Analysis of Heavy Metal Concentration in Soils of a Waste Disposal Site in Khulna using Artificial Intelligence Techniques

by

Shyamol Kumar Sarkar

A thesis submitted in partial fulfilment of the requirements for the Degree of

Master of Science in Civil Engineering

Khulna University of Engineering & Technology

Khulna, Bangladesh

September 2019

Declaration

This is to certify that the thesis work entitled *"Analysis of Heavy Metal Concentration in Soils of a Waste Disposal Site in Khulna using Artificial Intelligence Techniques"* has been carried out by Shyamol Kumar Sarkar in the Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.
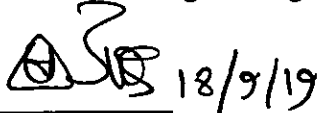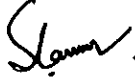

Dr. Md. Rafizul Islam

Professor

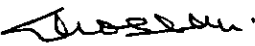Shyamol Kumar Sarkar

Roll No. 1701555

Approval

This is to certify that the thesis work submitted by Shyamol Kumar Sarkar entitled "*Analysis of Heavy Metal Concentration in Soils of a Waste Disposal Site in Khulna using Artificial Intelligence Techniques*" has been approved by the board of examiners for the partial fulfilment of the requirements for the degree of *Master of Science in Civil Engineering* in the Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh in September 2019.

BOARD OF EXAMINERS

1. _____
   Dr. Md. Rafizul Islam   Chairman
   Professor, Department of Civil Engineering   (Supervisor)
   Khulna University of Engineering & Technology, Khulna

2. _____   Member
   Dr. Md. Shahjahan Ali
   Professor and Head, Department of Civil Engineering
   Khulna University of Engineering & Technology, Khulna

3. _____
   Dr. Quazi Hamidul Bari   Member
   Professor, Department of Civil Engineering
   Khulna University of Engineering & Technology, Khulna

4. _____
   Dr. Md. Saiful Islam   Member
   Professor, Department of Civil Engineering
   Khulna University of Engineering & Technology, Khulna

5. _____
   Dr. Md. Delwar Hossain   Member
   Professor   (External)
   Department of Civil Engineering
   Bangladesh University of Engineering and Technology, Dhaka

Acknowledgement

To

My family

for their love, support and encouragement

and

Specially to my supervisor

for his unconditional support and inspiration,

without whom this research would not have been possible

Abstract

The collection of soil samples is labored and time consuming as well as the determination of heavy metal concentrations in laboratory was expensive. To these attempts, artificial intelligence techniques (AI) such as adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and artificial neural networks (ANN) were implemented for the analysis of heavy metal concentrations in soils of a selected waste disposal site at old Rajbandh, Khulna. The aim of this study was to fix the functions, algorithms, optimization methods for AI techniques based on their best performance and then select a best technique for the analysis of heavy metal concentrations in soils. In this study, soil samples were collected from eighty-five locations at a depth 0-30 cm from the existing ground surface from the selected disposal site. In the laboratory, the concentrations of heavy metals of Pb, Cu, Ni, Zn, Co, Cd, As, Sc, Hg, Mn, Cr, Ti, Sb, Sr, V and Ba in soils were measured.

Result reveals the model with SCP, gaussmf, linear and hybrid was the best-fitted model of ANFIS for the prediction of heavy metal concentrations in soils. In addition, in SVM analysis, the model SVM-RBF with 15 folds was selected for the prediction of heavy metal concentrations in soils. In ANN, the model LT (Levenberg-marquardt and Tansig functions) with neuron structure 2-10-1 was selected. The accuracy of the predicted results were checked based on the acceptable limits of prediction parameters like R value, RMSE, MAPE, GRI and percentage recovery. Among all heavy metals analysis in ANFIS, the maximum R-value 0.999 was found with the minimum RMSE 0.12 for Sc indicating the best correlation in prediction of Sc in soils. The others value of prediction parameters (MAPE= 36.00, GRI=1.50, percentage recovery=123.43%) for Sc were found within the acceptable limits. In addition, in SVM analysis, maximum R-value 0.73 with RMSE 2.03 was found for Cu; while, maximum R-value 0.88 with the minimum RMSE 1.01 for As was found in ANN. The results demonstrated that ANFIS model was a reliable technique than that of other counterparts of SVM and ANN to analyse the heavy metal concentrations in soils with the acceptable degree of robustness and accuracy. Therefore, the performance of AI techniques may be expressed by the sequence of ANFIS > SVM > ANN. Here it can be noted that one can easily be computed the concentration of a particular heavy metal in soils by inserting GPS values (latitude and longitude) only in the developed rule viewer of ANFIS. Therefore, this newly developed model will further be helpful for other researchers in this line to analysis heavy metal concentration in soils of selected waste disposal sites.

Table of Content

List of Tables

List of Figures

Nomenclature

AAS   : Atomic Absorption Spectroscopy

AI   : Artificial Intelligence

ANFIS   : Adaptive Neuro-Fuzzy Inference System

ANN   : Artificial Neural Networks

As   : Arsenic

Ba   : Barium

BBO   : Biogeography-Based Optimization

BP   : Back-Propagation

Ca   : Calcium

Cd   : Cadmium

Co   : Cobalt

Cr   : Chromium

Cu   : Copper

EC   : Electric Conductivity

FCM   : Fuzzy-c-means method

FDA   : Food and Drug Administration

Fe   : Iron

FFNN   : Feed-Forward Neural Network

FL   : Fuzzy Logic

Ga's   : Genetic Algorithms

Gaussmf   : Gaussian Membership Function

Gbellmf   : Generalized Bell- Shaped Membership Function

GIS   : Geographic Information System

GLP   : Good Laboratory Practice

GPS   : Global Positioning System

GRI   : Geometric Reliability Index

Hg   : Mercury

ICH   : International Council for Harmonisation

KBS's   : Knowledge-Based Systems

KCC   : Khulna City Corporation

KCPA   : Khulna City Planning Authority

LFG	: Land Fill Gas
LL	: Liquid Limit
LL	: Model combined with TRAINLM and Logsig
Logsig	: Log-Sigmoid Transfer Function
LP	: Model combined with TRAINLM and purelin
LT	: Model combined with TRAINLM and Tansig
MAPE	: Mean Absolute Percentage Error
MF	: Membership Function
Mg	: Manganese
Mg/kg	: Milligram per kilogram
Mn	: Manganese
MSW	: Municipal Solid Waste
Ni	: Nickel
NN	: Neural Network
OL	: Model combined with TRAINOSS and Logsig
OP	: Model combined with TRAINOSS and purelin
OT	: Model combined with TRAINOSS and Tansig
Pb	: Lead
PI	: Plasticity Index
Psigmf	: Product of Two Sigmoidal Membership Function
Purelin	: Linear Transfer Function
R	: Correlation Coefficient
RBF	: Radial Basis Function
RMSE	: Root Mean Square Error
Sb	: Antimony
Sc	: Scandium
SCP	: Sub-Clustering Partitioning
SL	: Model combined with TRAINSCG and Logsig
SP	: Model combined with TRAINSCG and Purelin
Sr	: Strontium
ST	: Model combined with TRAINSCG and Tansig
SVM	: Support Vector Machine
SVM-C	: Cubic Support Vector Machine

SVM-G   : Gaussian Support Vector Machine

SVM-L   : Linear Support Vector Machine

SVM-Q   : Quadratic Support Vector Machine

SVM-RBF   : Radial Basis Function Support Vector Machine

SW   : Surface Water

Tansig   : Tangent Sigmoid Transfer Function

TDS   : Total Dissolved Solid

Ti   : Titanium

TRAINLM   : Levenberg-Marquardt Training Function

TRAINOSS   : One-Step Secant Training Function

TRAINSCG   : Scaled Conjugate Gradient Training Function

Trapmf   : Trapezoidal-Shaped Membership Function

Trimf   : Triangular-Shaped Membership Function

V   : Vanadium

Zn   : Zinc

CHAPTER 1

INTRODUCTION

1.1 Research Background

Heavy metals are metallic elements that have relatively high density and toxic behavior even at low concentration (Alloway et al., 1990). The study of heavy metals in soils has increased in the last decades because of their adverse environmental and human health effects (Tahir et al., 2007). In waste disposal site, municipal solid waste (MSW) decomposes and produces three components of solid (degraded waste); liquid (leachate that is infiltrating into the underlying layer) and landfill gas (Sanjida and Rafizul, 2018). Open dumping facilities release huge quantity of harmful as well as poisonous chemicals like heavy metals to the surrounding water bodies as well as underlying soil layer, etc. Most of the environmental and human health problems come from the emission of heavy metals from the propagated leachate, contaminated soil, landfill gas (LFG), non-methanic volatile organic compounds as well as hazardous air pollutants in waste disposal site (Talib et al., 2008). In Khulna city, most of the MSWs were collected from door to door without any sorting and dumped in an open disposal site at Rajbandh. The emissions of toxic metal element from MSW, leachate and soil will be vulnerable to the environmental components and the nearby inhabitants. The evaluation of heavy metal distribution in soils is important to save the environment. Moreover, for various soil assessment techniques, heavy metal concentrations are needed. However, the collection of soil samples are labored, time consuming as well as the determination of heavy metal concentration in soils from laboratory is expensive.

The prediction of heavy metal concentration using artificial intelligence techniques (AI) may be the solution to solve this problem. In the literature, the AI techniques such as adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), artificial neural networks (ANN), fuzzy logic (FL), knowledge-based systems (KBSs), genetic algorithms (GAs), biogeography-based optimization (BBO) etc. are available and most usable. These AI techniques have many functions, algorithms, optimization methods, which can be used for the prediction of heavy metal concentration in soils. The aim of this study is to fix functions, algorithms and optimization methods for all AI techniques based on their best performance and select a best AI technique for the analysis of heavy metal concentrations in

soils. In this study, for the analysis of heavy metal concentrations in soils, the AI techniques such as ANFIS, SVM and ANN was performed.

A study stated that over the last few years or so, various AI techniques for analysis of heavy metals concentrations and other quality parameters; environmental modelling; water quality monitoring and assessment; estimation as well as forecasting in climatic sciences (Soyupak et al., 2003). The ANFIS had been studied to predict the concentrations of Cd in Filyos River, Turkey (Sonmez et al., 2018). In recent works on AI have resulted in finding a novel machine learning theory called SVM. The SVM method relies on the statistical learning theory, which enables learning machines to generalize the unknown data. It was introduced in the 1990s as a non-linear solution for classification and regression tasks (Behzad et al., 2009; Wei and Yang, 2010). In addition, the use of ANN has also increased in many areas of engineering field. ANN has been used successfully for modelling of heavy metal in soil, contaminated site characterization, settlement of structures, soil permeability, soil compaction, soil swelling and classification of soils (Rooki et al., 2011). In addition, the use of ANN has also increased in many areas of engineering field. ANN has been used successfully for modelling of heavy metal in soil, contaminated site characterization, settlement of structures, soil permeability, soil compaction, soil swelling and classification of soils (Rooki et al., 2011).

In this study, AI techniques such as ANFIS, SVM and ANN were implemented to analysis heavy metal concentrations in soils of a selected waste disposal site at old Rajbandh, Khulna. In ANFIS, the validation of models was performed by interchanging different input and output membership functions as well as optimization methods to select the best model of ANFIS. In addition, for SVM analysis various models with different kernel functions were formed to select best-fitted model of SVM. The cross-validation with different folds was also performed to control overfitting of the data. Furthermore, for selecting the best-fitted model of ANN; different neuron structures, different training functions as well as various transfer functions was implemented. The results of ANFIS, SVM and ANN model were also compared with the satisfactory values of correlation coefficient (R), root mean square error (RMSE), mean absolute percentage error (MAPE), geometric reliability index (GRI) and percent recovery. Therefore, the newly developed model of AI techniques will further be helpful for other researchers in this line to analysis heavy metal concentration in soils of selected waste disposal sites.

1.2 Objectives of the Study

Contamination of soils in MSW disposal site is a common incident in least developed Asian countries like Bangladesh. For the assessment and remediation of environmental pollutions and risk, it is very important to know the level of risk from contaminated soils. In addition, the heavy metal concentrations in soils implied the level of contamination of soils. The collection of soils samples from MSW disposal site are labored, time consuming and determination of heavy metal concentrations from laboratory is expensive. Prediction of heavy metal concentration may be the solution of this problem. In this study, AI techniques with various algorithms, kernel functions and membership functions were performed for the analysis of heavy metal concentrations in soils of a waste disposal site at old Rajbandh, Khulna. The main objectives of this study are summarised as follows:

1. To fix different models of artificial intelligence techniques for the prediction of heavy metal concentration in soils of a selected waste disposal site.

2. To check the validity of predicted results from fix models of artificial intelligence techniques.

3. To select best artificial intelligence technique for the analysis of heavy metal concentrations in soils.

1.3 Significance of the Study

Urbanization, industrialization, agriculture and exploitation of natural resources are basic activities associated with living in contemporary societies that have imposed pollutant loads especially toxic metals into natural cycles such as soils, water and air cycles (Gribble, 1994; Nasrabadi, 2015). Nowadays, throughout the world, heavy metals have been taken into consideration due to their ability to accumulate in the biota, toxicity and adverse health effects even at low concentrations (Hossein et al., 2014; Morillo et al., 2002). In this regard, heavy metals pollution of the soils is one of the serious problems for the population as well as the environment. The analysis and prediction of heavy metals of Pb, Cu, Ni, Zn, Co, Cd, As, Sc, Hg, Mn, Cr, Ti, Sb, Sr, V and Ba will be used to aware the general population of Khulna region. The study will also help to inform the authorities about the sources of spreading and distribution pattern of these heavy metals in soils of the disposal site as well

as help to take necessary steps to control the spread of heavy metals outside the disposal site area. The results from the study will also be used to find out the area needed to take under immediate remedial action to remove the heavy metals where the levels were too high. In addition, prediction of unknown points by simply putting their coordinate values will be very useful when initially a location has to be selected for various purposes and development. Otherwise, the laboratory tests can be very expensive and time consuming. Moreover, the prediction from different AI techniques is confusing for their various functions and algorithms. The study will help to select a best-fitted model with the particular functions and algorithms for different AI techniques by the model validation. The particular model of prediction of heavy metal concentrations with the best accuracy will help the researchers in this line.

