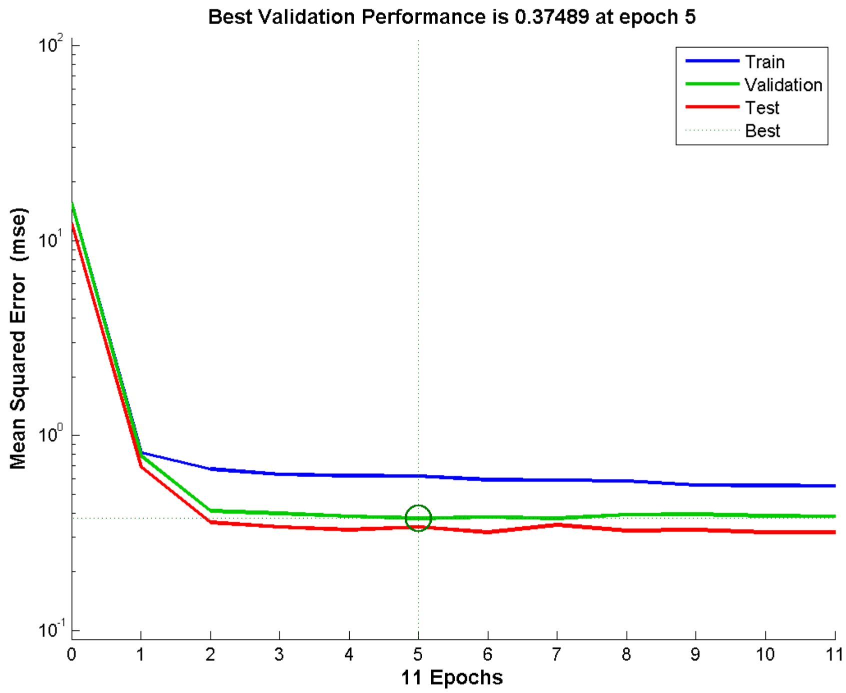


Figure 4.31. Dataset IV, MBV Prediction, Training performance of NN4

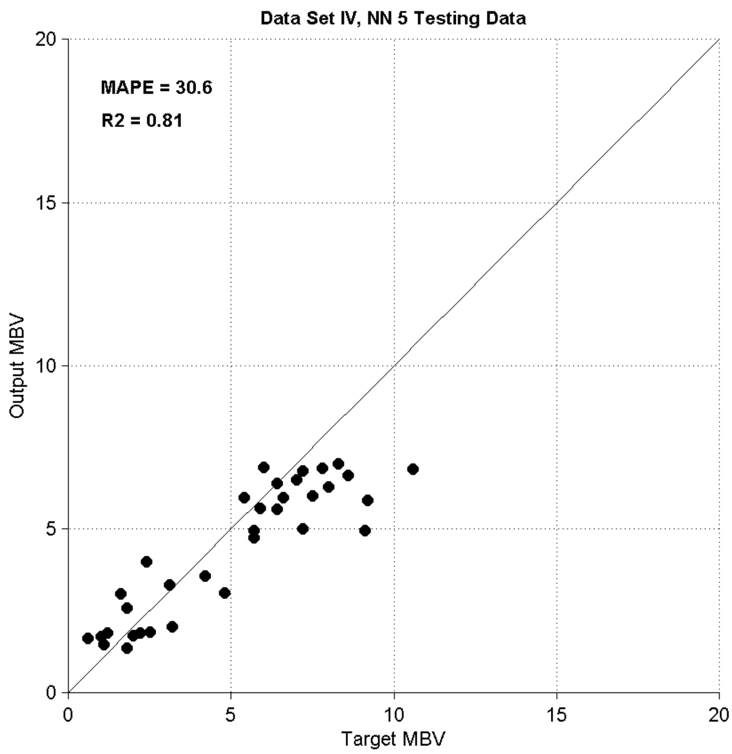


Figure 4.32. Dataset IV, MBV Prediction, Target vs. Output Graph of NN4 Testing Data

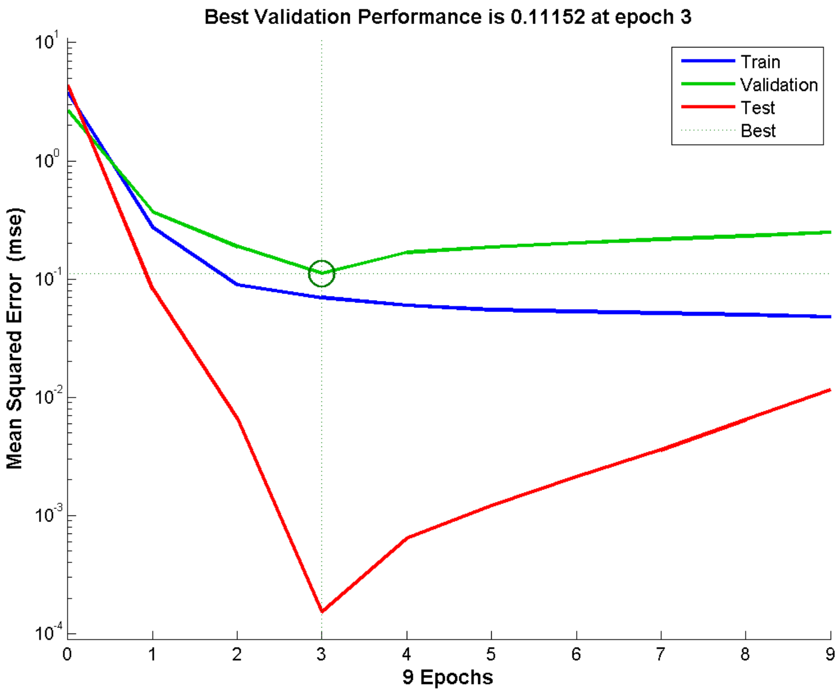


Figure 4.33. Dataset I, Swell Percent Prediction, Training performance of NN5

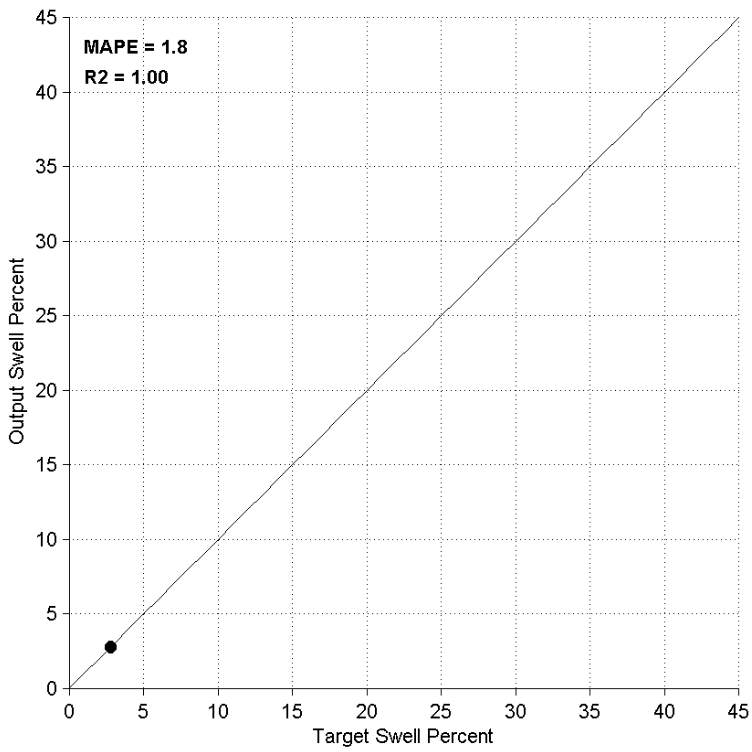


Figure 4.34. Dataset I, Swell Percent Prediction, Target vs. Output Graph of NN5 Testing Data

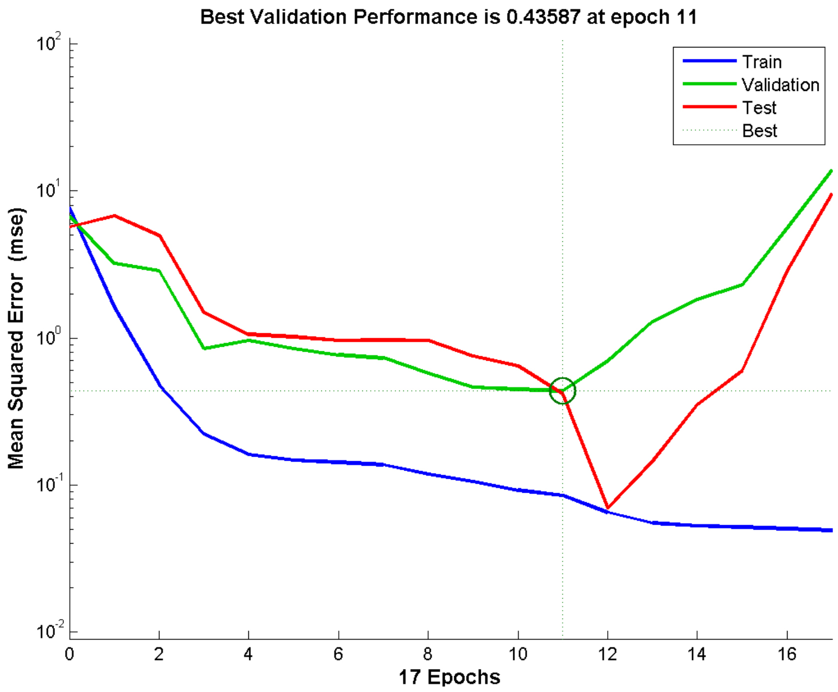


Figure 4.35. Dataset II, Swell Percent Prediction, Training performance of NN6

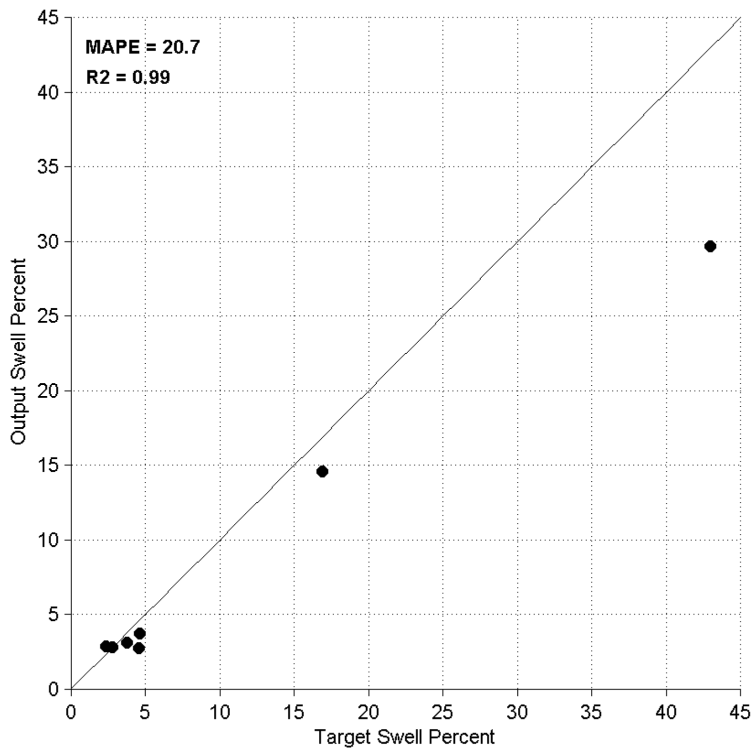


Figure 4.36. Dataset II, Swell Percent Prediction, Target vs. Output Graph of NN6 Testing Data

CHAPTER 5

SUMMARY, CONCLUSIONS AND FUTURE STUDIES

5.1.Summary

Identification of clayey soils with high swelling potential is an important engineering problem as these soils have serious impact on the sustainability of engineering infrastructure. Calculating the swelling potential of a clayey soil is also challenging as there are various factors affecting its potential such as environmental effects, the mineralogy, etc. and it may be an expensive practice to determine all in the laboratory. Engineers need practical means to identify the swelling potential of clayey soils.