1.4 Scope and Limitations

Rapid growth of population and industrialization surrounding the Khulna city tends to increase the generation of MSW creates additional load to MSW management system and finally contaminated the environmental components and surrounding soil layers. Contaminations of heavy metal in and around of the old Rajbandh disposal site possess a threat to the inhabitant. The spatial and temporal variation of heavy metals concentration in and around the soils of waste disposal site was described using conventional statistics, which shows the magnitude and pattern of contamination of heavy metals. AI techniques presented a developed network to predict the unknown values of heavy metal concentrations, if latitude and longitude of sampling were set as input. This helps to evaluate the contamination of heavy metal at larger distance and the spatial distribution can be displayed. Thus, it will be feasible to take necessary steps to manage the unintended disposal of waste to stop the spread of heavy metals. Further study can also be carried out where it is needed to implement in future. The main limitation of the study is the software (MATLAB), which predicts the values, has its limitations for predicting the unknown concentration points based on the input. Also no direct computation of index to assess the risk due to heavy metal contamination in and around the soils of waste disposal site. Some field conditions were ignored during the sampling procedure and the laboratory test values do not represent the in-situ condition of collected soils samples. More carefulness will help to get expected result.

1.5 Outline of the Study

The study has been presented in five distinct chapters comprising different aspects of this study. The outline and relations among these five chapters as depicted in Figure 1.1.

Chapter 1 described general knowledge on the background of waste disposal site, MSW, contaminated soils, heavy metals and the artificial intelligence (AI) techniques like adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and artificial neural network (ANN). In addition, objectives of the present research, scope and limitations of this study are also highlighted in this chapter.

Chapter 2 deals with the information related to the contamination of soils due to the presence of heavy metals in the soils of disposal site. This chapter also illustrates the background of different functions and algorithms of AI Techniques like ANFIS, SVM and ANN used in this study. The literature review presented in this chapter was collected from available previous research reports and technical papers related to this topic. The literature review begins with an introduction to MSW and waste disposal facilities first, and is followed by the generation of MSWs and contamination of soils due to heavy metals. The interrelationship between soils, MSW and heavy metals is illustrated in this chapter. After instituting, the literature review related to heavy metal and soils, general discussion on AI Techniques in field of prediction and comparison using MATLAB also discussed in this chapter.

Chapter 3 deals with the overall research methodology in this study. This chapter includes the information about the study area and soils condition. In this study, total eighty-five disturbed soils samples were considered from distinct locations in and around the waste disposal site at old Rajbandh in Khulna. Among the soils samples, some were collected from secondary sources. The method of soils sampling was also highlighted in this chapter. In the laboratory, the concentration of relevant heavy metal in soils were measured and monitored through standard test methods and hence highlighted in this chapter. Then, the measured concentrations of heavy metals were used to perform the descriptive conventional statistics using MS Excel to assess the basic features of soils data in a simpler. The AI techniques such as ANFIS, SVM and ANN were performed to predict the heavy metal concentrations, which were also highlighted in this chapter.

Chapter 4 deals with the model selection for ANFIS, SVM and ANN as well as prediction the concentrations of heavy metals such as Pb, Cu, Ni, Zn, Co, Cd, As, Sc, Hg, Mn, Cr, Ti, Sb, Sr, V and Ba in soils of waste disposal site for the unknown sampling points. In this chapter, regression analysis was performed using ANFIS, ANN and SVM through MATLAB to get the predicted heavy metal concentrations and accuracy parameters like correlation coefficient, R and root mean square error (RMSE). In addition, for the performance of various AI techniques were assessed by ascertaining the error in the predictions based on mean absolute percentage error (MAPE). For checking the accuracy level of prediction percent recovery and geometric reliability index (GRI) were also used and hence discussed in this chapter. In addition, this chapter represents the comparison of AI techniques based on various prediction parameters and finds the best one for prediction of heavy metal concentrations in soils.

Chapter 5 draws conclusions based on logical reasoning of the laboratory data and the analysis of outcomes as well as provides a few recommendations for future studies.
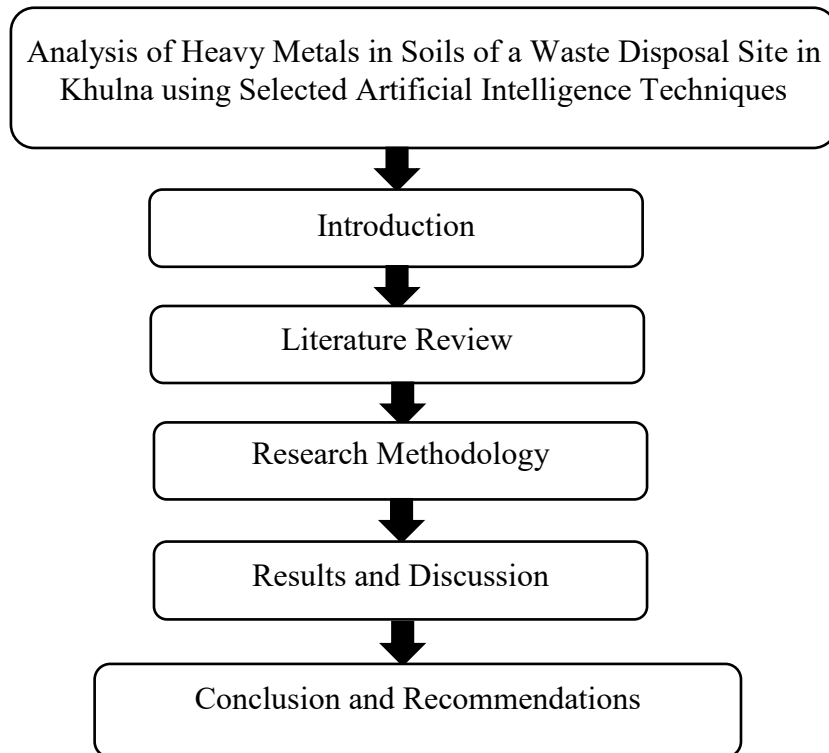
Figure 1.1: Outline and relations between the chapters of this study.

CHAPTER 2

LITERATURE REVIEW

2.1 General

This chapter deals with the information related to the contamination of soils due to the presence of heavy metals in the soils of disposal site. This chapter also illustrates the background of different functions and algorithms of Artificial Intelligence (AI) Techniques like adaptive neuro fuzzy inference system (ANFIS), support vector machine (SVM) and artificial neural network (ANN) used in this study. The literature review presented in this chapter was collected from available previous research reports and technical papers related to this topic. The literature review begins with an introduction to MSW and disposal facilities first, and is followed by the generation of MSWs and contamination of soils due to heavy metals. The interrelationship between soils, MSW and heavy metals is illustrated in this chapter. After instituting, the literature review related to heavy metal and soils, general discussion on AI Techniques in field of prediction and performance analysis using MATLAB also discussed in this chapter.

2.2 Concept of Municipal Solid Waste

Municipal Solid Waste (MSW) more commonly known as trash or garbage consists of everyday items that use and then throw away, such as product packaging, grass clippings, furniture, clothing, bottles, food scraps, newspapers, appliances, paint, and batteries. This comes from our homes, schools, hospitals, and businesses (epa.gov/epawaste/nonhaz/ municipal/web/html/). Many changes of MSW generation and composition have taken place due to urbanization including increase in the population. The rate of consumption has risen and the lifestyle of the people, too, has changed (Quina et al., 2008). Several studies have shown how increasing waste has affected society and the environment (Hashim and Chu, 2004). The amount of waste production is a sign of the level of industrialization or a degree of development of a country or a city (Ludwig and Keller, 2003).

2.3 Dumpling Facilities of Municipal Solid Waste

Landfill is one of the most widely used MSW management techniques; however, it needs high standard of environment protection in the operation of landfill (Oyeku and Eludoyin,

2010). The changing from dumping to high standard of environment protection needed time change technology, change in thinking and behaviour (Pugh, 1999). In addition, sanitary landfill is an engineering technique for the disposal of MSW on the land by spreading them in thin layers followed by compacting them to the smallest practical volume before covering them with soils at regular intervals (Figure 2.1). Sanitary landfilling involves placing MSW in lined pits with appropriate means of leachate and landfill gas control (Oyeku and Eludoyin, 2010). It is highly recognized as an environmentally and internationally desired technique of MSW disposal since it minimizes environmental damage and thus eliminates odors. The commonly practices landfills are as follows.

2.3.1 Sanitary Landfill

Sanitary landfills are sites where waste is isolated from the environment until it is safe. It is considered when it has completely degraded biologically, chemically and physically.
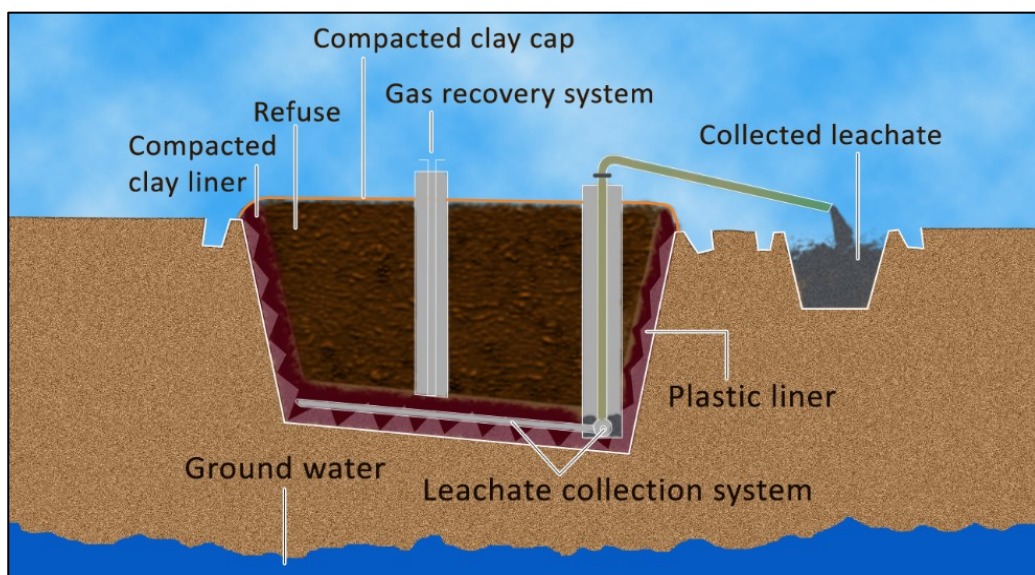
Figure 2.1: Sanitary landfill for MSW dumping (Source: https://pixfeeds. com/images/ sanitary-landfill-vs-open-dump.png)

In high-income countries, the level of isolation achieved may be high. However, such an expensive high level of isolation may not be technically necessary to protect public health. Four basic conditions should be met before a site can be regarded as a sanitary landfill (Figure 2.1). The ways of doing this should be adapted to local conditions. The immediate goal is to meet the best extent possible, the four stated basic sanitary landfill conditions, with a longer-term goal to meet them eventually in full. Small incremental improvements in

landfill design and operation over several years are more likely to succeed than attempts to make a single, large leap in engineering expectations (Thurgood, 1999).

2.3.2 Open Dumping of MSW

An open dumping is defined as a land disposal site at which solid wastes are disposed of in a manner that does not protect the environment, are susceptible to open burning, and are exposed to the elements, vectors, and scavengers (Figure 2.2). Whether it occurs on an open lot or pasture, down a ravine, along the roadside, in rural or urban communities, open dumping threatens the health and safety of everyone around it, especially children and older adults. It is also against the law. Besides looking bad and smelling worse, open dumps threaten surface and drinking water, provide breeding grounds for dangerous vectors such as rats (tetanus, rabies, etc.), and mosquitoes (St. Louis Encephalitis). Open dumps also present a variety of safety risks from sharp objects or needles, combustible levels of explosive methane, and unstable slopes, which can shift and potentially injure or kill.

Figure 2.2: Open dumping facility (Source: https://i0.wp. com/kashmirreader.com/wp-content/uploads/Dumping-site.jpg).

2.4 Contaminated Site

Contaminated land contains substances in or under the land that are actually or potentially hazardous to health or the environment. Areas with a long history of industrial production are known as brownfield land. Many such sites may be affected by their former uses such as mining, industry, chemical and oil spills and waste disposal

9

(https://en.wikipedia.org/wiki/Contaminated_land). Typically because of previous industrial or agricultural activities, it reaches a point where they become or have the potential to become dangerous to people or other organisms in the surrounding natural environment. The Environmental Protection Agency estimates that tens of thousands of contaminated sites remain in the United States that have the potential to cause environmental harm (https://www.epa.gov/report-environment/contaminated-land). The present environmental condition of Bangladesh is not equilibrium at all. Severe air, water and soils pollution are threatening human health, ecosystems and economic growth of Bangladesh. In Khulna, such a contaminated site at old Rajbandh (Figure 2.3) governs the environmental effects of Khulna city. Clean earth plays an integral part in the process of remediating and reclaiming these lands for future productive uses.

Figure 2.3: Contaminated site at old Rajbandh, Khulna
(Source:https://encryptedtbn0.gstatic.com/images?q=tbn:ANd9GcR3lHQJM
ZL_CglzCp9-rWhcFAo5AdPUmd-WMIUIyrIGS1qh4a9K).

2.5 Context of Heavy Metals

Metals are defined as any element that has a silvery lustre and is a good conductor of heat and electricity. Theoretically, there are many terms used to describe and categorize metals, including trace metals, transition metals, micronutrients, toxic metals and heavy metals. A metal having a specific gravity more than 5 gm/cm^3 is classified as heavy metal. In this study, to analyse and predict the heavy metal concentrations in soils different AI techniques

were performed through MATLAB. The providence of heavy metals in soils as well as environment and interrelationship between them are described in the following articles.