The research presented aims to determine the swelling percent of a clayey soil using the basic soil characterization test such as Atterberg limit tests, hydrometer tests as well as simple methylene blue (MB) tests. In addition, the methylene blue value (MBV) is also predicted with the same soil characteristic properties. Within this scope, a database of various clayey soils with properties such as Atterberg limits, clay content and MBV is compiled from numerous resources including the results of our own laboratory experiments.

The compiled dataset consists of 343 samples, which is divided into four groups and each group is used to identify different soil properties such as swell percent of our own soil samples, swell percent of soil samples collected from Turkey, MB value of our own soil samples, MB values of all soil samples collected from all over Turkey and from around the world, respectively. Dataset I contains the laboratory test result conducted within the scope of our study. 32 samples, obtained from different regions at Turkey are tested to obtain Atterberg limits, clay contents and MBVs. On a subset of these samples (20 samples) maximum dry density, optimum water content, swell percent and swell pressure tests are conducted. Dataset II contains both MBV's and swell percent data, and consists of 73 data samples. Dataset III is comprised of 125 data, combined from laboratory test results of samples gathered from Turkey only. Finally, Dataset IV is the one consisting of 343 data which is obtained from various sources worldwide. All the data are used for MBV prediction, on the other hand, only dataset I and II are used for swell percent determination studies.

In order to understand various relationships among the database, robust multivariate linear regression techniques are attempted first. Then advanced machine learning techniques genetic expression programming (GEP) and artificial neural networks (ANN) are used to model the nonlinear relationships in the existing database to determine (i) swell percent and (ii) MBV. The results proved that machine learning techniques proved to work well for estimation of swell percent and MBV for samples collected from Turkey and all over the world. The summary of the predictive models is given in Table 5.1 and Figure 5.1.

Table 5.1. Summary of Predictive Models

	MBV Prediction				Swell Percent Prediction	
MAPE (%)	Dataset I	Dataset II	Dataset III	Dataset IV	Dataset I	Dataset II
Linear Regression	17.4	14.0	19.8	73.3	44.6	57.9
GEP	8.3	14.5	13.2	46.3	11.5	22.1
Neural Networks	4.2	5.0	11.5	30.6	1.8	20.7

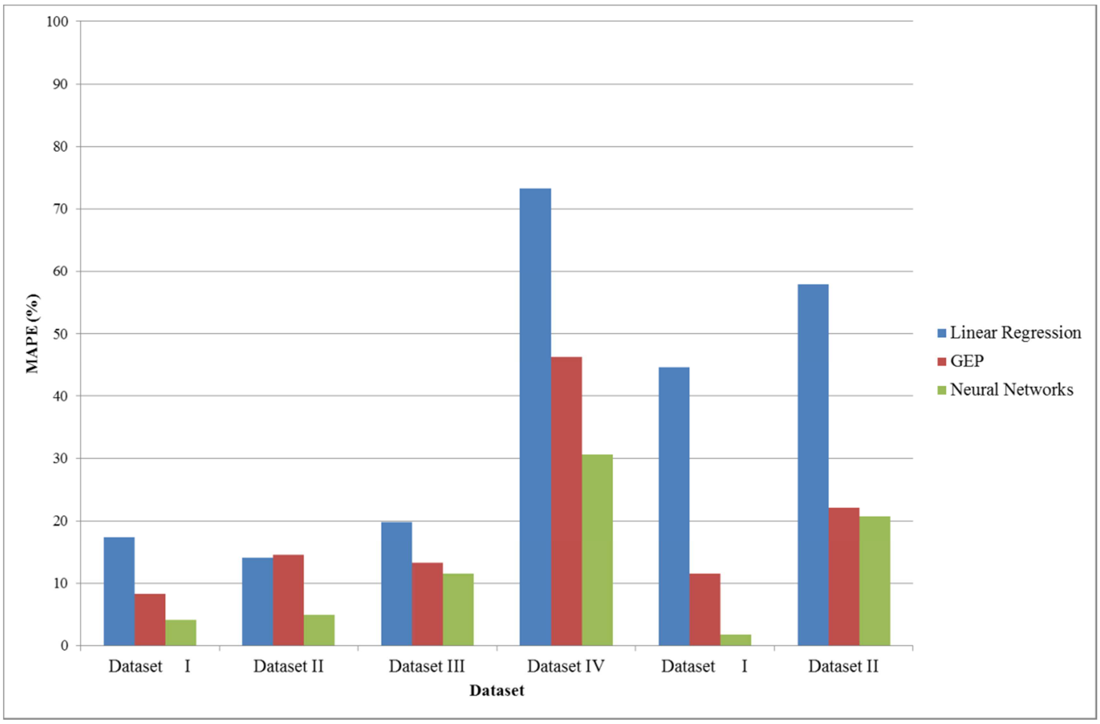


Figure 5.1. Summary of Predictive Models

5.2.Conclusions

Various conclusions inferred from our study are listed as follows:

- Multivariate linear regression analyses prove that, the variables in the database are not linearly correlated. The values of mean absolute percentage errors (MAPE) for the best predictive models for MBV prediction of Dataset I, II, III 17.4%, 14.0%, and 19.8%, respectively and the ones for determination of swell percent for Dataset I and II are 44.6% and 57.9% respectively. Therefore linear regression models may be ineffective for the given datasets, especially for swell percent determination.
- GEP analysis results in better performances than the performance of linear regression models. The values of MAPE for the best models of Dataset I, II, III, and IV for MBV prediction are 8.3%, 14.5%, 13.2%, and 46.3%,

respectively. The ones for Dataset I and II for the determination of swell percent are 11.5% and 22.1%, respectively.

- ANN analyses generate the best fitting models to the compiled database. The values of MAPE for the best models of Dataset I, II, III, and IV for MBV prediction are 4.2%, 5.0%, 11.5%, and 30.6%, respectively. The ones for the determination of swell percent for Dataset I, and II are 1.8% and 20.7%, respectively.
- The analysis results show that as the data in the database grows, the performances of predictive models gets worse. The worst analysis results are obtained from the most comprehensive dataset IV, consisting of 343 data points. Which suggests that the models lack other input/inputs that current database does not encounter for.
- Machine learning algorithms product complex input-output correlations where the models created with them will be available online, providing time efficient solutions for swell percent determinations.
- Even though models using only Atterberg limits and clay content perform well, when maximum dry density and optimum water content are used additively, MBV prediction models performances increase.

5.3.Future Studies

Our study emphasizes some missing aspects that can hopefully be highlighted in the future studies.

- Methylene blue test lacks a standard procedure for determining MBV of clayey soils. Therefore a standard procedure may be developed in the future. To eliminate the operator errors of MB test results an automated system is needed.
- Temperature and climate are two of the factors that determine swell potential. New models that include temperature and climate factors as a part of inputs can be generated.

- As there are various factors affecting the accuracy of the data obtained from the laboratory experiments, fuzzy based models can be used to understand the relations of the available dataset and predict swell percent and MBVs.
- Nonlinear mathematical regression models can also be developed in addition to machine learning models. It is expected that the performances can reach to those of machine learning based ones at most.

REFERENCES

(ANFOR), A. f. d. N. (1993). Mesure de la quantité et de l'activité de la fraction argileuse. La Défense, Paris, France: pp. 68-94

Abidin Erguler, Z. and R. Ulusay (2003). "A simple test and predictive models for assessing swell potential of Ankara (Turkey) Clay." Engineering Geology 67(3): 331-352.

Aringheri, R., Pardini, G., Gispert, M. (1992) Testing a simple methylene blue method for surface area estimation in soils. Agrochimica Vol. XXXVI-N.3, Agrochimica, Italy

ASTM Standard D422, 2014e2, "Standard Test Method for Particle-Size Analysis of Soils" ASTM International, West Conshohocken, PA, 2014, DOI: 10.1520/D422 E02, www.astm.org.

ASTM Standard .D854, 2014e1, "Standard Test Method for Specific Gravity of Soil Solids by Pycnometer" ASTM International, West Conshohocken, PA, 2014, DOI: 10.1520/D854 , www.astm.org.

ASTM Standard D4318, 2014e1, "Standar Tes Methods for Liquid Limit, Plastic Limit and Plasticity Index of Soils" ASTM International, West Conshohocken, PA, 2014, DOI: 10.1520/D4318 E01, www.astm.org.

Brindley, G. and T. Thompson (1970). "Methylene Blue Absorption by Montmorillonites. Determinations of Surface Areas and Exchange Capacities with Different Initial Cation Saturations (Clay-Organic Studies XIX)." Israel Journal of Chemistry 8(3): 409-415.

Brooks, C. (1964). "Mechanism of Methylene Blue Dye Adsorption on Siliceous Minerals." Kolloid-Zeitschrift und Zeitschrift für Polymere 199(1): 31-36.

Chen, F. H. (1975,1988). Foundations on Expansive Soils. Amsterdam; Newyork, Elsevier.

Chiappone, A., Marello, S., Scavia, C., & Setti, M. (2004). Clay mineral characterization through the methylene blue test: comparison with other experimental techniques and applications of the method. Canadian geotechnical journal, 41(6), 1168-1178.

Claudia, M. (2000). Predicting Swelling/Shrinkage Potential Using the Methylene Blue Method: Some Examples in Italian Clayey Soils. ISRM International Symposium, International Society for Rock Mechanics.

Cokca, E. (1991). "Swelling Potential of Expansive Soils with a Critical Appraisal of the Identification of Swelling of Ankara Soils by Methylene Blue Tests." Fen Bilimleri Enstitüsü: 325.

Cokca, E. and A. Birand (1993). "Determination of Cation Exchange Capacity of Clayey Soils by the Methylene Blue Test." Geotechnical Testing Journal 16: 518-518.

Craig, R. F. (2004). Craig's soil mechanics, CRC Press.

Çokça, E. (2002). "Relationship between Methylene Blue Value, Initial Soil Suction and Swell Percent of Expansive Soils." Turkish Journal of Engineering and Environmental Sciences 26: 521-529.

Das, B. M. (2013). Advanced soil mechanics, CRC Press.

DiMillio, A. F. (1999). A Quarter Century of Geotechnical Research.

Emodi, B. S. (1949). "The Adsorption of Dyestuffs by Montmorillonite." Clay Minerals Bulletin 3: 76-79.