2.5.1 Providence of Heavy Metal in Soils and Environment

The incapability to determine metal species in soils obstructs efforts to understand the mobility, bioavailability and fate of contaminant metals in the environmental systems together with the assessment of health risks posed by metal elements, and the development of methods to remediate metal contaminated sites. In soils, metals are found in one or more of the following several "pools" in the soils:

  i.   Dissolved in the soils solution;
 ii.   Occupying exchange sites in inorganic soils constituents;
iii.   Specially adsorbed in inorganic soils constituents;
 iv.   Associated with insoluble soils organic matter;
  v.   Precipitated as pure or mixed solids;
 vi.   Present in the structure of secondary minerals; and/or
vii.   Present in the structure of primary minerals

However, in some natural soils, up to 30 to 60% of heavy metals can occur in unstable forms developed from parent materials rich in metal contamination as well as in the contaminated soils. Natural and anthropogenic sources are one of the root cause of heavy metal contamination which has caused widespread and variable the hazardous possibilities of environmental and health effect (Tahir et al., 2007). According to Walker et al., (2003), the anthropogenic sources of metal contamination can be divided into five main groups: (1) metalliferous mining and smelting (As, Cd, Pb and Hg); (2) industry (As, Cd, Cr, Co, Cu, Hg, Ni and Zn); (3) atmospheric deposition (As, Cd, Cr, Cu, Pb and Hg); (4) agriculture (As, Cd, Cu, Pb and Zn); and (5) MSW disposal (As, Cd, Cr, Cu, Pb, Hg and Zn). (Vaalgamaa and Conley, 2008) also stated a natural activity is another cause of heavy metal contamination. Industries such as plating, ceramics, glass, mining and battery manufacture are considered the main sources of heavy metals in local water systems causing the contamination of groundwater with heavy metals. Furthermore, heavy metals which are frequently found in high concentrations of waste landfill leachate are also potential source of pollution for groundwater and underlying soils layer (Xiao et al., 2013).

2.5.2 Background on Soil Quality Standards of Heavy Metals used in this Study

Preventing of heavy metal contamination is critical; because cleaning of contaminated soils is extremely expensive and difficult to achieve. In order to evaluate soil quality, soil functions and response properties (concentration of heavy metal, climate, hydrology etc.) must be assessed. There are different guidelines according to land use; because people use land differently and this affects who and how people may be exposed to soil contamination. In addition, different organizations and countries provides the maximum allowable limits of heavy metal concentrations in soils shown in Table 2.1. This table describes the soil quality guidelines for the protection of environment and human health from contaminated sites.

Table 2.1: Maximum alowable limits of heavy metal concentrations (mg/kg) in soil of different sources (After: Fahmida and Rafizul, 2017)

Heavy Metal	WHO standard	CCME	Austria	Canada	Poland	Japan	UK	Germany	U.S.A
Pb	15-20	70	100	200	100	400	100	500	200
Cu	-	63	100	100	100	125	100	50	100
Ni	0-100	45	100	100	100	100	50	100	500
Zn	20-300	200	300	400	300	250	300	300	300
Co	-	40	50	25	50	50	-	-	40
Cd	0-30	1.4	5	8	3	-	3	-	0.7
As	-	12							
Hg	-	6.6							
Mn	200-9000	-							
Cr	0-85	64	100	75	100		50	200	1000
Sb	-	20							
V	-	130							
Ba	-	750							

A sustainable soil is referred to as the use of soil as a natural resource on a way that does not exert any negative effects on the environment. In this study, soil quality guidelines from different countries like Canadian Council of Ministers of the Environment (CCME), WHO, Poland, U.K, U.S.A, Austria, Germany and Japan were used to check the quality of soil.

2.5.3 Effects of Heavy Metal in Soils

Heavy metals exhibit toxic effects towards soils biota by affecting key microbial processes and decrease the number and activity of soils microorganisms. Even low heavy metal concentrations may inhibit the physiological metabolism of plant. Uptake of heavy metals

by plants and subsequent accumulation along the food chain is a potential threat to animal and human health (Singh et al., 2011). The chemical behavior of heavy metals in soils is controlled by a number of processes including metal cation release from contamination source materials (e.g., fertilizer, sludge, smelter dust, ammunition, and slag), cation exchange and specific adsorption onto surfaces of minerals and soils organic matter, and precipitation of secondary minerals. Increased anthropogenic inputs of Zn in soils have caused considerable concern relative to their effect on water contamination (Zhang et al., 2012). In addition, oxidizing conditions generally increase the retention capacity of metals in soils, while reducing conditions will generally reduce the retention capacity of metals. Soils reduction has been shown to result in the coincident release of metals associated with minerals that are susceptible to reductive dissolution, in particular Mn and Fe oxides. The MSW that disposed without proper planning and treatment are rich in different types of heavy metals. When MSW is dumped in a disposal site, they take a long time to disintegrate depending on their nature. For example, paper towel takes 2-4 weeks to disintegrate, whereas plastic bag takes 200-1000 years. The metals of the greatest concern due to their extensive use, their toxicity and their widespread distribution is Hg, Pb, Cd, Cr and As (Addae, 2013). The toxicant may also cause effects on the microorganisms and soils fauna. However, the risk for further dispersion of the pollution to other recipients must also be considered.

2.6 Artificial Intelligence Techniques

In the literature, the father of Artificial Intelligence named "John McCarthy" stated that "The science and engineering of making intelligent machines, especially intelligent computer programs". The AI is accomplished by studying how human brain thinks and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems. AI field has a very wide scope in computation and automation the world. The algorithms and methods studied in AI include adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), artificial neural network (ANN), fuzzy logic (FL), knowledge-based systems (KBSs), genetic algorithms (GAs) and biogeography-based optimization (BBO) etc. (Mayfield and Fairbrother, 2013). Recent investigations have highlighted the application of the AI techniques such as ANFIS, SVM and ANN to the geotechnical engineering problems.

2.6.1 Adaptive Neuro Fuzzy Inference System

(Jang, 1993) first introduced the adaptive neuro-fuzzy inference system (ANFIS). The neuro-fuzzy approach combines ANN and fuzzy logic. It effectively integrates the learning capability of neural networks into a fuzzy inference system (Sonmez et al., 2018). It can be used to approximate any real continuous function on a compact set to any degree of accuracy (Jang et al., 1997). Depending on the types of inference operations upon if–then rules, most FIS can be classified into three types: Tsukamoto's system, Mamdani's system and Sugeno's system (Kişi and Öztürk, 2007). In this study, the first-order Sugeno fuzzy model is used because it has been used widely in engineering problems. The ANFIS model is able to use two different optimization methods (hybrid and back-propagation) to tune membership function (MF) and generate fuzzy rules. The hybrid method is a combination of least squares estimation combined with back-propagation method (https://www.mathworks.com/help/fuzzy/train-adaptive-neuro-fuzzy-inference-systems-gui.html).

2.6.1.1 Basic of ANFIS Model Structure

ANFIS model structure represents the overall view of ANFIS operation. It consists of five layers, and the basic functions of each layer are the input, fuzzification, rule inference, normalization and defuzzification (Emamgholizadeh et al., 2014). The equivalent ANFIS architecture of the first order is shown in Figure 2.4.
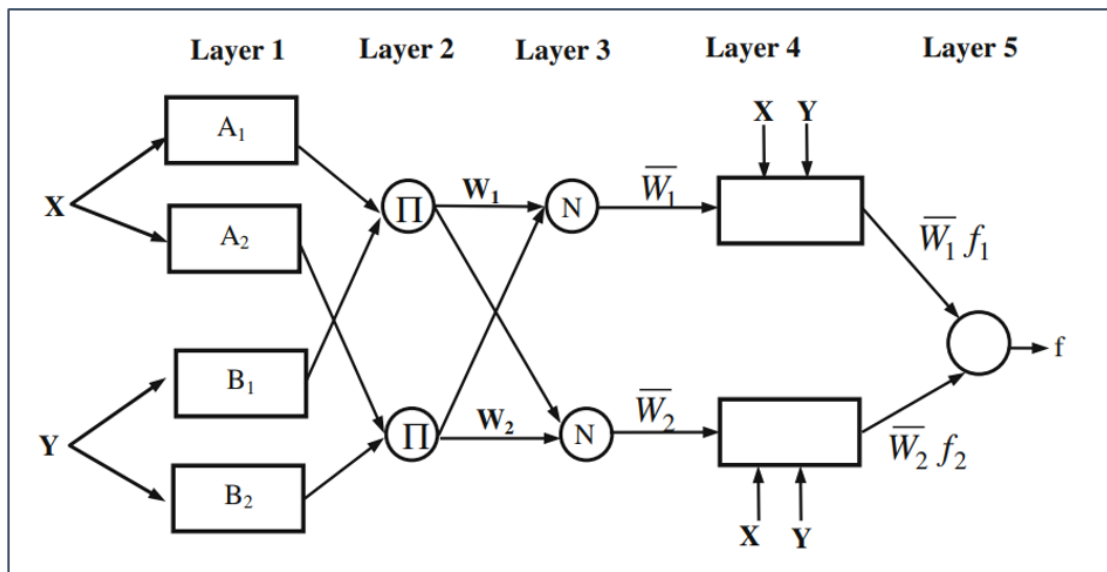
Figure 2.4: Equivalent ANFIS structure (Source: Emamgholizadeh et al., 2014).

The relationship between the input and output of each layer is summarized as follows.

Layer 1 (Input notes):

Every node (I) in this layer is an adaptive node, and each node in this layer generates membership grades of crisp inputs variable, which belong to each of convenient fuzzy sets by using the membership functions. Parameters in this layer are called premise parameters. The output of each node in this layer O_i^1 is defined by the following Equation (1).

$$O_i^1 = \mu_{Ai}(x) \ for \ i = 1,2 \ or$$

$$O_i^1 = \mu_{Bi-2}(y) \ for \ i = 3,4 \ ... ... \ (1)$$

Where x and y are the crisp input to the ith node, and A_i and B_{i-2} are the linguistic labels characterized by MFs μ_{Ai} and μ_{Bi}, respectively. Another major consideration that should be taken into account while designing ANFIS is the selection of proper MF. Several MFs were reported in the literature such as Gaussian, generalized bell-shaped, trapezoidal and triangular. Selection among the abovementioned MFs is generally based on the trial and error. Assuming a generalized bell-shaped as the MF, the node output in first layer O_i^1 can then be calculated by the following Equation (2).

$$O_i^1 = \mu_{Ai} = \frac{1}{1+(\frac{x-c_i}{a_i})^{2b_i}}$$

$$O_i^1 = \mu_{Bi-2} = \frac{1}{1+(\frac{y-c_i}{a_i})^{2b_i}} \ ... ... \ (2)$$

Layer 2 (Rule nodes):

Every node in this layer is a fixed node labelled π, and the output is the product of all the incoming signals. Each node output that represents the firing strength of a rule is computed by the following Equation (3):

$$O_i^2 = \mu_{Ai}(x)\mu_{Bi}(y), \quad i = 1 \ and \ 2 \ ... ... \ (3)$$

Layer 3 (Average nodes):

In this layer, the nodes calculate the ratio of the ith rule's firing strength to the sum of all rules' firing strengths; the firing strength in this layer is normalized and taken as $\overline{w_i}$. It is represented by the Equation (4):

$$O_i^3 = \overline{w_i} = \frac{w_i}{\sum w_i}, \quad i = 1 \ and \ 2 \ldots \ldots \ldots \ldots \ldots \ldots \ldots (4)$$

Layer 4 (Consequent nodes):

This layer's nodes are adaptive with node functions like the following Equation (5):

$$O_i^4 = \overline{w_i}.f_i = w_i(p_i x + q_i y + r_i), \quad i = 1 \ and \ 2 \ldots (5)$$

Where w_i the ith node is output from the previous layer and f_i is a linear function of input variables.

Layer 5 (Output nodes):

This layer's single fixed node computes the final output as the summation of all incoming signals. It is represented by the Equation (6):

$$O_i^5 = \overline{w_i} = \frac{\sum \overline{w_i}.f_i}{\sum \overline{w_i}} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . (6)$$

2.6.1.2 Rules of ANFIS

The Rules show how the shape of certain membership functions influences the overall result. Figure 2.5 shows the rules for first-order Takagi–Sugeno fuzzy model with a two input and one output system.
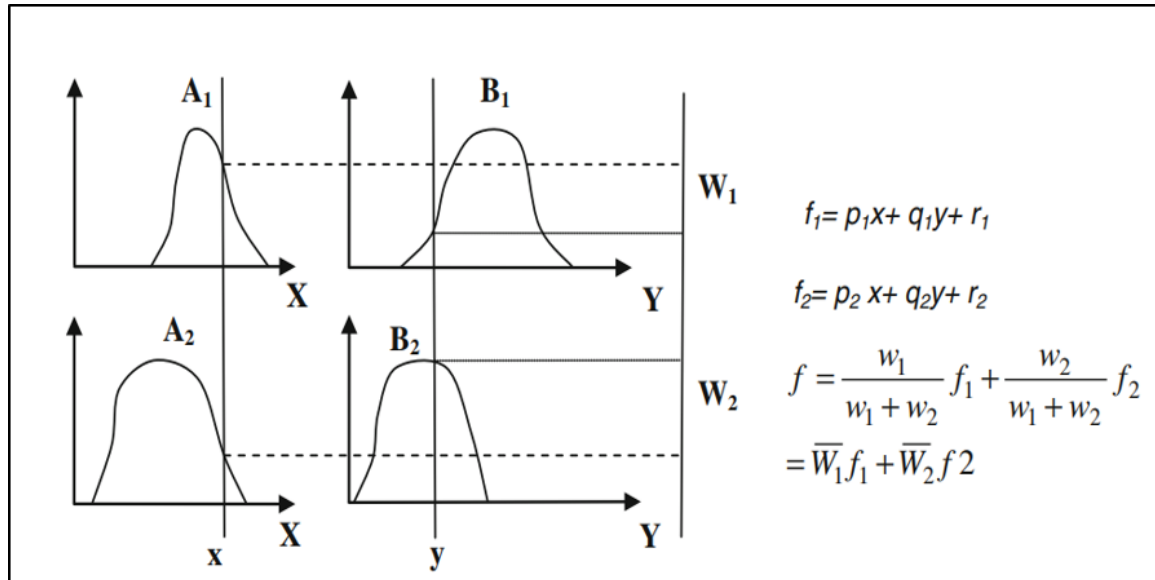
Figure 2.5: Sugeno's fuzzy if–then rule and fuzzy reasoning mechanism (Source: Emamgholizadeh et al., 2014).

Where x and y are input variables. Ai and Bi are the linguistic labels (low, medium, high, etc.) characterized by convenient membership functions, fi are the outputs within the fuzzy region specified by the fuzzy rule; p_i, q_i and r_i are the parameters of the output function (i = 1 or 2). In first-order Sugeno's system, if FIS has two inputs x and y and one output f, a typical rule set with two fuzzy IF/THEN rules. It can be expressed as:

$$\text{Rule 1: If x is } A_1 \text{ and y is } B_1; \text{ then } f_1 = p_1x + q_1y + r_1$$
$$\text{Rule 2: If x is } A_2 \text{ and y is } B_2; \text{ then } f_2 = p_2x + q_2y + r_2$$

2.6.1.3 Rule Viewer of ANFIS

The Rule Viewer displays a roadmap of the whole fuzzy inference process shown in Figure 2.6. The three plots across the top of the figure represent the antecedent and consequent of the first rule. Each rule is a row of plots, and each column is a variable. The rule numbers are displayed on the left of each row.
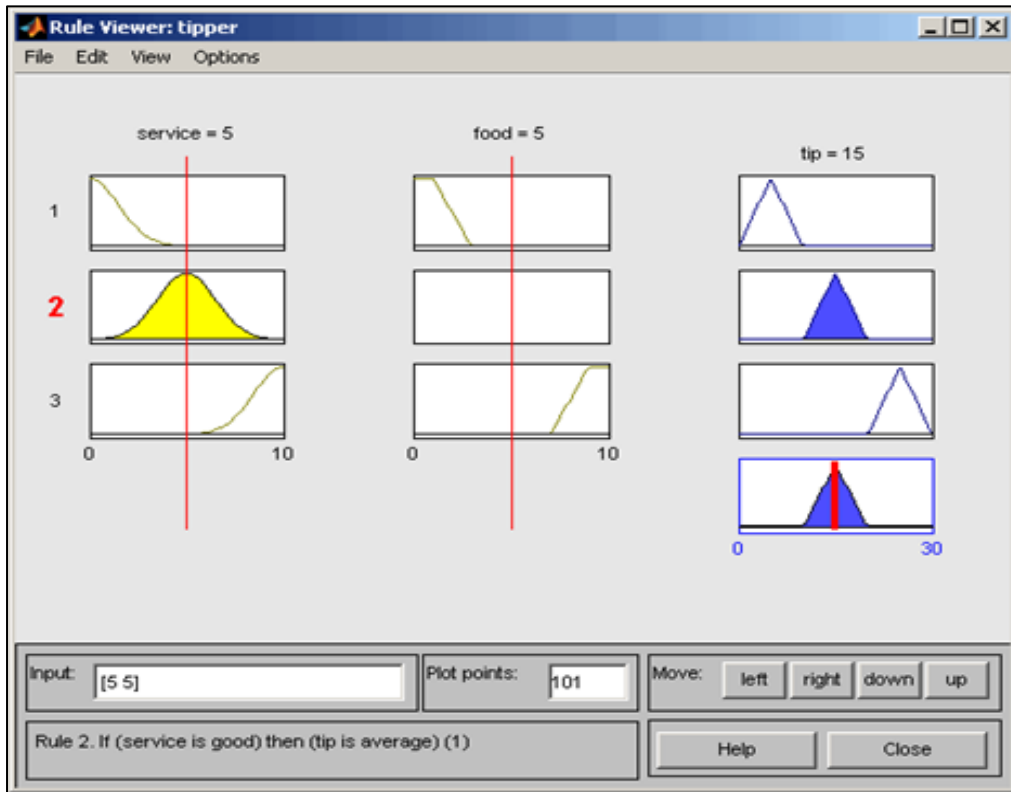
Figure 2.6: Rule viewer of food and service for tip (Source: Matlab Documentation, 2017a).

1. The first two columns of plots (the six yellow plots) show the membership functions referenced by the antecedent, or the if-part of each rule.
2. The third column of plots (the three blue plots) shows the membership functions referenced by the consequent, or the then part of each rule.
3. The fourth plot in the third column of plots represents the aggregate weighted decision for the given inference system.

This decision will depend on the input values for the system. The defuzzified output is displayed as a bold vertical line on this plot. The variables and their current values are displayed on top of the columns. In the lower left, there is a text field Input in which specific input values can be inserted. For the two-input system, it can be inserted as an input vector like [9 8] and then pressing Enter, it is adjusted these input values which provides new output. The red index line can be moved horizontally in order to change the input values. The plots can be shifted using left, right, down, and up to visualize the entire rule viewer.

18

2.6.1.4 Surface Viewer of ANFIS

The Surface Viewer is a graphical interface that examine the output surface of an FIS for any one or two inputs. Surface Viewer is a read-only editor. Using the drop-down menus, two desired input variables can be selected to the two input axes (X and Y); as well as the output variable can be assigned to the output (or Z) axis. Color bar shows the output range with colour variation (Figure 2.7). For creating a smoother plot, the Plot points field is specified the number of points on which the membership functions are evaluated in the input or output range. This field defaults to the minimum number of plots, 101. The different input membership functions of ANFIS is depicted in Table 2.2.
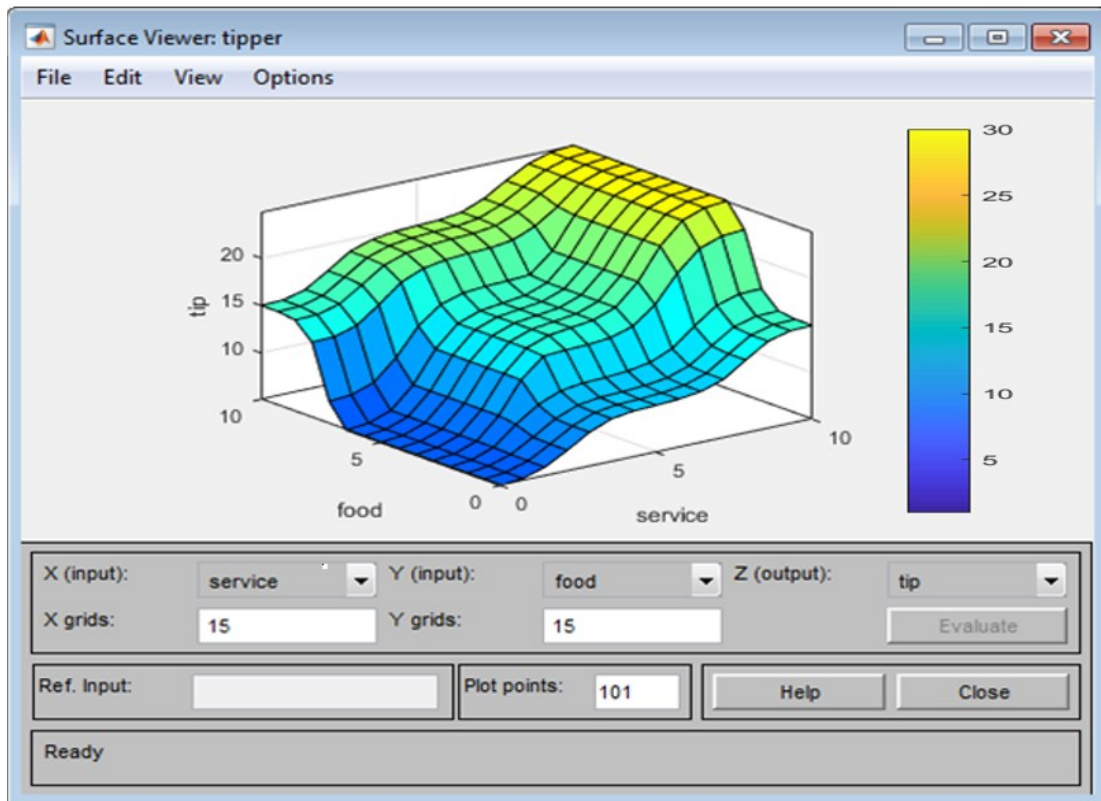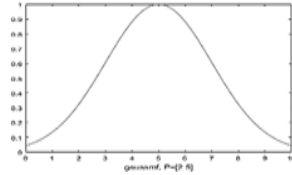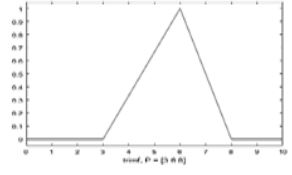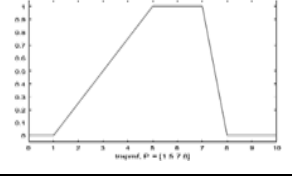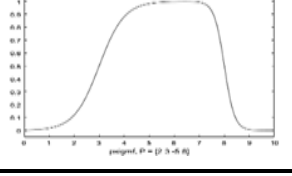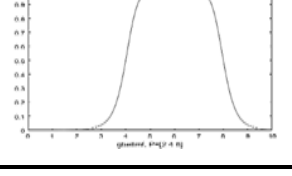
Figure 2.7: Surface viewer of food and service for tip (Source: Matlab Documentation, 2017a).

Table 2.2:Description of different input membership functions of ANFIS (After: Matlab Documentation, 2017a).

Membership Function	Full Meaning	Syntax	Description	Figure	Reference
gaussmf	Gaussian membership function	y = gaussmf(x,[sig c])	The symmetric Gaussian function depends on two parameters σ and c which listed in order in the vector [sig c].		Matlab 2017a Documentation
trimf	Triangle-Shaped Membership Function	y = trimf(x,[a b c])	The triangular curve is a function of a vector, x, and depends on three scalar parameters a, b, and c. The parameters a and c locate the "feet" of the triangle and the parameter b locates the peak.		Matlab 2017a Documentation
trapmf	Trapezoidal-shaped membership function	y = trapmf(x,[a b c d])	The trapezoidal curve is a function of a vector, x, and depends on four scalar parameters a, b, c, and d. The parameters a and d locate the "feet" of the trapezoid and the parameters b and c locate the "shoulders."		Matlab 2017a Documentation
psigmf	Product of two sigmoidal membership functions	y = psigmf(x,[a1 c1 a2 c2])	The psigmf is simply the product of two such curves plotted for the values of the vector x. $f1(x; a1, c1) \times f2(x; a2, c2)$ The parameters are listed in the order [a1 c1 a2 c2].		Matlab 2017a Documentation
gbellmf	Generalized bell-shaped membership function	y =gbellmf(x,params)	The generalized bell function depends on three parameters a, b, and c where the parameter b is usually positive. The parameter c locates the center of the curve.		Matlab 2017a Documentation

2.6.2 Support Vector Machine

Support vector machine (SVM) is a popular machine learning tool both for classification and regression, first identified by Vladimir Vapnik and his colleagues in 1992 (Cortes and Vapnik, 1995). SVMs are very specific class of algorithms, characterized by usage of kernels, absence of local minima, sparseness of the solution and capacity control obtained by acting on the margin, or on number of support vectors, etc. SVM regression contains all the main features that characterize maximum margin of algorithm. A non-linear function is leaned by linear learning machine mapping into high dimensional kernel induced feature space. The capacity of the system is controlled by parameters that do not depend on the dimensionality of feature space (Cherkassky and Ma, 2002). SVM regression and classification are considered as nonparametric technique because they rely on kernel functions. This functions are defined the loss function that ignores errors, which are situated within the certain distance of the true value. Statistics and Machine Learning Toolbox™ implements linear epsilon-insensitive SVM (ε-SVM) regression, which is also known as L1 loss. In ε-SVM regression, the set of training data includes predictor variables and observed response values. The goal is to find a function f(x) that deviates from y_n by a value no greater than ε for each training point x, and at the same time is as flat as possible.

2.6.2.1 SVM Classification

SVM classification is such a machine learning algorithm use to find a hyperplane in an N-dimensional space (N-the number of features) that distinctly classifies the data points. To separate the two classes of data points, many possible hyperplanes could be chosen (Figure 2.8). The objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