Ergüler, Z. A. (2001). An Investigation on the Swelling Behavior of the Ankara Clay and Effect of Disturbance on Swelling Determination of Swelling Potential by Empirical Relations. Geological Engineering. Ankara, Hacettepe University. M. Sc.

Fairbairn, P. and R. H. Robertson (1957). "Liquid Limit and Dye Adsorption." Clay Minerals Bulletin 3: 129-136.

Ferreira, C. (2001). "Gene Expression Programming: A New Adaptive Algorithm for Solving Problems." Complex Systems 13(2): 42.

Fityus, S. G., Smith, D. W., & Jennar, A. M. (2000, November). “Surface Area Using Methylene Blue Adsorption As A Measure Of Soil Expansivity”, ISRM International Symposium. International Society for Rock Mechanics.

Graupe, D. (2007). Principles of Artificial Neural Net, World Scientific.

Hang, P. T. and G. Brindley (1970). "Methylene Blue Absorption by Clay Minerals. Determination of Surface Areas and Cation Exchange Capacities (Clay-Organic studies XVIII)." Clays and clay minerals 18(4): 203-212.

Haykin, S. S. (2009). “Neural networks and learning machines”, Pearson Education Upper Saddle River.

Hergül, T. (2012). “An Experimental study on the treatment of expansive soils by granular materials”, Ankara, METU. Ph.D.

Hills, J. and G. Pettifer (1985). "The Clay Mineral Content of Various Rock Types Compared with the Methylene Blue Value." Journal of Chemical Technology and Biotechnology. Chemical Technology 35(4): 168-180.

Karahan, F. (1999). “Relationship Between the Methylene Blue Test and Swelling Percentage, Swell Pressure and Suction of an Expansive Soil”, Civil Engineering Ankara, METU. M. Sc.

Koza, J. R. (1995). "Survey of Genetic Algorithms and Genetic Programming. Wescon Conference Record, Western Periodicals Company.

Koza, J. R. and R. Poli (2005). Genetic programming. Search Methodologies, Springer: 127-164.

Kriesel, D. (2007). "A brief introduction to neural networks." Retrieved August 15: 2011.

Lambe, T. W. and R. V. Whitman (1969). "Soil Mechanics, Series in Soil Engineering." John Wiley & Sons.

Li, J., Cameron, D. A., & Ren, G. (2014). Case study and back analysis of a residential building damaged by expansive soils. Computers and Geotechnics,56, 89-99.

Locat, J., Lefebvre, G., & Ballivy, G. (1984). Mineralogy, chemistry, and physical properties interrelationships of some sensitive clays from Eastern Canada. Canadian Geotechnical Journal, 21(3), 530-540.

Mitchell, J. K. and K. Soga (1976). Fundamentals of Soil Behavior, Wiley New York.

Nelson, J. and D. Miller (1992). "Expansive Soils." Willy, New York.

Nikolaides, A., Manthos, E., & Sarafidou, M. (2007). Sand equivalent and methylene blue value of aggregates for highway engineering. Foundations of civil and environmental engineering, (10), 111-121.

Outhwaite, J. G. and A. D. Morgan (1972). "Methylene-Blue Test for Determination of Active Clay Content of a Green Moulding Sand " Foundry Trade Journal: 7.

Ozer, M., Ulusay, R., & Isik, N. S. (2012). Evaluation of damage to light structures erected on a fill material rich in expansive soil. Bulletin of Engineering Geology and the Environment, 71(1), 21-36.

Poli, R. and J. Koza (2014). Genetic Programming, Springer.

Euliano, N., Principe, J., & Lefebvre, W. C. (1999). "Neural and Adaptive Systems: Fundamentals Through Simulations " John Wiley and Sons, New York 119: 656.

Raj, P. P. (2008). Soil Mechanics & Foundation Engineering, Pearson Education India.

Robertson, R. H. and R. Ward (1951). "The Assay of Pharmaceutical Clays." Journal of Pharmacy and Pharmacology 3(1): 27-35.

Terzaghi, K. (1996). Soil Mechanics in Engineering Practice, John Wiley & Sons.

Türköz, M. (2007). Determination of Swelling Potential of High Plasticity Clays with Direct Tests and a Comparative Study. Civil Engineering. Eskişehir, Eskişehir Osmangazi University Phd.

Velde, B. (1995). Origin and Mineralogy of Clays: Clays and the Environment, Springer-Verlag.

Walker, M. (2001). "Introduction to Genetic Programming." Tech. Np: University of Montana.

Worrall, W. (1958). "Adsorption of Basic Dyestuffs by Clays." Trans. Brit. Ceram. Soc 57: 210-217.

Yenes, M., Nespereira, J., Blanco, J. A., Suárez, M., Monterrubio, S., & Iglesias, C. (2012). Shallow foundations on expansive soils: a case study of the El Viso Geotechnical Unit, Salamanca, Spain. Bulletin of Engineering Geology and the Environment, 71(1), 51-59.

Yukselen, Y. and A. Kaya (2008). "Suitability of the methylene blue test for surface area, cation exchange capacity and swell potential determination of clayey soils." Engineering Geology 102(1): 38-45.

APPENDIX A

EXPERIMENT RESULTS

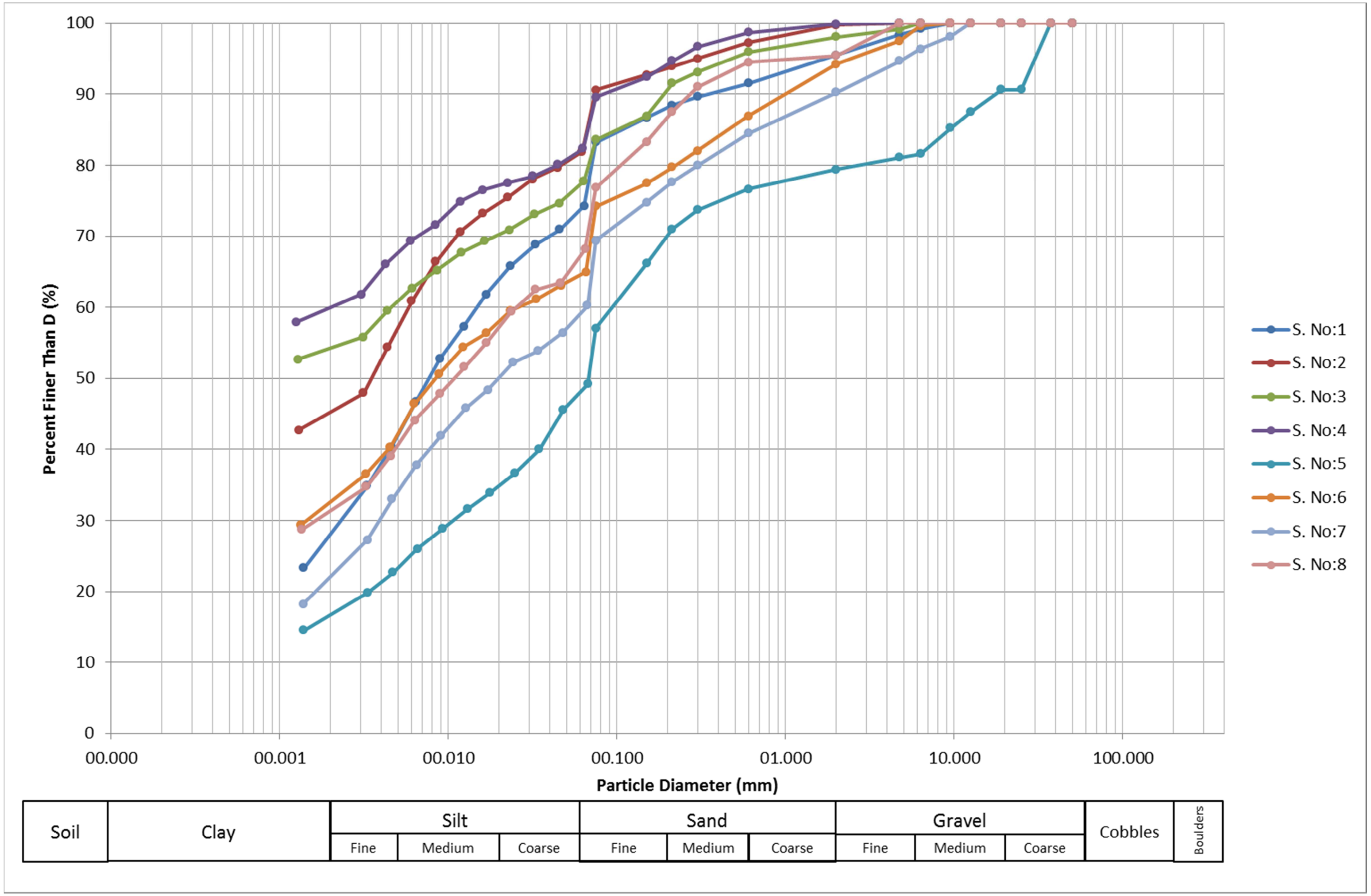


Figure A.1. Sieve Analyses and Hydrometer Test Results (1/4)