2.6.2.2 Hyperplanes

Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. In addition, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line (Figure 2.9a).
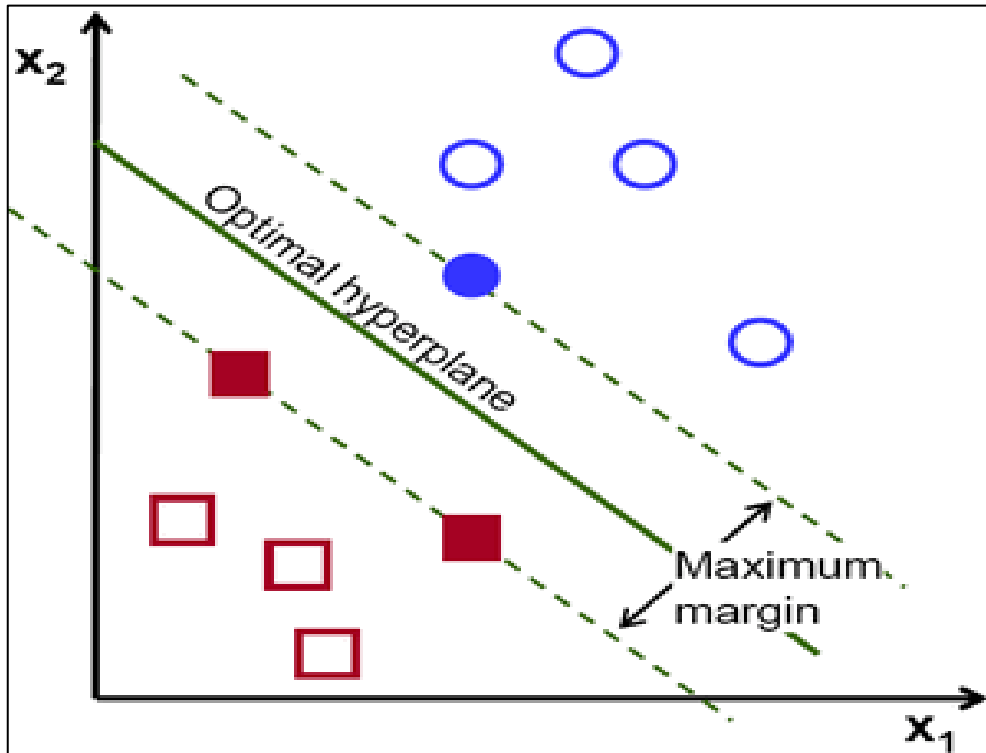
Figure 2.8: Selected optimal hyper-plane of SVM classification analysis (Source: https://cdn images1.medium.com/max/750/0*0o8xIA4k3gXUDCFU.png.
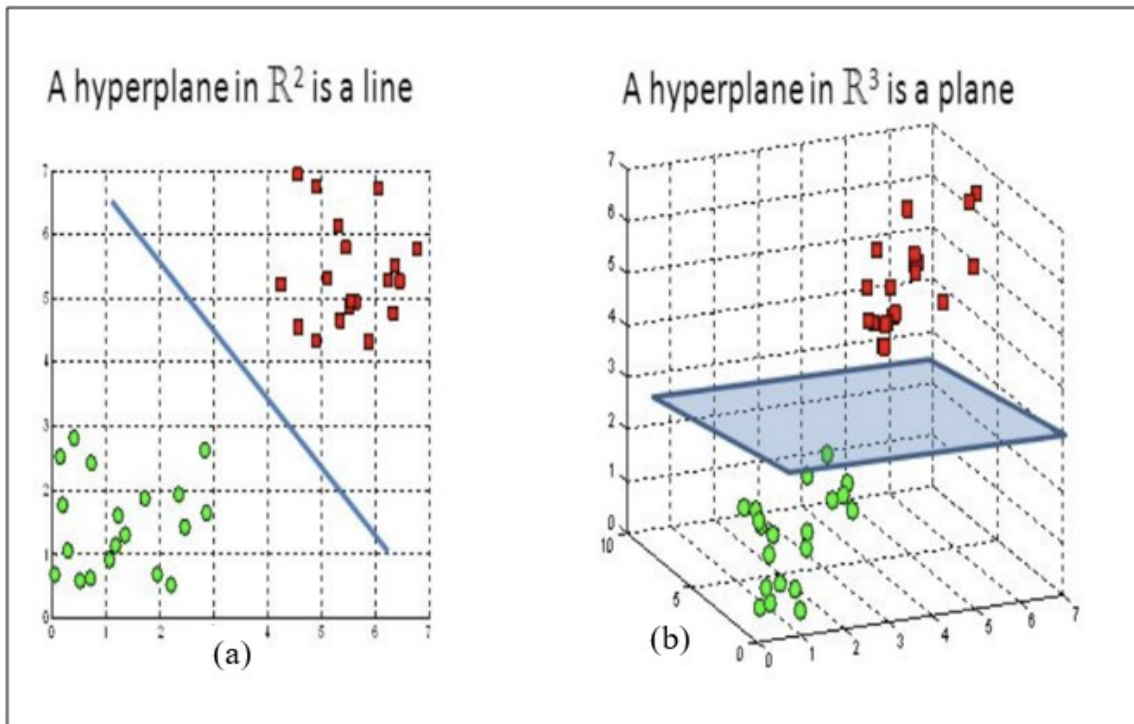
Figure 2.9: Hyperplanes in (a) 2D and (b) 3D feature space (Source: https://cdn-images-1.medium.com/max/2000/1*ZpkLQf2FNfzfH4HXeMw4MQ.png).

If the number of input features is 3, then the hyperplane becomes a two-dimensional plane (Figure 2.9b). It becomes difficult to imagine when the number of features exceeds 3.

2.6.2.3 Support Vectors

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, it is maximized the margin of the classifier (Figure 2.10). Deleting the support vectors will change the position of the hyperplane. These are the points that help to build the SVM.
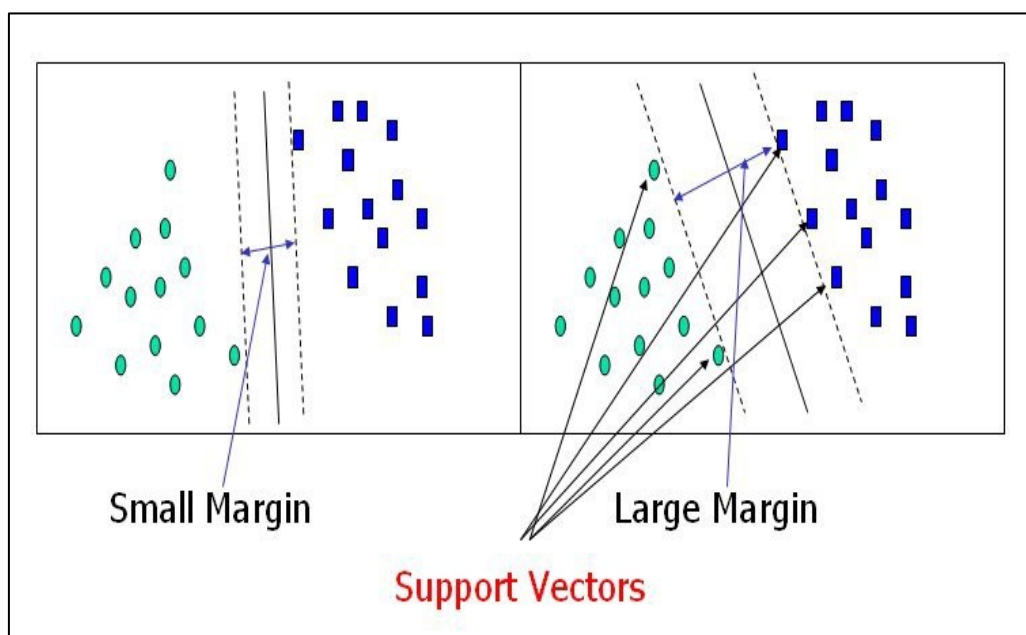
Figure 2.10: Support vectors margin (Source: https://cdn-images-1.medium.com/max/1600/0*ecA4Ls8kBYSM5nza.jpg).

2.6.2.4 SVM Regression

Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences. First of all, being output is a real number; it becomes very difficult to predict the information at hand, which has infinite possibilities.
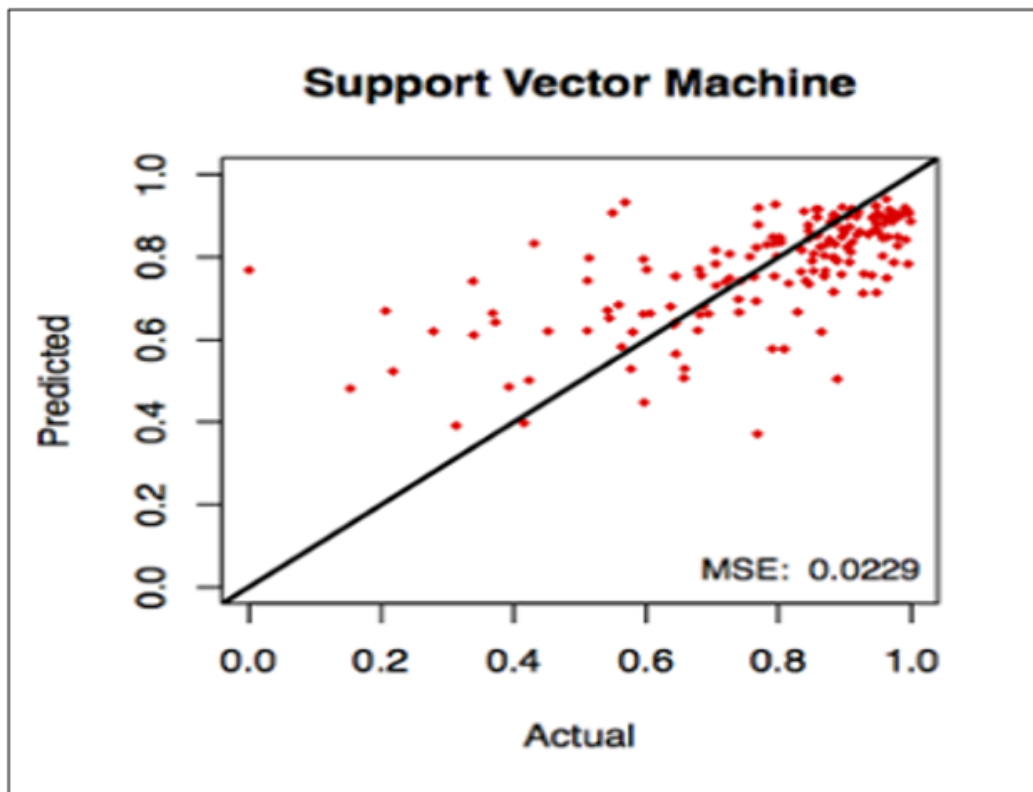
Figure 2.11: Support vectors margin (Source: https://cdn-images-
1.medium.com/max/1600/0*ecA4Ls8kBYSM5nza.jpg).

In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM, which would have already requested from the problem. Besides this fact, there is also a more complicated reason; the algorithm is more complicated therefore to be taken in consideration. However, the main idea is always the same to minimize error, individualizing the hyperplane that maximizes the margin, keeping in mind that part of the error is tolerated (https://www.saedsayad.com/support_vector_machine_reg.htm). Figure 2.11 shows the SVM regression for the actual and predicted data.

2.6.2.5 Modelling of SVM

The support vector machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labelled training data. The main task of support vector machine for regression model is to analyse the model with various kernel functions and then provide a predicted value against measured data. The model produced by Support Vector Regression has to be close to the prediction line shown in Figure 2.12. Moreover in this modelling different kernel functions can be specified for the decision function and so SVMs

are also said to be a "kernel methods" (Wang, 2005). For avoiding over-fitting, choosing kernel functions, number of cross validation and regularization term is essential. SVM do not directly provide probability estimation, these are calculated using a particular number of fold for cross-validation (Smola and Schölkopf, 2004).
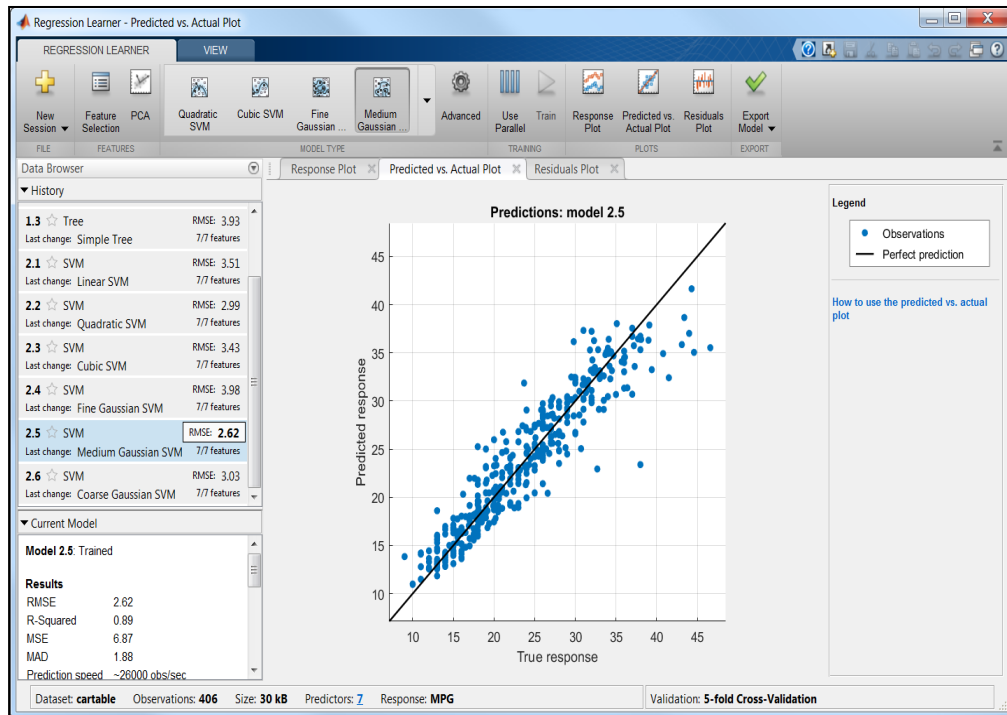
Figure 2.12: SVM Regression model in MATLAB (Source: Wang, 2005).

2.6.2.6 Application of SVM

SVMs have demonstrated highly competitive performance in numerous real-world applications such as bio-informatics, text mining, face recognition and image processing which has established SVM as one of the state-of- the-art tools for machine learning and data mining, along with other soft computing techniques, e.g., neural networks and fuzzy systems (Wang, 2005). SVM has also been applied for development of prediction models, in geotechnical engineering. SVM has been applied for settlement prediction of foundations on cohesionless soils (Samui, 2008), swelling pressure of expansive soils (Das et al., 2010), MDD and unconfined compressive strength of stabilized soils (Das et al., 2011), liquefaction of soils (Lee and Chern, 2013), angle of shearing resistance of soils (Roy and Dass, 2014), prediction of the concentration of Ni and Fe (Gholami et al., 2011). Prediction of heavy metals in soils using SVM is limited in literature.

2.6.2.7 Kernel Function of SVM

The kernel function determines the correlation in the response as a function of the distance between the predictor values. The SVM with different kernel functions like linear support vector machine (SVM-L), quadratic support vector machine (SVM-Q), cubic support vector machine (SVM-C) and gaussian or radial basis function of support vector machine (SVM-RBF) is used to predict the heavy metal concentrations in soils. From the MATLAB 2017a, it can get use from "Regression Learner App" to train the regression models. After performing the models for predictor and response data, it provides the value of R, R², MAE, MSE and RMSE value to compare with each other of SVM-L, SVM-Q, SVM-C and SVM-RBF.