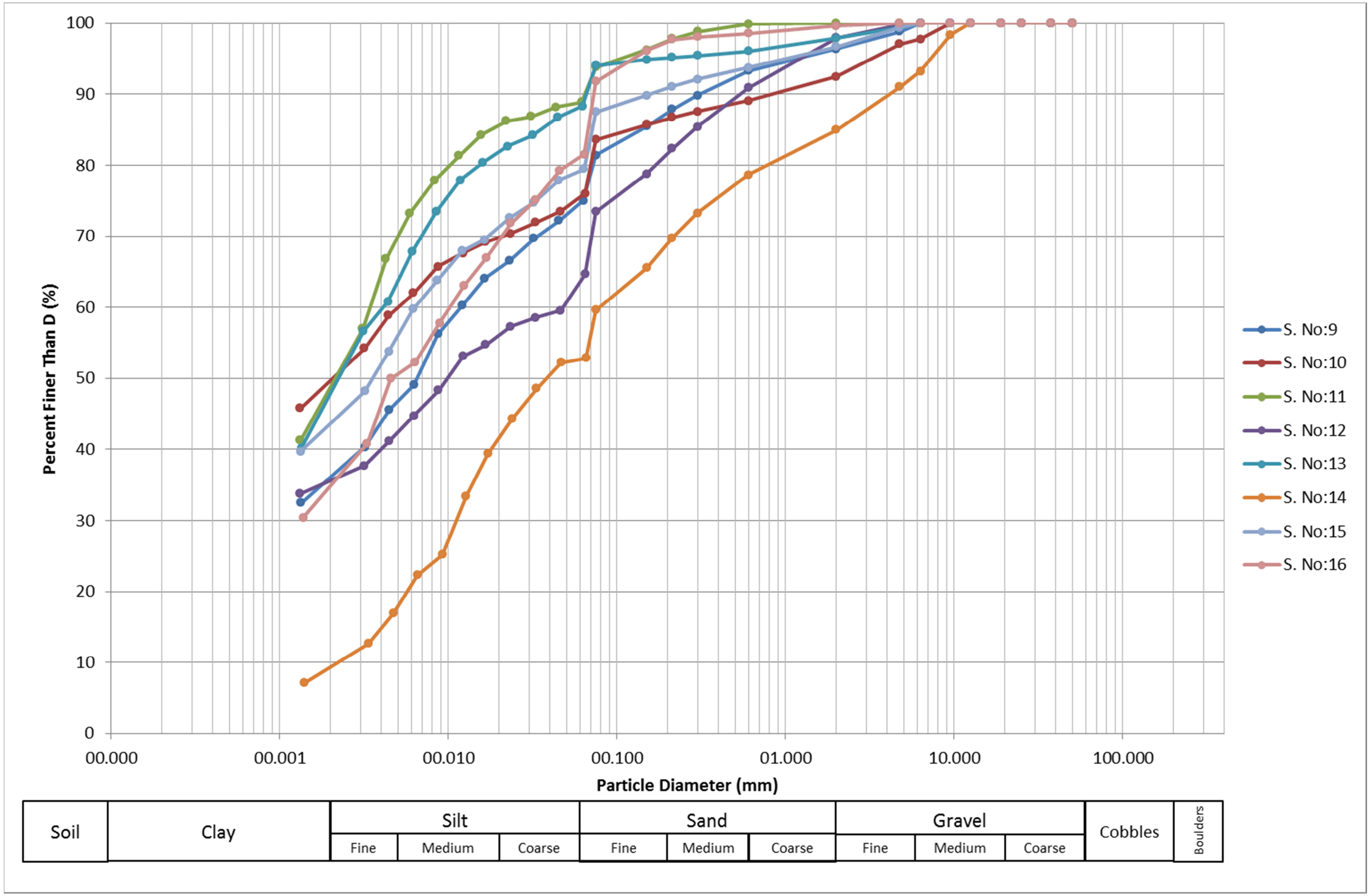


Figure A.2. Sieve Analyses and Hydrometer Test Results (2/4)

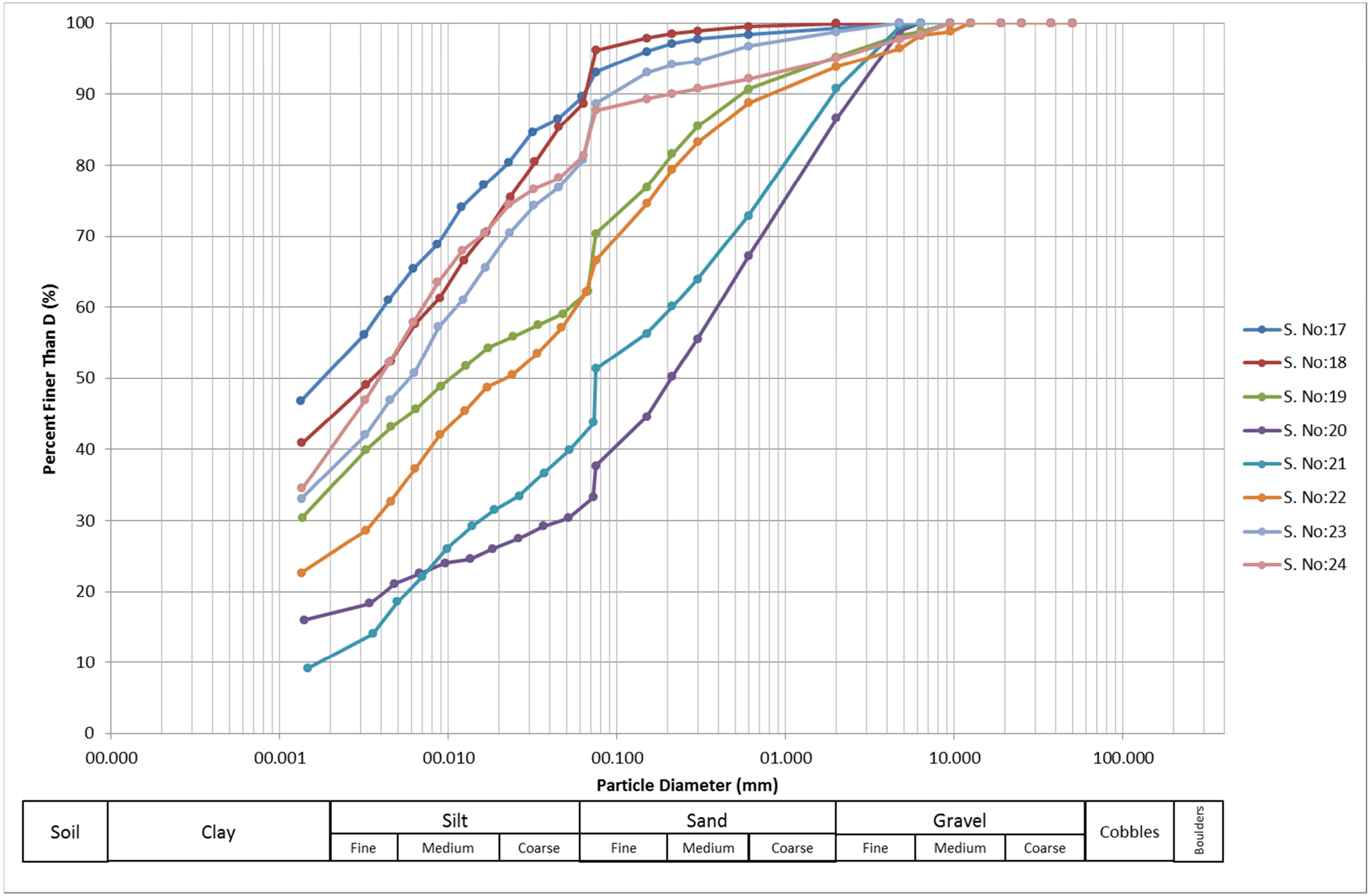


Figure A.3. Sieve Analyses and Hydrometer Test Results (3/4)

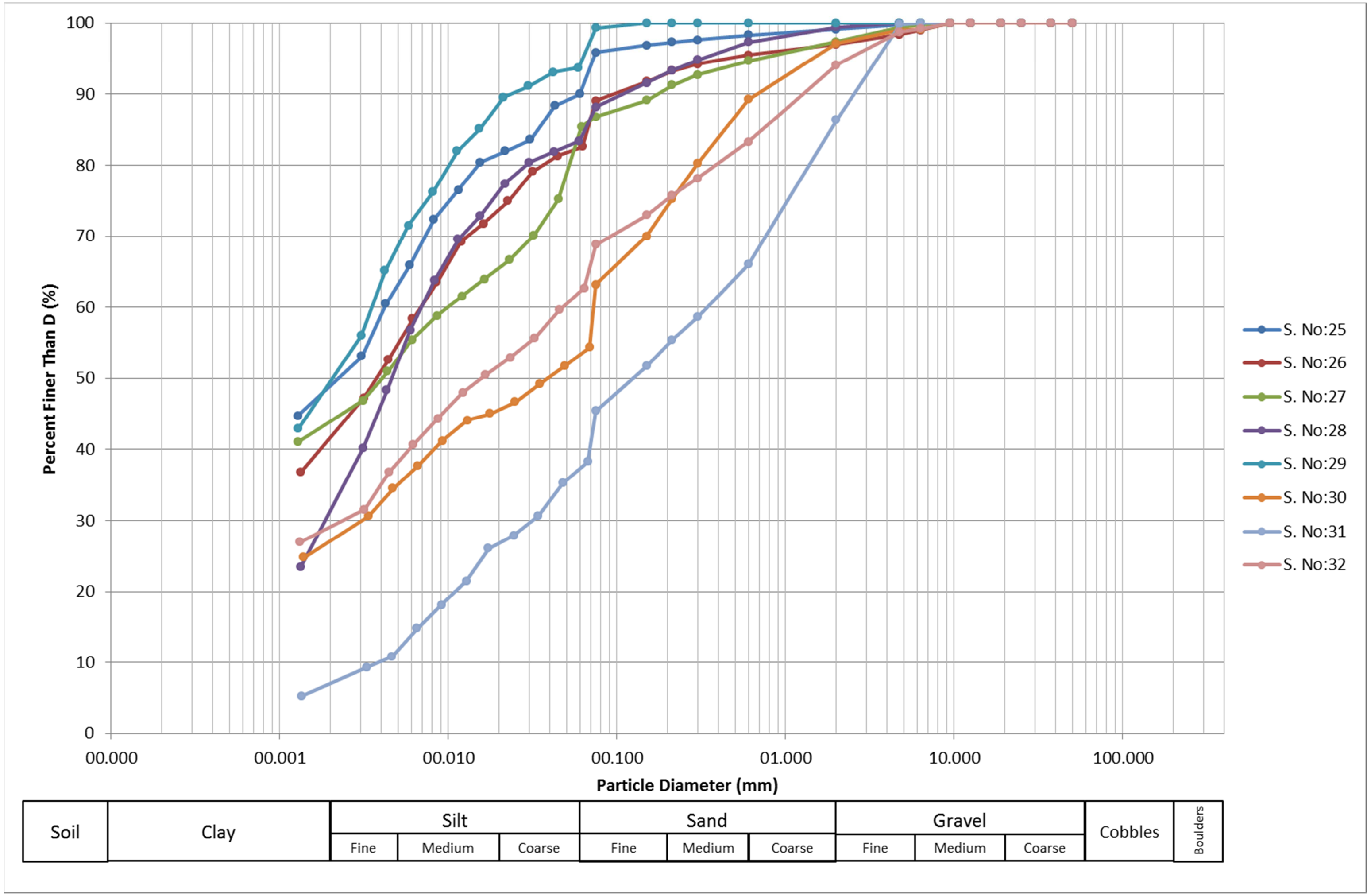


Figure A.4. Sieve Analyses and Hydrometer Test Results (4/4)