2.6.2.7.1 Linear Support Vector Machine

The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra. This is where the kernel plays role. For **linear kernel,** the equation for prediction of a new input using the dot product between the input (x) and each support vector is calculated by the following Equation (7).

$$f(x) = B_0 + \sum \{a_i (x.x_i)\} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (7)$$

This equation involves calculating the inner products of a new input vector (x) with all support vectors in training data. The coefficients B0 and ai (for each input) must be estimated from the training data by the learning algorithm (Ghadimi, 2014).

2.6.2.7.2 Quadratic Support Vector Machine

A new quadratic kernel-free non-linear support vector machine (which is called SVM-Q) is introduced. The SVM optimization problem can be stated as follows: Maximize the geometrical margin subject to all the training data with a functional margin greater than a constant. The functional margin is equal to WTX + b, which is the equation of the hyper-plane used for linear separation. The geometrical margin is equal to1‖W‖. In addition, the constant in this case is equal to one. To separate the data non-linearly, a dual optimization form and the Kernel trick must be used. In this paper, a quadratic decision function that is

26

capable of separating non-linearly the data is used. The geometrical margin is proved equal to the inverse of the norm of the gradient of the decision function. The functional margin is the equation of the quadratic function. SVM-Q is proved to be put in a quadratic optimization setting. This setting does not require the use of a dual form or the use of the Kernel trick (https://link.springer.com/article/10.100, 2018).

2.6.2.7.3 Cubic Support Vector Machine

The Cubic SVM is a kernel function commonly used with support vector machines (SVM) and other kernel zed models, that represents the similarity of vectors (training samples) in a feature space over cubic of the original variables, allowing learning of non-linear models. Cubic SVM looks not only at the given features of input samples to determine their similarity, but also combinations of these. In the context of regression analysis, such combinations are known as interaction features. The (implicit) feature space of a cubic kernel is equivalent to that of cubic regression, but without the combinatorial blow up in the number of parameters to be taught (Chang et al., 2010)

2.6.2.7.4 Radial Basis Function of Support Vector Machine

The RBF is by far the most popular choice of kernel types used in support vector machines. This is mainly because of their localized above equation is taken for each class variables and finite responses across the entire range of the real x-axis. A radial basis function is defined by the following Equation (8).

$$Z(x) = \Phi(\| x - \mu \|) \dots \dots \dots \dots \dots \dots \dots . . \dots \dots \dots \dots \dots \dots (8)$$

Where x is an n-dimensional vector, μ is an n-dimensional vector called the centre of the radial basis function, $\|. \|$ denotes Euclidean distance, and is a univariate function, defined for positive input values that would be referred to as the profile function.

The model is built up as a linear combination of N radial basis functions with N distinct centers. By giving an input vector x, the output of the RBF network is the activity vector $\hat{y}$ given by the following Equation (9).

$$\hat{y}(x) = z \sum_{j=1}^{N} \beta_j \, z_j(x) \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (9)$$

Where β_j is the weight associated with the jth radial basis function, centered at μ_j, and $z_j = \Phi(\parallel x - \mu_j \parallel)$. The output $\hat{y}$ approximates a target set of values denoted by y.

The different kernel functions of SVM with its characteristics are provided in Table 2.3.

Table 2.3: Description of different kernel functions of SVM (After: https://in.mathworks.com/help/stats/choose-a-classifier.html, 2018)

Classifier type	Prediction speed	Memory usage	Interpretability	Model flexibility
Linear SVM	Binary: Fast Multiclass: Medium	Medium	Easy	Low Makes a simple linear separation between classes.
Quadratic SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	Medium
Cubic SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	Medium
RBF SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	High Decreases with kernel scale setting. Makes finely detailed distinctions between classes, with kernel scale set to sqrt (P)/4.

2.6.3 Artificial Neural Network

Artificial neural networks (ANN) are one of the main tools used in machine learning systems vaguely inspired by the biological neural networks that constitute human brains. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use. They are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize. (https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/). ANNs are also capable of classifying

patterns, clustering, approximating functions, forecasting, optimizing results and controlling inputs such that a system follows the desired trajectory of solving problems by using, modifying and extrapolating acquired knowledge.

2.6.3.1 Historical Overview

ANN is a type of Artificial Intelligence technique that mimics the behaviour of the human brain (Haykin, 2009). A study conducted by Choobbasti et al., (2015) and stated that over the last few years or so, the use of ANN has increased in many areas of engineering. In particular, ANN has been applied to many geotechnical engineering problems and have demonstrated some degree of success. A review of the literature reveals that ANN has been used successfully in pile capacity prediction, modelling of soils behavior, site characterization, settlement of structures, liquefaction, soils permeability and hydraulic conductivity, soils compaction, soils swelling and classification of soils (Rooki et al., 2011). (Choobbasti et al., 2015) found that the Marquardt-Levenberg method, based on Gauss-Newton's equations could be used for training the network with a neuron in the output layer to minimize the square sum of the nonlinear objective function. It was found that training the NN models by using this algorithm to converge in several cases where training by using other algorithms are failed to do so (Hagan and Menhaj, 1994). (Rooki et al., 2011) made a research on prediction of heavy metal in acid mine drainage using ANN. (Alkaiem and Sternberg, 2016) used of artificial intelligence techniques to predict distribution of heavy metals in groundwater of Lakan lead-zinc mine in Iran. In that study, the ANN was developed to estimate the heavy metals concentrations in groundwater using SO4, Cl, and TDS as input parameters, and Fe, Mn, Pb, and Zn as output parameters and the performance of ANN was satisfactory. In the present study, ANN was used to predict values of indices of using its latitude and longitude only as well as the comparison of predicted values of indices are made to check the accuracy of the ANN model.

2.6.3.2 Network Function

Network Function – a functional building block within a network infrastructure, which has well-defined external interfaces and a well-defined functional behaviour. In practical terms, a Network Function is today often a network node or physical appliance. Mathematically, a neuron's network function $(x) \, gi$ is defined as a composition of other functions $g_i(x)$ which

can further be defined as a composition of other functions. This can be conveniently represented as a network structure, with arrows depicting the dependencies between variables. A widely used type of composition is the nonlinear weighted sum which represented by the following Equation (10).

$$f(x) = K \sum_{i=0}^{i} wigi(x) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (10)$$

Where K is some predefined function, such as the hyperbolic tangent. A typical Network function is shown in Figure 2.13.
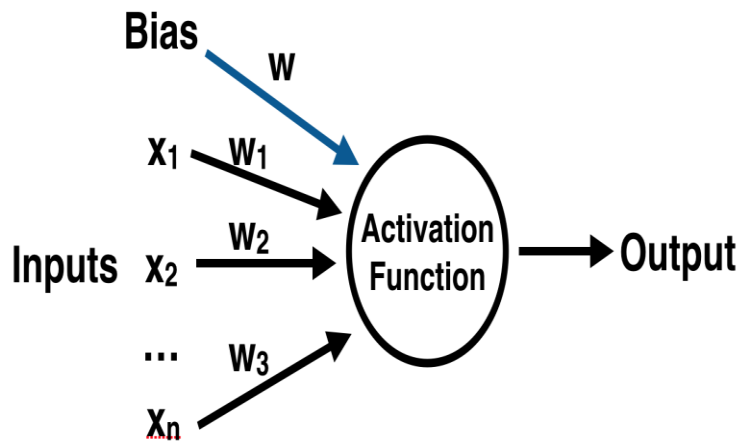
Figure 2.13: A typical Network Function (Source: https://i.stack.imgur.com/iIcbq.gif).

2.6.3.3 ANN Modelling

ANNs can describe nonlinear and complex relationships using a part of the input and output training patterns from the dataset. These approaches establish a non-linear relationship between inputs and outputs (Hornik et al., 1990). An ANN can be demonstrated based on architecture that shows the connection pattern between nodes, connection weights method determination, and the activation function (Kişi, 2008). Because of their ability to learn a system's dynamics from data, ANNs are able to solve large-scale complex problems. The most commonly used neural network architecture is the feed-forward neural network (FFNN). The structure of a three-layered FFNN is based on some neurons in each layer and

elements, which link them (Markopoulos, 2008). The training of a network is based on the optimization process for weights to obtain the appropriate weights to minimize errors; this process continues until the values of the output layer are as close as possible to the actual outputs (Hornik et al., 1990). In this study, the LM, OSS and SCG training algorithms as well as transig, logsig and purelin transfer functions were utilized to tune the weights. Figure 2.14 shows the feed-forward network for this study, having one hidden layer with several nodes between the input and output layers.
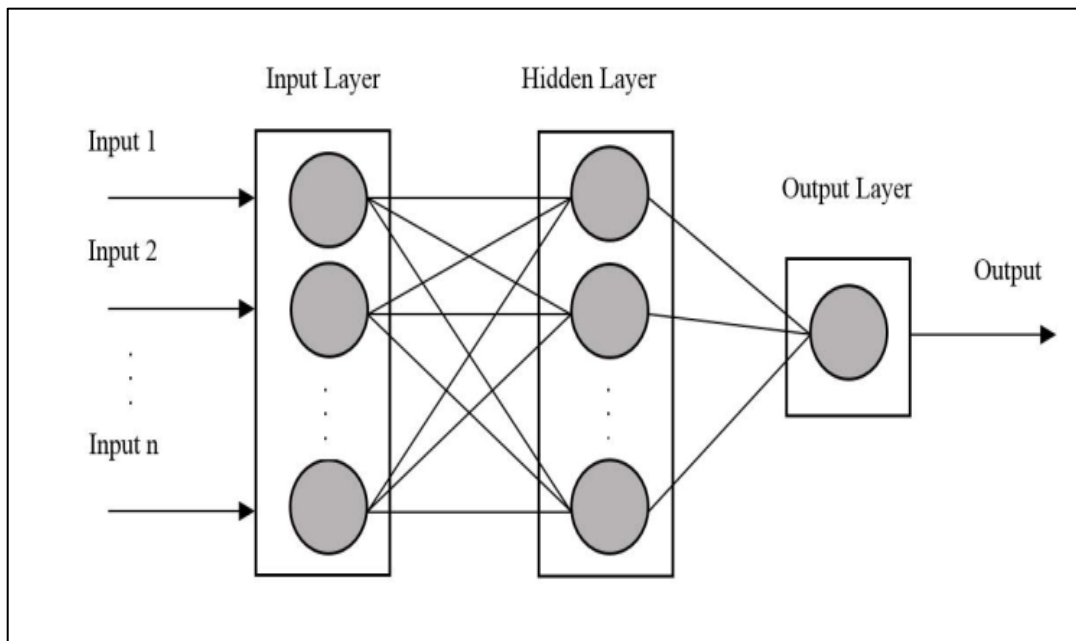
Figure 2. 14: A typical Artificial Neural Network model (Source: Alizamir & Sobhanardakani, 2017).

a) Input layer - It contains those units (Artificial Neurons) which receive input from the outside world on which network will learn, recognize about or otherwise process.

b) Output layer - It contains units that respond to the information about how it's learned any task.

c) Hidden layer - These units are in between input and output layers. The job of hidden layer is to transform the input into something that output unit can use in some way.

2.6.3.4 Training Functions of ANN Model

Levenberg-Marquardt Neural Network (LMNN), One Step Secant Neural Network (OSSNN), and Scaled Conjugate Gradient Neural Network (SCGNN) algorithms are used

in training processes for analysis of ANN model called as LMNN model, OSSNN model and SCGNN model respectively.

2.6.3.4.1 Levenberg- Marquardt Neural Network

The levenberg-marquardt (LMNN) algorithm adaptively varies the parameter updates between the gradient descent update and the Gauss-Newton update. The relationship is showing in Equation (11).

$$[J^T W J + \lambda I] h_{lm} = J^T W (y - \hat{y}) \dots \dots \dots \dots \dots \dots (12)$$

Where small values of the algorithmic parameter λ result in a Gauss-Newton update and large values of λ result in a gradient descent update. The parameter λ is initialized to be large so that first updates are small steps in the steepest-descent direction. If any iteration happens to result in a worse approximation($X^2(p + h_{lm}) > X^2(p)$), then λ is increased. Otherwise, as the solution improves, λ is decreased, the Levenberg-Marquardt method approaches the Gauss-Newton method, and the solution typically accelerates to the local minimum (Suri, 2017). In Marquardt's update, relationship is showing in Equation (12).

$$[J^T W J + \lambda \, diag(J^T W J)] h_{lm} = J^T W (y - \hat{y}) \dots \dots . (12)$$

The values of λ are normalized to the values of $J^T W J$. The Levenberg-Marquardt algorithm implemented in the Matlab function lm.m.

2.6.3.4.2 One Step Secant Neural Network

The one-step secant (OSSNN) method can train any network as long as its weight, net input, and transfer functions have derivative functions. Backpropagation is used to calculate derivatives of performance perf with respect to the weight and bias variables X. Each variable is adjusted according to the following Equation (13).

$$X = X + a * dX \qquad \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . \dots (13)$$

Where dX is the search direction. The parameter "a" is selected to minimize the performance along the search direction. The line search function searchFcn is used to locate the minimum

point. The first search direction is the negative of the gradient of performance. In succeeding iterations, the search direction is computed from the new gradient and the previous steps and gradients, according to the following Equation (14).

$$dX = -gX + Ac^*X_{step} + Bc^*dgX \ \dots\dots\dots\dots\dots (14)$$

Where gX is the gradient, Xstep is the change in the weights on the previous iteration, and dgX is the change in the gradient from the last iteration (Battiti and Tecchiolli, 1995).

2.6.3.4.3 Scaled Conjugate Gradient Neural Network

The scaled conjugate gradient (SCGNN) methods are also based on the above general optimization strategy, but chooses the search direction and the step size more carefully by using information from the second order approximation by the following Equation (15).

$$E(w + y) \approx E(w) + E'(w)^T y + \frac{1}{2}y^T E''(w)y \ \dots\dots\dots (15)$$

Quadratic functions have some nice properties that general functions not necessarily have. Denoting the quadratic approximation to E in a neighbourhood of a point w by $E_{qw}(y)$, so that $E_{qw}(y)$ is given by the following Equation (16).

$$E(w + y) \approx E(w) + E'(w)^T y + \frac{1}{2}y^T E''(w)y \ \dots\dots\dots (16)$$

In order to determine minima to $E_{qw}(y)$ the critical points for $E_{qw}(y)$ must be found. The critical points are the solution to the linear system defined by the following Equation (17).

$$E'_{qw}(y) = E''(w)y + E'(w) = 0 \ \dots\dots\dots\dots\dots\dots\dots\dots (17)$$

If a conjugate system is available, the solution can be simplified considerable. (Johansson, Dowla, and Goodman, 1991) shows in a very understandable way how.