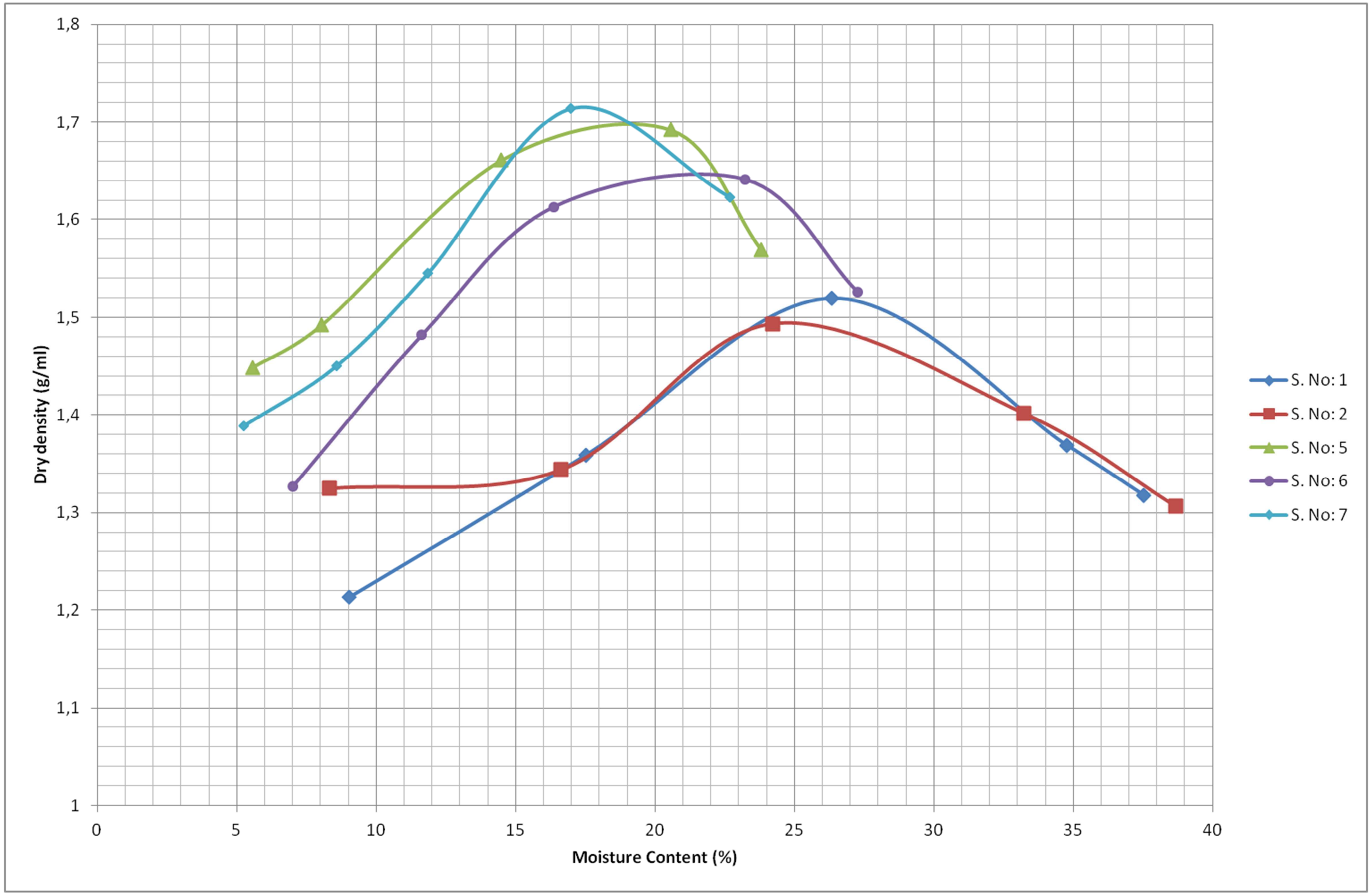


Figure A.5. Dry Density vs. Moisture Content Graph (1/4)

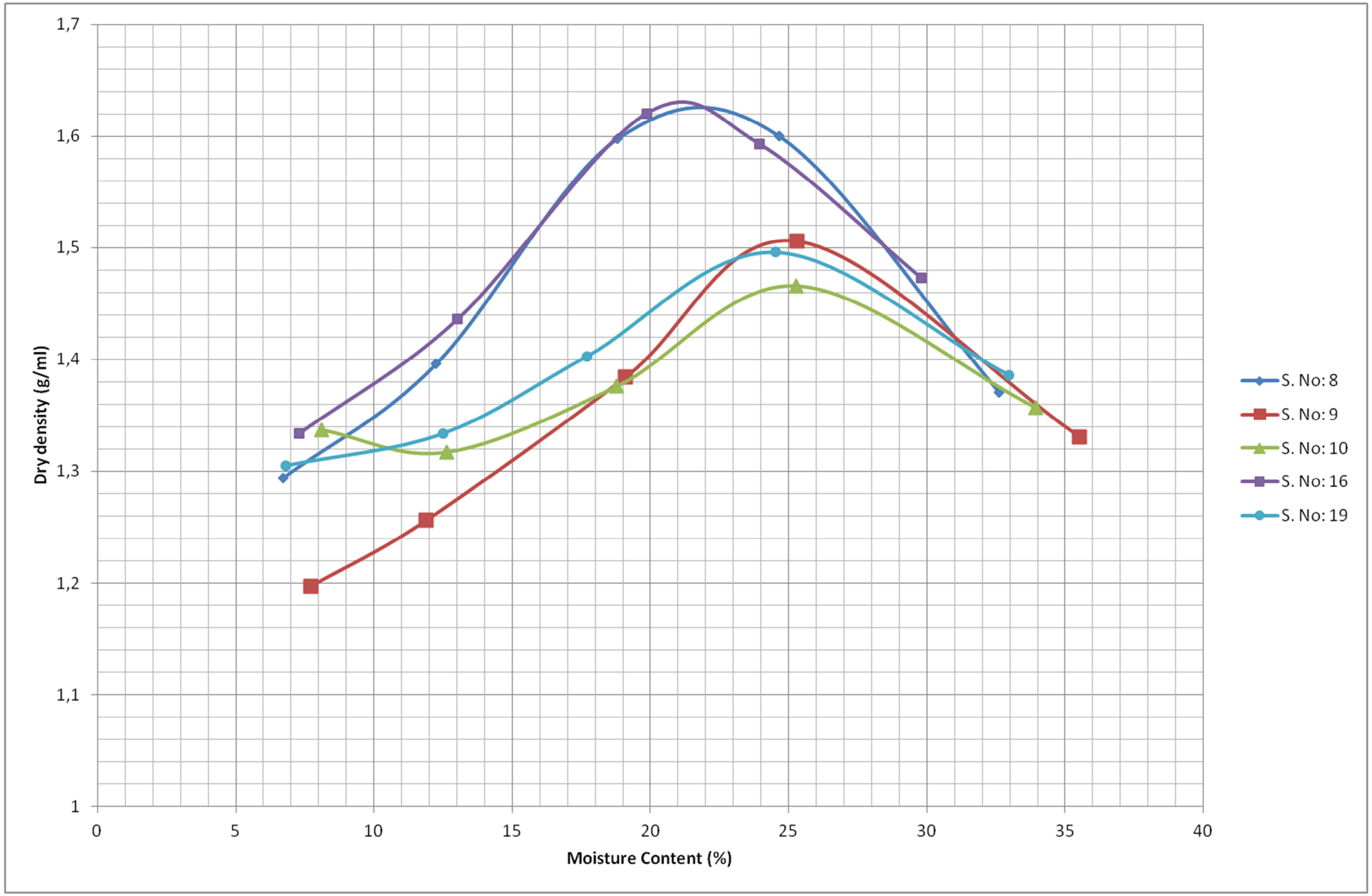


Figure A.6. Dry Density vs. Moisture Content Graph (2/4)

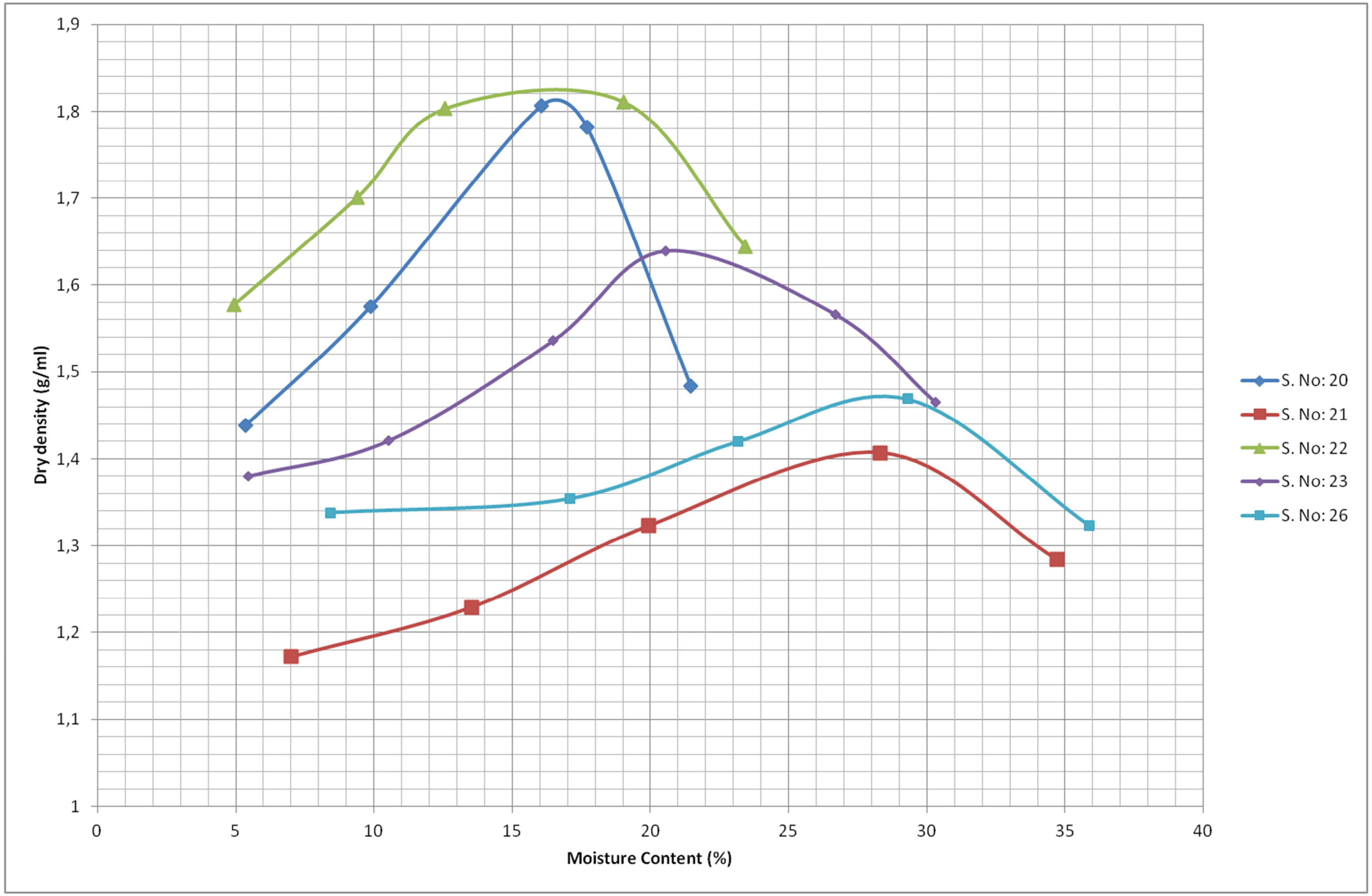


Figure A.7. Dry Density vs. Moisture Content Graph (3/4)

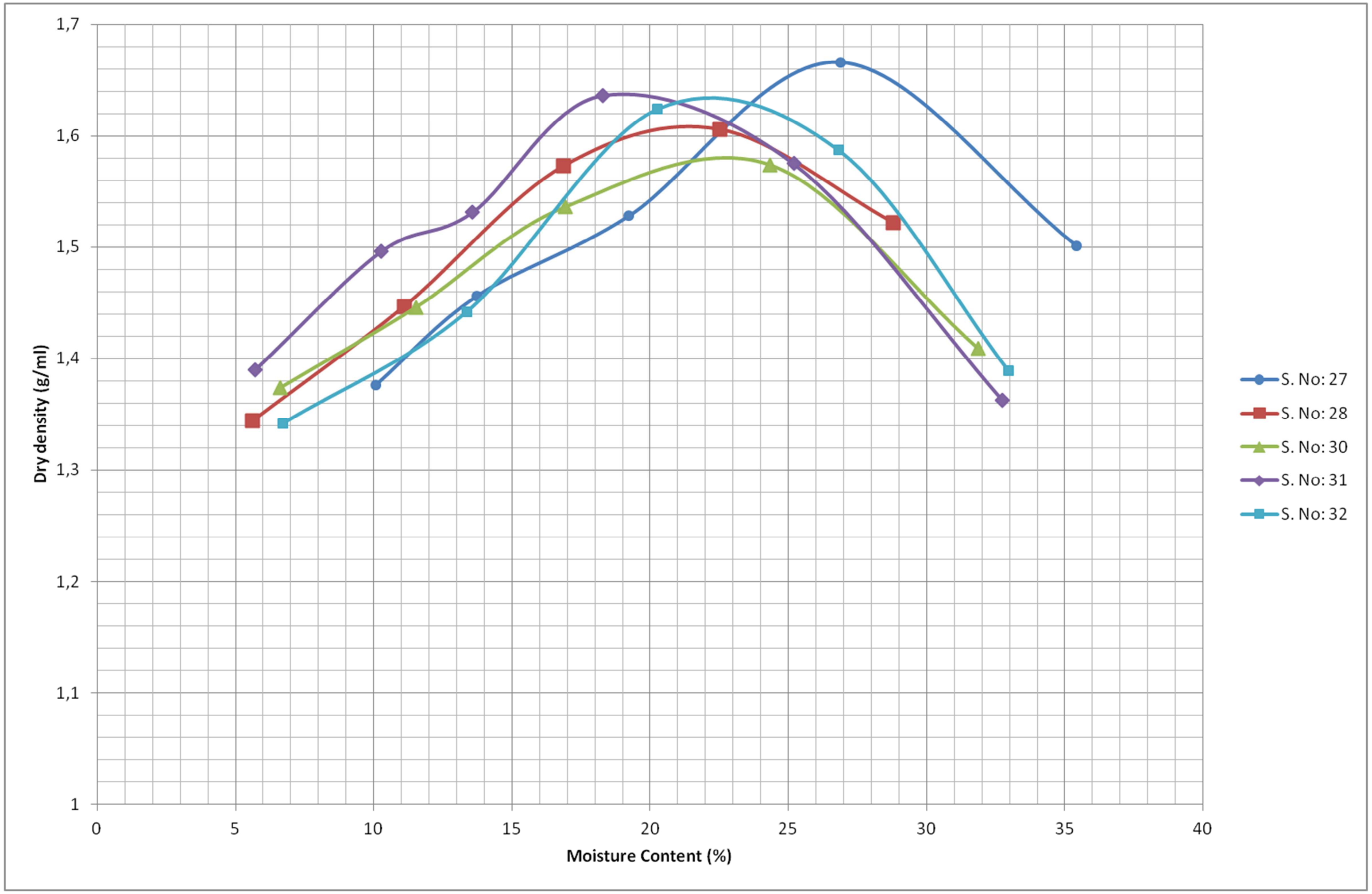


Figure A.8. Dry Density vs. Moisture Content Graph (4/4)

APPENDIX B

LINEAR REGRESSION RESULTS

5.4.B.1. Dataset I, MBV Prediction:

REG 3

```
.  reg MBV PI CC, level (85) robust

Linear regression                          Number of obs =      32
                                           F(  2,    29) =   58.25
                                           Prob > F      =  0.0000
                                           R-squared     =  0.7957
                                           Root MSE      =  1.4566

------------------------------------------------------------------------------
             |               Robust
         MBV |      Coef.   Std. Err.      t    P>|t|     [85% Conf. Interval]
-------------+----------------------------------------------------------------
          PI |   .0962497   .0236067     4.08   0.000     .0613424     .131157
          CC |   .1079267   .0219808     4.91   0.000     .0754236    .1404298
       _cons |   1.254334   .6869504     1.83   0.078     .2385375    2.270131
------------------------------------------------------------------------------

.        vif

    Variable |       VIF       1/VIF
-------------+----------------------
          CC |      1.67    0.598417
          PI |      1.67    0.598417
```

```
-------------+----------------------
    Mean VIF |      1.67

.       swilk MBV PI CC

                   Shapiro-Wilk W test for normal data

    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
         MBV |     32    0.97745      0.752    -0.591    0.72278
          PI |     32    0.86877      4.378     3.065    0.00109
          CC |     32    0.96398      1.201     0.381    0.35163
```

5.5.B.2. Dataset II, MBV Prediction:

REG 23

```
. reg MBV LL PL wopt, level(85) robust


Linear regression                               Number of obs =      73
                                                F(  3,    69) =   65.06
                                                Prob > F      =  0.0000
                                                R-squared     =  0.6814
                                                Root MSE      =  1.2315
------------------------------------------------------------------------------
             |               Robust
         MBV |      Coef.   Std. Err.      t    P>|t|     [85% Conf. Interval]
-------------+----------------------------------------------------------------
          LL |   .2000173   .0168855    11.85   0.000     .1754364    .2245982
```

```
          PL |  -.2289242   .0542422    -4.22   0.000    -.3078866   -.1499617
        wopt |   .0127079   .0081173     1.57   0.122     .0008912    .0245245
       _cons |   1.696933   1.033518     1.64   0.105     .1924006    3.201466
------------------------------------------------------------------------------
.       vif

    Variable |       VIF       1/VIF
-------------+----------------------
          PL |      3.42    0.292316
          LL |      2.62    0.381291
        wopt |      1.74    0.573867
-------------+----------------------
    Mean VIF |      2.60


.       swilk MBV LL PL wopt
                   Shapiro-Wilk W test for normal data
    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
         MBV |     73    0.98188      1.154     0.312    0.37743
          LL |     73    0.97141      1.821     1.306    0.09576
          PL |     73    0.96496      2.232     1.750    0.04006
        wopt |     73    0.69589     19.368     6.460    0.00000
```

5.6.B.3. Dataset III, MBV Prediction:

REG 3

```
. reg MBV PI CC, level (85) robust

Linear regression                    Number of obs =     125
                                     F(  2,   122) =   47.60
                                     Prob > F      =  0.0000
                                     R-squared     =  0.5046
                                     Root MSE      =  1.7683

------------------------------------------------------------------------------
             |               Robust
         MBV |      Coef.   Std. Err.      t    P>|t|     [85% Conf. Interval]
-------------+----------------------------------------------------------------
          PI |   .1142464   .0162389     7.04   0.000     .0907219    .1377709
          CC |   .0226243   .0125502     1.80   0.074     .0044434    .0408053
       _cons |    2.61919   .4754603     5.51   0.000     1.930413    3.307967
------------------------------------------------------------------------------

.        vif

    Variable |       VIF       1/VIF
-------------+----------------------
          CC |      1.69    0.591436
          PI |      1.69    0.591436
-------------+----------------------
    Mean VIF |      1.69

.        swilk MBV PI CC
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
MBV	125	0.97360	2.630	2.171	0.01496
PI	125	0.80476	19.448	6.663	0.00000
CC	125	0.98519	1.476	0.874	0.19117

5.7.B.4. Dataset IV, MBV Prediction::

REG 1

. reg MBV LL PL, level(85) robust

Linear regression   Number of obs =   342

F(2,   339) = 103.70

Prob > F   = 0.0000

R-squared   = 0.3782

Root MSE   = 2.5193

MBV	Coef.	Robust Std. Err.	t	P>|t|	[85% Conf. Interval]	
LL	.1588111	.0128599	12.35	0.000	.1402568	.1773654
PL	-.2041947	.0361478	-5.65	0.000	-.2563487	-.1520407
_cons	1.414536	.5580012	2.54	0.012	.6094517	2.21962

```
.     vif

   Variable |       VIF       1/VIF

-------------+----------------------

         LL |      4.01    0.249592

         PL |      4.01    0.249592

-------------+----------------------

   Mean VIF |      4.01

.     swilk MBV LL PL

                  Shapiro-Wilk W test for normal data

   Variable |    Obs       W           V         z       Prob>z

-------------+--------------------------------------------------

        MBV |    342    0.94705     12.674     5.999    0.00000

         LL |    342    0.97185      6.738     4.506    0.00000

         PL |    342    0.96091      9.357     5.282    0.00000
```

5.8.B.5. Dataset I, Swell Percent Prediction:

REG 8

```
.  reg Sper A LL PL DW MBV, level(85) robust



Linear regression                               Number of obs =      20

                                          F(  5,    14) =   16.38

                                          Prob > F      =  0.0000

                                          R-squared     =  0.8425

                                          Root MSE      =  1.9008

------------------------------------------------------------------------------
```

```
             |               Robust
        Sper |      Coef.   Std. Err.      t    P>|t|     [85% Conf. Interval]
-------------+----------------------------------------------------------------
           A |   2.771119   1.357203     2.04   0.060       .70397    4.838268
          LL |   .4876147    .093127     5.24   0.000     .3457734    .6294559
          PL |  -.5225423   .1655015    -3.16   0.007    -.7746169   -.2704677
          DW |   12.81028   6.728957     1.90   0.078     2.561438    23.05912
         MBV |  -.5590666   .3657077    -1.53   0.149    -1.116074    -.002059
       _cons |  -27.79408   13.35517    -2.08   0.056    -48.13527   -7.452892
------------------------------------------------------------------------------

. vif

    Variable |       VIF       1/VIF
-------------+----------------------
          LL |      7.77    0.128692
         MBV |      4.83    0.206868
          PL |      4.43    0.225858
          DW |      2.78    0.359372
           A |      1.41    0.711400
-------------+----------------------
    Mean VIF |      4.24

. swilk Sper A LL PL DW MBV
```

```
                   Shapiro-Wilk W test for normal data

    Variable |    Obs       W           V         z       Prob>z
-------------+-------------------------------------------------
        Sper |     20    0.64230      8.467     4.305    0.00001
           A |     20    0.78627      5.059     3.267    0.00054
          LL |     20    0.93355      1.573     0.913    0.18070
          PL |     20    0.96524      0.823    -0.393    0.65283
          DW |     20    0.97247      0.652    -0.863    0.80588
         MBV |     20    0.94687      1.258     0.462    0.32207
```

5.9.B.6. DataSet Iı, Swell Percent Prediction:

REG 28

```
. reg Sper LL DW wopt MBV, level(85) robust

Linear regression                               Number of obs =      73
                                                F(  4,    68) =   24.35
                                                Prob > F      =  0.0000
                                                R-squared     =  0.7887
                                                Root MSE      =  3.7998

------------------------------------------------------------------------------
             |               Robust
        Sper |      Coef.   Std. Err.      t    P>|t|     [85% Conf. Interval]
-------------+----------------------------------------------------------------
          LL |   .1080296   .0700903     1.54   0.128     .0059796    .2100796
          DW |   8.300297   5.504735     1.51   0.136     .2855164    16.31508
```