Let p1…. pN be a conjugate system. Because p1,…, pN form a basis for $\Re$N, the step from a starting point y1 to a critical point y* can be expressed in Equation (18) as a linear combination of p1,…, pN.

$$y_* - y_1 = \sum_{i=1}^{N} \alpha_i p_i, \alpha_i \in \Re \ \dots\dots\dots\dots\dots\dots\dots\dots\dots (18)$$

Multiplying (18) with $p_j^T E''(w)$ and substituting E'(w) for $- E''(w) y_*$ gives the following Equation (19).

$$p_j^T(-E'(w) - E(w)y_1) = \alpha_j p_j^T E(w)p_j$$

$$\Rightarrow \quad \alpha_j = \frac{p_j^T(-E'(w) - E''(w)y_1)}{p_j^T E''(w)p_j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (19)$$

The critical point y_* can be determined in N iterative steps using (18) and (19). Unfortunately y_* is not necessarily a minimum, but can be a saddle point or a maximum. Only if the Hessian matrix E'' (w) is positive definite then $E_{qw}(y)$ has a unique global minimum. This can be realized by the following Equation (20).

$$E_{qw}(y) = E_{qw}\big(y_* + (y - y_*)\big)$$

$$= E(w) + E'(w)^T\big(y_* + (y - y_*)\big) + \frac{1}{2}\big(y_* + (y - y_*)\big)^T E''(w)\big(y_* + (y - y_*)\big)$$

$$= E(w) + E'(w)^T y_* + E'(w)^T(y - y_*) + \frac{1}{2}y_*^T E''(w)y_* + \frac{1}{2}y_*^T E''(w)(y - y_*)$$

$$+ \qquad \frac{1}{2}(y - y_*)^T E''(w)y_* + \frac{1}{2}(y - y_*)^T E''(w)(y - y_*)$$

$$= 3E_{qw}(y_*) + (y - y_*)^T\big(E''(w)y_* + E'(w)\big) + \frac{1}{2}(y - y_*)^T E''(w)(y - y_*)$$

$$= 4E_{qw}(y_*) + \frac{1}{2}(y - y_*)^T E''(w)(y - y_*) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (20)$$

It follows from (20) that if y_* is a minimum then $\frac{1}{2}(y - y_*)^T E''(w)(y - y_*) > 0$ for every y, hence E''(w) has to be positive definite. The Hessian E'' (w) will in the following if not told otherwise be assumed to be positive definite. The intermediate points $y_{k+1} = y_k + \alpha_k p_k$ given by the iterative determination of y_* are in fact minima for $E_{qw}(y)$ restricted to every k-plane in Equation (21).

$$\pi_k : y = y_1 + \alpha_1 p_1 + \cdots + \alpha_k p_k \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

The description of various training functions such as LMNN, OSSNN and SCGNN of ANN is provided in Table 2.4.

Table 2.4: Description of various training functions of ANN (After: Matlab Documentation, 2017a)

Neural network algorithm	Full meaning	Required memory or time	Stop of the training time
LMNN	Levenberg-Marquardt Neural Network	More memory and less time	Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples
OSSNN	One Step Secant Neural Network	More time	Training stops according to adaptive weight minimization
SCGNN	Scaled Conjugate Gradient Neural Network	Less memory	Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples

2.6.3.5 Transfer Functions of ANN Model

Transfer functions are commonly used in the analysis of systems such as single-input single-output filters in the fields of signal processing, communication theory, and control theory. That is usually the case for signal processing and communication theory. Transfer functions also calculate a layer's output from its net input. Three types of transfer functions are used in ANN model named hyperbolic tangent sigmoid transfer function (tansig), linear transfer function (purelin) and Log-sigmoid transfer function (logsig). These are briefly discussed in the following article.

2.6.3.5.1 Tangent Sigmoid Transfer Function

The tangent sigmoid (tansig) is a transfer algorithm which represented by the following Equation (22).

$$a = tansig(n) = \frac{2}{1 + exp(-2 * n)} - 1 \quad ...\,...\,... (22)$$

This is mathematically equivalent to tanh (N). It differs in that it runs faster than the MATLAB implementation of tanh, but the results can have very small numerical differences. This function is a good trade-off for neural networks, where speed is important but not the exact shape of the transfer function (Vogl et al., 1988).

2.6.3.5.2 Linear Transfer Function

Linear transfer function (purelin) is typically used for function approximation or regression tasks. This is intuitive because step and logistic functions give binary results where the linear function gives continuous results. The algorithm of purelin is represented by the following Equation (23).

$$a = purelin(n) = n \quad ...\,...\,...\,...\,...\,...\,...\,...\,...\,...\,...\,... (23)$$

2.6.3.5.3 Log-sigmoid Transfer Function

Logarithm of sigmoid (logsig) states, it is a modified version. It produces outputs in scale of (-∞, 0]. The algorithm of logsig is represented by the following Equation (24).

$$logsig(n) = \frac{1}{1 + exp(-n)} \quad ...\,...\,...\,...\,...\,...\,...\,...\,...\,...\,...\,.. (24)$$

2.7 Prediction parameters of AI Techniques

The testing data (unknown 15 sampling points) was assessed with the following prediction parameters.

2.7.1 Correlation Coefficient

A correlation coefficient (R) is a statistical measure of the degree to which changes to the value of one variable predict change to the value of another. Simply, a correlation coefficient is the statistical measure of the linear relationship (https://whatis.techtarget.com/definition/correlation -coefficient) between a dependent variable and an independent variable. In positively correlated variables, the value increases or decreases in tandem. In negatively correlated variables, the value of one increases and the value of the other decreases. The "R" represents it that displays in the following Equation (25).

$$R = \frac{n(\sum y \cdot y_p)\ (\sum y)(\sum y_p)}{\sqrt{[n \sum y^2\ (\sum y)^2][n \sum y_p^2\ (\sum y_p)^2]}} \quad \dots \dots \dots \dots \dots (25)$$

Where y = observed value, yp = predicted value, n = number of observations.

By (Rumsey, 2015) in statistics, the correlation coefficient R measures the strength and direction of a linear relationship between two variables on a scatterplot. The value of R is always between +1 and –1. To interpret its value, it is categorised in the following manner.

1. Exactly –1. A perfect downhill (negative) linear relationship
2. –0.70. A strong downhill (negative) linear relationship
3. –0.50. A moderate downhill (negative) relationship
4. –0.30. A weak downhill (negative) linear relationship
5. 0. No linear relationship
6. +0.30. A weak uphill (positive) linear relationship
7. +0.50. A moderate uphill (positive) relationship
8. +0.70. A strong uphill (positive) linear relationship
9. Exactly +1. A perfect uphill (positive) linear relationship

A research conducted by Smith (1986) suggested that the value of R lies between 0 to 1. It is also suggested some guidelines for deciding the performance of the model. If $|R| \geq 0.8$: a strong correlation exists, $0.2 < |R| < 0.8$: correlation exists and $|R| \leq 0.2$: a weak correlation exists. When the value of $|R|$ is greater than 0.9, then a very strong correlation exists between the variables.

2.7.2 Root Mean Square Error

The root mean square error (RMSE) is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. In other words, it measures the quality of the fit between the actual and the predicted data of a model. It is represented by the following Equation (26).

$$RMSE = \sqrt{\frac{\sum_1^n (y - y_p)^2}{n}} \quad \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (26)$$

Where y = observed value, y_p = predicted value, n = number of observations.

RMSE is one of the most frequently used measures of the goodness of fit of generalized regression models. According to Schweizer (2010), lower values of RMSE indicate better fit and zero means no error. RMSE is a good measure of how accurately the model predicts the response, and it is the most important criterion for fit if the main purpose of the model is prediction.

2.7.3 Mean Absolute Percentage Error

The mean absolute percent error (MAPE) measures the size of the error in percentage terms. It is calculated as the average of the unsigned percentage error, as shown in the following Equation (27):

$$MAPE = \frac{1}{n} \sum \frac{|y - y_p|}{|y|} \times 100\% \quad \dots \dots \dots \dots \dots \dots \dots . (27)$$

Where y = observed value, yp = predicted value, n = number of observations.

For the MAPE, extreme values occur only at the high end because it is typically based on a right-skewed distribution of absolute percent errors (APE) bounded on the left by zero and unbounded on the right. In a comprehensive analysis of county-level projections, the MAPE was on average higher by about 30–40% than robust measures of central tendency for most methods and projection horizons (Rayer, 2007).

2.7.4 Geometric Reliability Index

A version of the geometric reliability index (GRI) was defined as the inverse of the coefficient of variation. The reliability index is the shortest distance from the origin of reduced variables (Hasofer and Lind, 1974). Using geometry, the reliability index can be measured by the following Equation (29).

$$GRI = \frac{1 + \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\frac{\hat{y}_t - y_t}{\hat{y}_t + y_t})^2}}{1 - \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\frac{\hat{y}_t - y_t}{\hat{y}_t + y_t})^2}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\ (29)$$

Where y_t = observed value, $\hat{y}_t$ = predicted value, n = number of observations.

According to Leggett and Williams (1981), GRI is a statistical method to determine the reliability of a model. The index is a number GRI ≥ 1. The formula of geometry index also expresses that when the error will be zero, then the index value will be 1. In contrary, the more error alienate the index value from 1. Therefore, it can be said that GRI value must be greater or equal 1 and GRI value 1 represents the perfectness and reliability of model.

2.7.5 Percent Recovery

The percentage recovery means what percentage of measured value is recovered by the predicted value. It is represented by the following Equation (28).

$$Persent\ Recovery = \frac{y_p}{y} \times 100 \dots\dots\dots\dots\dots\dots\dots\ (28)$$

Where y = observed value, y_p = predicted value.

The importance of recovery can be found in the International Council for Harmonisation (ICH) guidelines as well as in the Good Laboratory Practice (GLP) guidelines for analytical method validation (Branch, 2005). According to Walfish (2006), recoveries in the range of 20-200% for internal standard are considered 'acceptable' (depending on the jurisdiction). Food and Drug Administration (FDA), Investopedia declares that recovery should not need to be 100% but should be reproducible. FDA approved variability limit for Lower Limit of Qualification (LLOQ) is +/- 20%. Therefore, the ideal frame of recovery is 80-120% that represents the robustness of the model.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 General

This chapter deals with the overall research methodology in this study. This chapter includes the information about the study area and soils condition. In this study, total eighty-five disturbed soils samples were considered from distinct locations in and around the waste disposal site at old Rajbandh in Khulna. The method of soils sampling was also highlighted in this chapter. In the laboratory, the concentrations of relevant heavy metal in soils were measured and monitored through standard test methods and hence highlighted in this chapter. Then, the measured concentrations of heavy metals were used to perform the descriptive conventional statistics using MS Excel to assess the basic features of soils data in a simpler. The AI techniques such as ANFIS, SVM and ANN were performed to predict the heavy metal concentrations, which were also highlighted in this chapter.

3.2 Description of the Study Site

Khulna is the third largest established metropolitan city after Dhaka and Chittagong in Bangladesh. It is located in the Khulna Division. It has an area of 4394.45 km² and is bordered on the north by the Jashore and the Narail district, on the south by the Bay of Bengal, on the east by the Bagerhat District, and on the west by the district. The geological location of Khulna is 22.350N and 89.300E, surrounded by Rupsa, Arpangachhia, Shibsa, Pasur, and the Koyra. Urban development is dribbling into neighbouring zones to the North and West results a huge amount of municipal solid waste (MSW) generation. The areas of KCC and KCPA are 45.65sq.km and 69.50sq.km, respectively (https://en.wikipedia.org/wiki/Khulna).The increasing population in Khulna city tends to dispose increasing amount of MSW as well as liquid waste termed as leachate. These MSW contain a large amount of metal elements, which get direct contact to the environment. This may result a great thread to the environment and human health. The selected waste disposal site, old Rajbandh is the only certified waste dumping site of Khulna shown in Figure 3.1. Based on above-mentioned authenticities, it has become