```
     wopt |   .2188653   .0293519     7.46   0.000     .1761295     .261601
      MBV |   .6734742   .3035222     2.22   0.030     .2315521    1.115396
    _cons |  -25.87537   9.655891    -2.68   0.009    -39.93415   -11.81659
------------------------------------------------------------------------------
.       vif
    Variable |       VIF       1/VIF
-------------+----------------------
          LL |      2.89    0.346115
         MBV |      2.36    0.423196
        wopt |      1.36    0.736515
          DW |      1.09    0.916518
-------------+----------------------
    Mean VIF |      1.93
.       swilk Sper LL DW wopt MBV

                   Shapiro-Wilk W test for normal data

    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
        Sper |     73    0.65567     21.930     6.731    0.00000
          LL |     73    0.97141      1.821     1.306    0.09576
          DW |     73    0.92698      4.651     3.350    0.00040
        wopt |     73    0.69589     19.368     6.460    0.00000
         MBV |     73    0.98188      1.154     0.312    0.37743
```

APPENDIX C

GENETIC EXPRESSION PROGRAMMING AND KARVA LANGUAGE PROGAM CODES

5.10. C.1.Data Set I, MBV Prediction, Run1 Analysis Output Karva Language

Tan.c1.Acot.X3.d1.c3.d1.d1.d1.c5.d1.d0.d0.d1.d2.d1.d0.d1.d2.c0.d1.c4.d1.d0.d0.c8.d1.c3.d0.c1.d2.d1.d1.d1.d0.d0.c9.d1.c6.d2.d0

+

Coth.c9.Cosh.c1.d0.Csch.Add3.*.Sec.Add3.d1.d1.d1.d0.d0.c8.d2.d1.d1.d1.d0.d0.d0.d2.d0.c4.d2.d0.d2.d1.c6.d1.c4.d1.c6.d2.d0.d2.c0.d1.c8

+

Tan.c1.Add4.c4.d1.d2.Mul3.d2.*.X5.d1.d2.d1.c0.d1.c3.c8.d2.d1.d0.d1.c8.d2.d0.d0.c8.d0.d1.c8.d0.d2.d2.d2.d2.d1.d1.c6.d1.c6.d1.c9

+

Log.Cosh./.*.+.c9.d2.Cot.-.c4.c9.d2.c8.d2.c9.d1.c5.d1.d2.d2.d2.d2.c6.d1.c5.d1.d1.c2.c1.d2.d2.d2.c2.d0.d2.d0.d2.c8.c6.d1.c0

+

E.Sin.d0.-
.d2.Tanh.d0.Sinh.Mul4.d2.d2.c4.d0.d1.d1.d2.c2.d0.d2.d0.d0.c2.d2.d1.d1.c2.c2.d2.c6.d1.d0.d0.d2.d0.c6.d1.d2.d0.d0.d1.d2

+

Coth.-.+.+.d0.+.Abs.c4.+.Add3.d2.c0.d2.d2.c2.d1.d2.c4.d0.c7.d1.d1.d1.d0.d1.d2.d0.c5.c3.c2.d2.c7.d2.c5.d1.d2.c8.c0.c3.d2.c6

+

Atan.d1.Cos.c0.c1.d2.Neg.Csch.d2.d1.d2.d0.d2.d2.c2.c3.c8.d1.c4.d2.c3.d0.d0.d2.d1.d2.d1.d0.c1.d1.d0.d1.d0.c7.c7.d2.d2.c6.c0.d1.d0

+

Log.-.d1.c8.3Rt.Sech.Sech.E.*.-.d1.d2.c7.d1.d2.d2.c3.d2.d1.d1.c2.d0.d0.c0.d1.d0.d0.d1.d0.c0.d0.d0.c5.d0.c0.d1.d0.d0.c2.c1.d1

+

Log.-.d1.c8.3Rt.Pi.Sech.d0.+.d2.d2.c2.d0.c4.d2.d2.c1.c9.d0.d1.d0.c0.c1.c7.d0.d2.d0.d1.d2.c1.d0.d2.c1.d2.c9.d1.c6.d0.d2.c0.c4

+

E.d1.Exp.Sin.-
.Atan.d2.X2.X4.c7.d1.c6.d0.d0.d1.d0.d0.d2.d2.d0.d2.d1.c1.d1.d0.d0.c9.c9.c0.c7.c5.d0.c6.d2.d0.d2.c3.c1.c9.d1.d0

+

Mul4.c9.c4.Csch.Div4.Mul4.d2.c1.c8.+.c3.d1.d0.c2.c7.d2.c0.d0.d2.c7.d2.d1.c8.c5.d0.c9.c1.d1.c4.d2.d2.d1.c3.c4.c4.c3.c7.d1.c5
.d1.d2

Numerical Constants:

Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
c0 = 0.465118	c0 = -2.08728	c0 = 0.049499	c0 = -9.521423	c0 = 7.532166
c1 = 8.69928	c1 = -0.816437	c1 = -1.304046	c1 = 9.1091	c1 = 6.452209
c2 = -2.919434	c2 = -3.552124	c2 = -9.914063	c2 = 4.253692	c2 = -3.397919
c3 = -6.934479	c3 = -9.043609	c3 = 8.037139	c3 = -3.768036	c3 = -2.674072
c4 = -2.283844	c4 = -4.226165	c4 = 5.765351	c4 = -2.204651	c4 = 8.512573
c5 = 3.65213	c5 = -4.388031	c5 = 5.008728	c5 = -9.07608	c5 = -6.029174
c6 = -0.516846	c6 = -0.223388	c6 = -3.519928	c6 = 6.674774	c6 = 9.010803
c7 = 3.565918	c7 = -3.782715	c7 = 9.640747	c7 = -7.794464	c7 = 0.295135
c8 = 9.780915	c8 = -8.240997	c8 = 7.134063	c8 = -6.204132	c8 = 8.892456
c9 = -3.299774	c9 = -9.656891	c9 = 3.915253	c9 = -1.013488	c9 = 8.402313
Gene 6	Gene 7	Gene 8	Gene 9	Gene 10
c0 = -3.628174	c0 = 0.495392	c0 = -0.683167	c0 = 7.961396	c0 = -9.524536
c1 = 3.227722	c1 = 7.758209	c1 = 0.533386	c1 = -1.016967	c1 = -3.848786
c2 = 1.813019	c2 = -0.241211	c2 = -8.459991	c2 = -7.931061	c2 = -9.914063
c3 = 7.268677	c3 = 4.034881	c3 = -5.631439	c3 = -5.631439	c3 = 9.492401
c4 = 4.654755	c4 = 9.010803	c4 = 3.162109	c4 = -9.6203	c4 = 0.52356
c5 = -5.380859	c5 = 5.160186	c5 = 0.572205	c5 = 0.572205	c5 = -5.318512
c6 = -5.900788	c6 = -4.287506	c6 = -0.271057	c6 = 9.513367	c6 = 6.196014
c7 = -9.21344	c7 = -8.943329	c7 = -5.570648	c7 = -1.983918	c7 = -3.982514
c8 = -6.639282	c8 = 8.512573	c8 = -2.707764	c8 = -2.707764	c8 = -9.559631

c9 = -5.31604	c9 = 3.047424	c9 = -3.299682	c9 = -4.267975	c9 = -9.621491

Gene 11

c0 = -8.925415	c3 = 6.934418	c6 = -8.191711	c9 = 3.388031
c1 = -1.168823	c4 = -6.019531	c7 = 5.818451	
c2 = -3.531494	c5 = -0.151489	c8 = 1.61615	

5.11. C.2. Data Set II, MBV Prediction, Run1 Analysis Output Karva Language

d0.d0.Div4.d2.d0.d4.3Rt.d4.d3.c1.d1.c8.d0.c2.c2.d0.c7.d2.d0.d2.c4.d1.d4.d3.d2.d0.d0.d4.d2

+

E.d3.d1.c8.d4.d4.d4.d3.d0.d3.d4.d4.c1.d1.d3.d1.c2.d4.d1.d4.d4.c4.d4.d3.d4.d0.d1.d4.c5

+

Sqrt.X4.Csch.5Rt.Sqrt.Abs.+.c7.d1.d1.d2.d3.d0.c2.d1.d0.d0.c3.c8.d3.c2.c0.c5.d0.c6.d0.d3.d2.c8

+

Sinh.Sinh.5Rt.Csch./.-.d1.c6.d2.d4.d1.c2.c0.c5.d1.d0.c9.d2.c6.d3.d1.c2.c9.d3.d2.c6.d3.c6.c9

+

Sec.4Rt.Cosh.Pi.X4.X5.*.d0.d0.c7.d1.c7.d4.d4.d0.c2.c5.d2.c6.d3.d0.c2.c1.d3.c7.c2.d3.c1.d1

Numerical Constants:

Gene 1	Gene 2	c0 = -3.907319	c1 = 6.610657	Gene 5
c0 = -8.541504	c0 = 3.309784	c1 = -3.695801	c2 = 9.218017	c0 = 3.30978
c1 = -3.985138	c1 = -8.94983	c2 = -1.720856	c3 = 2.481476	c1 = -9.43143
c2 = 4.482513	c2 = -3.656098	c3 = -4.633423	c4 = -3.907319	c2 = 2.44846
c3 = 7.253723	c3 = -4.242187	c4 = 7.489349	c5 = 1.76355	c3 = -4.24219
c4 = 0.484772	c4 = -3.861817	c5 = 3.831055	c6 = -6.177032	c4 = -3.77295
c5 = 2.345367	c5 = 3.831055	c6 = 5.747132	c7 = -6.768524	c5 = 3.83106
c6 = -6.825562	c6 = -3.157623	c7 = 1.112518	c8 = 8.97617	c6 = -3.15762
c7 = -6.230774	c7 = 0.525696	c8 = 9.651978	c9 = 9.30829	c7 = -8.36145
c8 = 8.433868	c8 = 6.270538	c9 = -1.726349		c8 = -4.89713
c9 = 3.970002	c9 = -6.233062			c9 = -2.42215
		Gene 4		
	Gene 3	c0 = 1.661713		

5.12. C.3. Data Set III, MBV Prediction, Run3 Analysis Output Karva Language

Neg.d1.Asec.d2.Asec.Sinh.d2.c2.d1.d2.c5.c4.d0.d0.c3.d0.c2.c3.d1.d1.c7.d1.d1.d1.c0.c6.d1.c7.d0

+

*.d1.Sec.+.Tanh.c8.d0.d1.d1.d2.d1.c7.d1.d2.d2.d2.d1.d1.c1.d0.d0.d1.d1.c6.d0.d2.d2.d0.c0

+

Sin./.Mul4.d0.d1.Sin.-.d0.d1.d2.d2.d0.d2.d0.c9.c9.d0.d1.d2.d1.d0.d0.d1.c2.c2.d0.d0.d1.d2

+

*.Sec.d1.-.c9.c4.Neg.c2.d0.d1.d2.c1.d2.d0.c2.c0.c2.d2.c2.d1.d2.d0.d1.c1.c4.c7.d2.c8.c6

+

d1.Sech.d0.Tanh.Atan.d0.c9.c4.d1.d1.d2.d2.d2.d0.d0.c2.c1.d2.c2.d1.d2.d0.d1.c0.c2.c2.d2.c1.d2

+

Sin./.Mul4.d0.d1.Sin.-.d0.d1.d2.d2.d1.c4.d1.d0.d1.d1.c9.d1.d1.c8.d2.d1.d1.c8.d1.c9.c2.d2

+

*.d1.Sec.Tan.E.Tanh.Log.c0.d2.d1.d1.c3.d0.d0.c8.d0.d0.c4.d1.d1.c7.d0.c7.d2.d0.d1.d0.d0.d2

+

Mul4.-.Tan./.Tan.Neg.4Rt.c8.c2.c3.d1.d2.c9.d0.d1.d0.c5.d1.d0.d1.d1.d2.d0.d2.d0.d1.d2.d0.d1

Numerical Constants:

Gene 1	Gene 2	Gene 3	Gene 4
c0 = -0.313751	c0 = -9.300293	c0 = 4.602448	c0 = -5.905396
c1 = 6.688782	c1 = 1.81958	c1 = 8.000335	c1 = 4.169677
c2 = 6.908417	c2 = -9.355957	c2 = -7.283203	c2 = -1.659638
c3 = -5.035553	c3 = -9.406525	c3 = 3.992737	c3 = 8.321777
c4 = 6.856812	c4 = 6.968995	c4 = 1.214874	c4 = -1.382141
c5 = 8.826569	c5 = -6.933777	c5 = -0.640289	c5 = 9.290588
c6 = -1.037109	c6 = -6.287384	c6 = -0.007354	c6 = -4.096618
c7 = 3.641297	c7 = -9.824097	c7 = -9.624756	c7 = -0.029937
c8 = -4.01416	c8 = -0.020111	c8 = 7.21643	c8 = -8.575073
c9 = 2.769013	c9 = 9.204406	c9 = 9.409485	c9 = -5.379974

Gene 5	Gene 6	Gene 7	Gene 8
c0 = -2.325867	c0 = 0.187164	c0 = 5.07788	c0 = -7.461487
c1 = -2.418487	c1 = 5.288727	c1 = 6.170623	c1 = -4.392029
c2 = 3.840729	c2 = 5.210846	c2 = -6.964417	c2 = -2.912598
c3 = 5.900605	c3 = 3.992737	c3 = 0.850952	c3 = -7.998749
c4 = 0.513214	c4 = 4.550415	c4 = 8.406952	c4 = -5.341247
c5 = 8.826569	c5 = -9.071014	c5 = -5.811829	c5 = -9.148254
c6 = -2.135528	c6 = -9.023773	c6 = 6.052978	c6 = -7.74115
c7 = 6.032867	c7 = -7.717255	c7 = -7.708435	c7 = 3.299804
c8 = 6.494232	c8 = 3.272491	c8 = 3.410492	c8 = -6.321594
c9 = -0.916474	c9 = 9.550903	c9 = -5.210968	c9 = 0.884735

5.13. C.4. Data Set IV, MBV Prediction, Run3 Analysis Output Karva Language

Asin.Atan.Sech.*.Sech.d0.d2.d1.d1.d2.d0.c7.c6.c8.c7.c9.d2.d1.d1.c1.d2

+

..Exp.-.Exp.c6.d0.c0.d2.d0.d1.d0.d1.c6.d1.d1.d2.c6.d0.c4.c4

+

Cos.Nop.-.Nop.c8.d1.c8.c9.d0.d0.c4.c0.d0.c0.c5.d1.d2.c5.d0.d2.d1

+

d1.d1.E.-.d1.d1.c6.c0.d2.d2.d1.c6.d1.c0.c4.d1.d1.d1.d2.c4.c5

+

Nop.Csch.Csch.Cos.d1.d1.c7.c7.d2.d0.d2.d1.d0.c2.d2.d2.d2.d1.d2.d0.d0

+

Csch.c5.d0.c4.c0.d2.d0.d0.c0.d0.d2.d1.c1.d1.d1.d0.c1.c8.d0.c1.d2

+

Csch.Sub4.c1.Div3.X4.c3.d0.d0.d2.d0.c9.d1.c6.c7.c7.d2.c9.c1.c1.c4.c9

Numerical Constants:

Gene 1	Gene 2	Gene 3	Gene 4
c0 = -6.575012	c0 = 2.667908	c0 = 0.691773	c0 = -1.649323
c1 = 3.318603	c1 = 3.965149	c1 = -7.633484	c1 = 5.250763
c2 = -8.74997	c2 = 3.016845	c2 = 6.413727	c2 = -8.04898
c3 = -9.858826	c3 = -1.113006	c3 = 1.268982	c3 = -6.571106
c4 = -1.873627	c4 = 8.199463	c4 = -6.42279	c4 = 5.317138
c5 = 3.019073	c5 = 5.209564	c5 = 9.51947	c5 = -8.196319
c6 = 1.098571	c6 = -3.265503	c6 = 8.669953	c6 = -3.675263
c7 = -5.628906	c7 = 1.540497	c7 = -3.433624	c7 = -0.869202
c8 = -5.4841	c8 = -9.672089	c8 = -8.939454	c8 = 9.25235
c9 = 1.402954	c9 = -4.545502	c9 = -2.688019	c9 = 2.669556

Gene 5	Gene 6	Gene 7
c0 = 2.913422	c0 = 7.169128	c0 = -1.313293
c1 = 9.155761	c1 = -1.29956	c1 = 9.738404
c2 = -6.628479	c2 = -4.955018	c2 = -1.191131
c3 = -4.809662	c3 = -4.571228	c3 = -3.695496
c4 = -7.616821	c4 = -4.053375	c4 = 1.937134
c5 = -3.72757	c5 = -1.304413	c5 = -1.529877
c6 = 7.280578	c6 = 2.587891	c6 = -0.865265
c7 = -3.045623	c7 = 0.529755	c7 = 7.912995
c8 = -7.556091	c8 = -5.496704	c8 = -0.496033
c9 = -2.633392	c9 = 8.351044	c9 = 9.909546

5.14. C.5. Data Set I, Swell Percent Prediction, Run4 Analysis Output Karva Language

Asec.+.X5.c6.d4.d1.d0.c3.d4.c2.d4.d1.d4

+

Coth.+.Sinh.d1.c7.c3.d0.c0.d4.c4.d2.c8.d5

+

Coth.+.Sinh.d4.c0.d2.d0.c5.c3.c2.d1.c0.d5

+

Sech.-.4Rt.d1.c6.d2.c7.d2.d0.c9.d1.d3.d2

Numerical Constants

Gene 1	Gene 2	Gene 3	Gene 4
c0 = 5.813965	c0 = -2.2836	c0 = -1.71463	c0 = 6.262756
c1 = 4.766022	c1 = -6.104766	c1 = 9.782288	c1 = -1.523315
c2 = 1.746948	c2 = 9.612122	c2 = -5.266113	c2 = 8.931611
c3 = -0.589386	c3 = 0.369751	c3 = 4.917878	c3 = -3.804596
c4 = -1.678101	c4 = 1.743714	c4 = 2.231109	c4 = 7.085601
c5 = 8.102325	c5 = -6.968812	c5 = -2.119567	c5 = 8.215393
c6 = 9.505341	c6 = -8.590302	c6 = 8.495911	c6 = 8.166534
c7 = 8.257599	c7 = -1.598663	c7 = 7.358459	c7 = 3.226501
c8 = 2.715393	c8 = -0.435638	c8 = 9.514526	c8 = -9.132629
c9 = 9.951294	c9 = 8.638947	c9 = -3.960328	c9 = -5.459351

5.15. C.6. Data Set II, Swell Percent Prediction, Run5 Analysis Output Karva Language

Sinh.Div4./.-.Inv./.d1.c9.c3.d0.d1.c6.d0.d1.d2.c9.d1.d2.d4.d4.d0.c4.d2.d4.d2

+

Div4.Coth.Sub3.Exp.Sub4.+.d0.d2.c9.d0.c3.c6.d0.d3.d3.d3.c1.d0.d1.d4.d1.c2.d4.d3.d0

+

Div3.Sech.-.Inv.Sin.d0.c6.c7.c9.d2.d0.c0.d1.d0.c3.d2.d2.c1.d0.d3.d0.d2.d2.d0.d0

+

Sin.Div4./.-.Inv./.d1.c2.c1.d0.c2.d1.d4.d1.d0.d1.d1.d0.c3.d1.c2.d2.d0.d3.c6

+

Sin.5Rt.d3.d3.c3.Pi.c2.c6.c0.c8.d1.d4.d3.c1.d1.d0.d1.c0.c4.c2.d2.c1.d0.d3.d4

Numerical Constants

Gene 1

Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
c0 = -5.820923	c0 = -1.056274	c0 = -2.241272	c0 = 1.457733	c0 = 6.620606
c1 = 0.666779	c1 = 9.129577	c1 = 3.286316	c1 = 4.73999	c1 = -0.591522
c2 = 3.730683	c2 = -5.227691	c2 = 6.863617	c2 = -0.012329	c2 = 7.294159
c3 = 5.581818	c3 = 3.317108	c3 = 6.89206	c3 = 4.496155	c3 = 5.206695
c4 = -0.35672	c4 = -7.241302	c4 = -5.599609	c4 = 1.251007	c4 = -1.372375
c5 = 2.134002	c5 = 0.508484	c5 = 9.169433	c5 = -8.398041	c5 = 9.07077
c6 = -0.71402	c6 = 9.300689	c6 = 5.797852	c6 = -9.778778	c6 = 0.014129
c7 = 5.104156	c7 = -5.599609	c7 = -2.049743	c7 = 6.82251	c7 = 5.955658
c8 = -2.706177	c8 = 9.965882	c8 = -2.241272	c8 = 9.566132	c8 = -7.490601
c9 = -4.301025	c9 = -6.786621	c9 = -1.126709	c9 = -0.223022	c9 = 5.974152

Where;

Addition	+		
Substraction	-		
Multiplication	*		
Division	/		
Power	Pow		
Square Root	Sqrt		
Exponential	Exp		
10^x	Pow10		
Logarith of base 10	Log		
Inverse	Inv		
Absolute value	Abs		
Negation	Neg		
No operation	Nop		
X 0f the power of 2	X2		
X 0f the power of 3	X3		
X 0f the power of 4	X4		
X 0f the power of 5	X5		
Cube root	3Rt		
Quartic root	4Rt		
Quantic root	5Rt		
Addition with 3 inputs	Add3		
Substraction with 3 inputs	Sub3		
Multiplication with 3 inputs	Mul3		
Division with 3 inputs	Div3		
Addition with 4 inputs	Add4		
Substraction with 4 inputs	Sub4		
Multiplication with 4 inputs	Mul4		
Division with 4 inputs	Div4		
Number Pi	Pi		
Euler's number	E		
Sine	Sin		
Cosine	Cos		
Tangent	Tan		
Cosecant	Csc		
Secant	Sec		
Cotangent	Cot		
Arcsine	Asin		
Arccosine	Acos		
Arctangent	Atan		
Arccosecant	Acsc		
Arcsecant	Asec		
Arccotangent	Acot		
Hyperbolic Sine	Sinh		
Hyperbolic cosine	Cosh		
Hyperbolic tangent	Tanh	Hyperbolic secant	Sech
Hyperbolic cosecant	Csch	Hyperbolic cotangent	Coth