inevitable of comprehensive study of distribution of heavy metals in soils ascends approximately the Rajbandh waste disposal site.
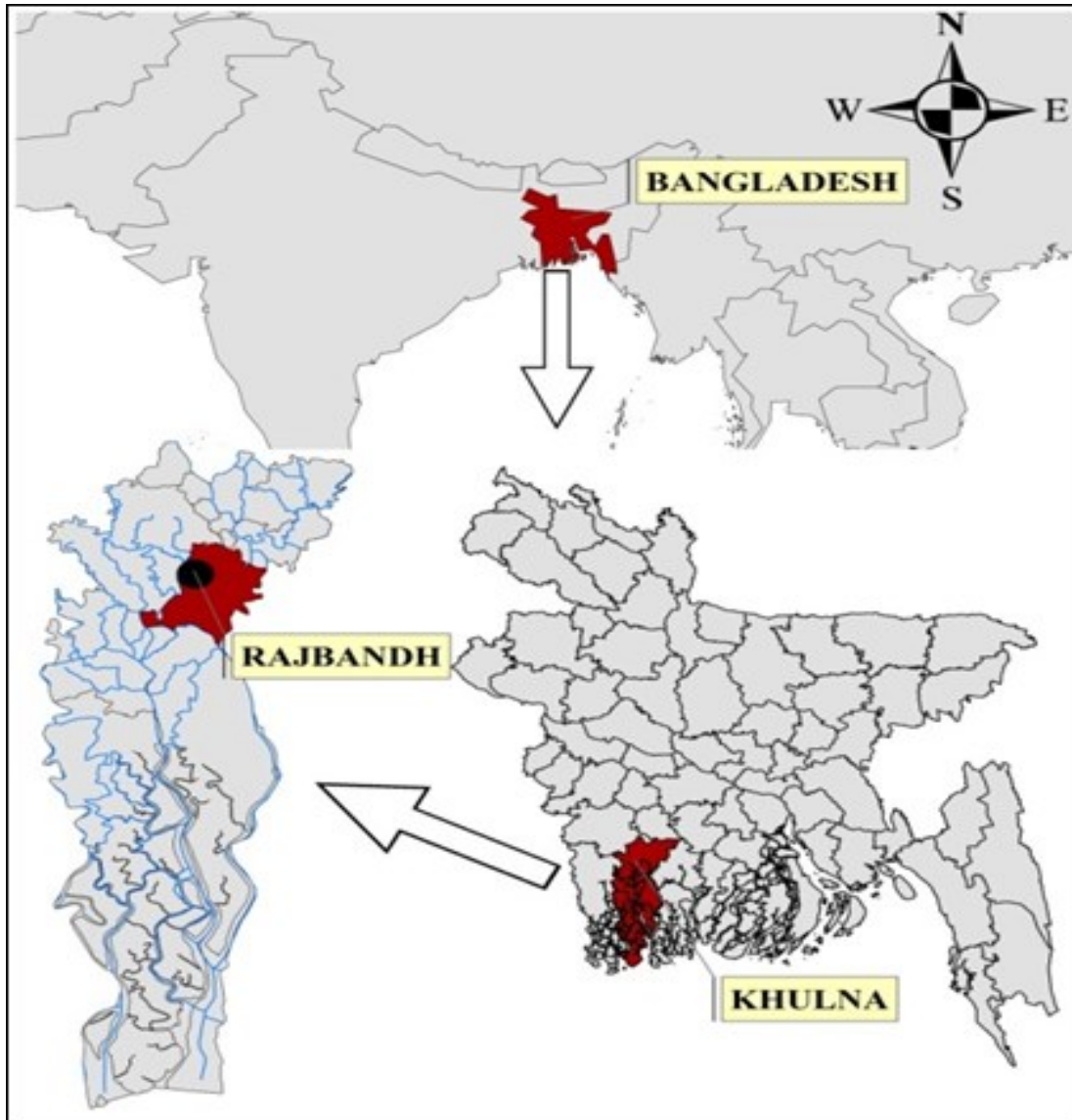

Figure 3.1: Location map of Rajbandh at Khulna city of Bangladesh.

(Source: Fahmida and Rafizul, 2017)

3.3 Location and Soils Conditions of Waste Disposal Site

The waste disposal site located at old Rajbandh, Khulna with an area of 5 acres, is 8km far from the city centre i.e. Royal and Castle Salam Square of Khulna city and situated along the North-

side of Khulna-Satkhira highway. The percolation and seepage capacity of leachate from MSW in disposal sites depends mainly on the basic characteristics of the soils (Daniel and Koerner, 1995). It is therefore important to know the physical and mechanical properties of underlying soils as thoroughly as possible before assessing their physio-chemical or hydro-mechanical behavior. Based on these concepts, in the laboratory through standard ASTM (2004) methods, some relevant physical and mechanical properties of soils from this selected disposal site were determined by soils moisture content, plastic limit, liquid limit, plasticity index and shrinkage limit of clay soils used as CCL were found 22, 22, 43, 21 and 16 %, respectively. In addition, the percentages of soils constituents were found as sand, silt and clay of 10, 56.6 and 33.4%, respectively. Then the value of soils pH, optimum moisture content, maximum dry density and coefficient of permeability were found 6.7, 18 %, 16 kN/m^3 and 1.90x10-7 cm/sec, respectively (Sanjida and Rafizul, 2018).

3.4 Soils Sampling

In this study, total eighty-five disturbed soils samples were considered from distinct locations in and around the waste disposal site at old Rajbandh in Khulna (Figure 3.2). Among the samples, some (twenty) were collected from secondary sources (Sanjida and Rafizul, 2018). These samples were collected at a depth of 0-30 cm from the existing ground surface of the selected waste disposal site. The latitude and departure of all the soils-sampling locations was recorded using GPS device, which were later imported into a geographic information system (ArcGIS 10.1). The sampling points were selected maintaining gradual addition of about 10 m distance from the 1st sampling point (SS-1) by the subsequent sampling points. The first sampling point, SS-1 is located at the centre of the selected waste disposal site. Proper care was taken to remove any loose material, debris, coarse aggregates from the bottom of the excavated pit. Samples were taken in large polythene bags and eventually transported to the laboratory. Figure 3.2 depicted the soils sampling locations in waste disposal site at old Rajbandh, Khulna. The overall research methodology of this study is illustrated in Figure 3.3.
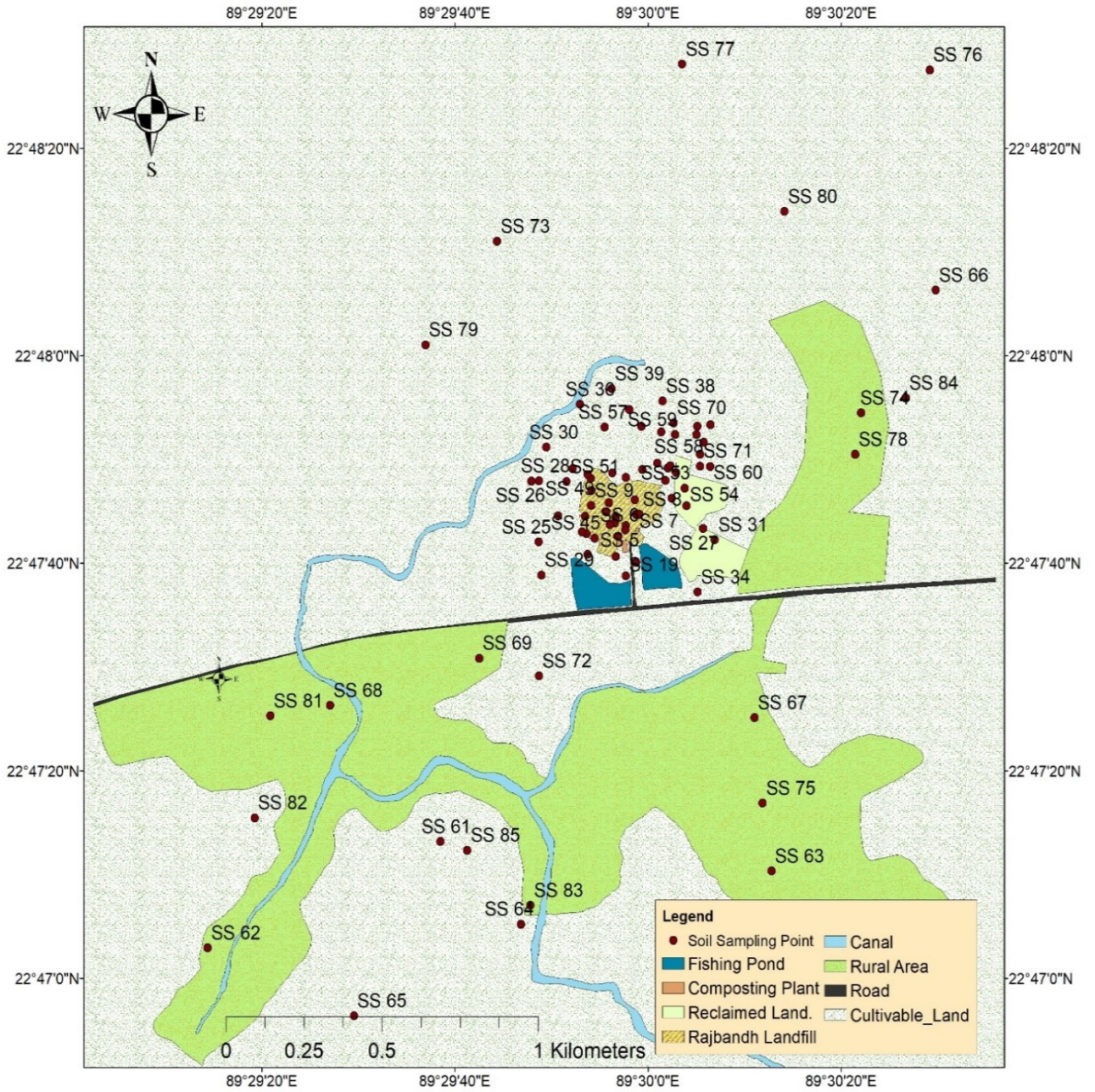
Figure 3.2: Map showing soil-sampling points of the selected waste disposal site at old Rajbandh, Khulna

```
                  ┌─────────────────────────────┐
                  │  Conceptualization of study │
                  └─────────────────────────────┘
                                 │
                                 ▼
                  ┌─────────────────────────────┐
                  │     Inventory of the study  │
                  └─────────────────────────────┘
                                 │
                                 ▼
              ┌─────────────────────────────────────┐
              │ Reconnaissance survey of the study area │
              └─────────────────────────────────────┘
                                 │
                                 ▼
                  ┌─────────────────────────────┐
                  │   Collection of soil samples │
                  └─────────────────────────────┘
                                 │
                                 ▼
                  ┌─────────────────────────────┐
                  │      Collection of data      │
                  └─────────────────────────────┘
```

Figure 3.3: Overall research methodology of this study.

45

3.5 Laboratory Investigations

The soils samples were carried in the laboratory to measure the heavy metal concentrations of lead (Pb), copper (Cu), nickel (Ni), zinc (Zn), cobalt (Co), cadmium (Cd), arsenic (As), scandium (Sc), mercury (Hg), manganese (Mn), chromium (Cr), titanium (Ti), antimony (Sb), strontium (Sr), vanadium (V) and barium (Ba) in soils samples. Moreover, the values of some heavy metal concentrations were collected from secondary sources (Sanjida and Rafizul, 2018). The procedure of acid digestion and atomic absorption spectrophotometer (AAS) analysis are described in the following articles.

3.5.1 Acid Digestion

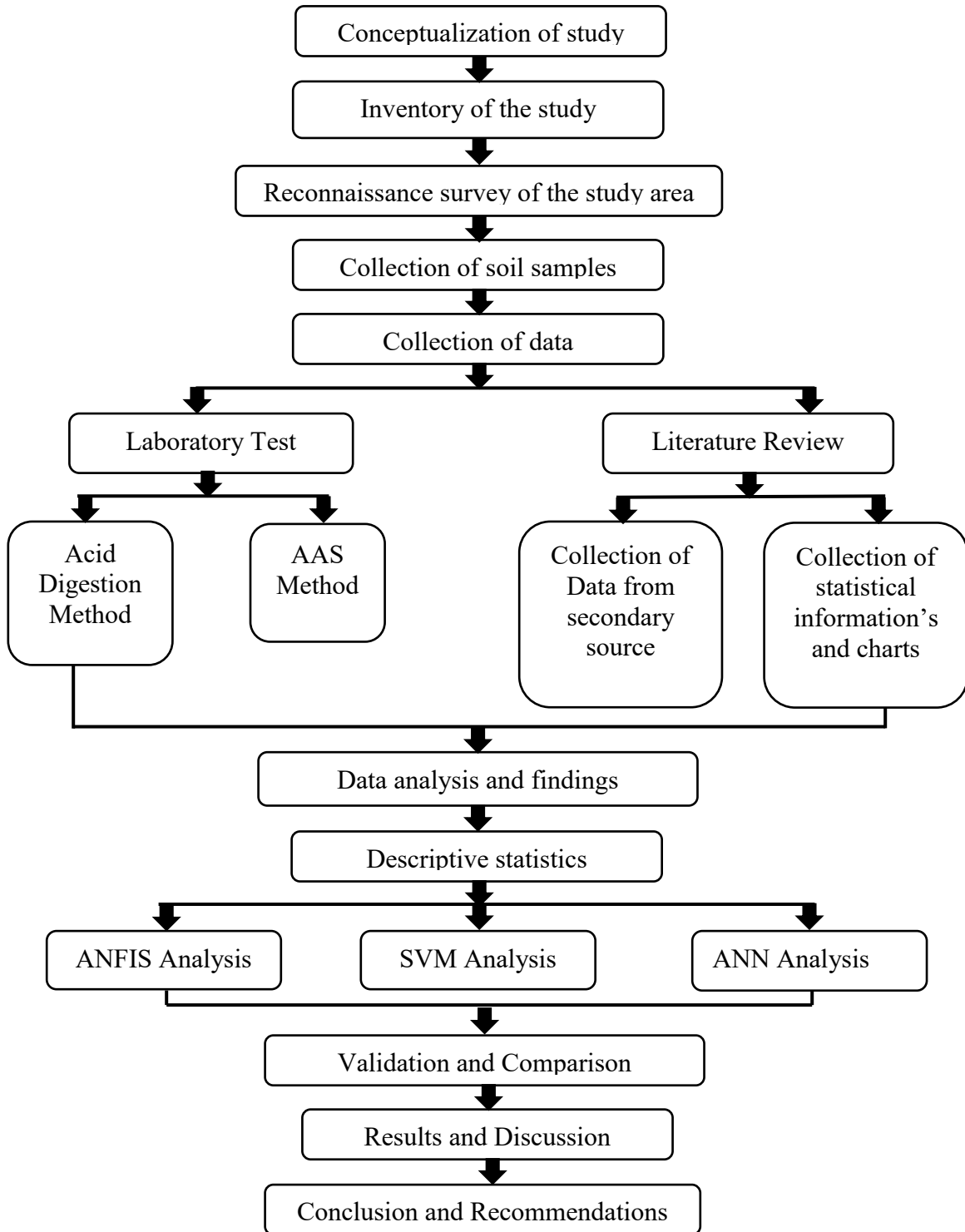To measure the heavy metal concentrations in soils, laboratory work was done following the standard test method. In laboratory investigation, at first 10 g of each soils sample was taken into a 100 mL conical flask. Already, the flask had been washed with deionized water prepared by adding 6 mL $HNO_3/HClO_4$ acid in ratio 2:1 and left overnight. Each sample was kept into the temperature of 150°C for about 90 minutes. Later, temperature was raised to 230°C for 30 minutes. Subsequently, HCl solution was added in ratio 1:1 to the digested sample and re-digested again for another 30 minutes. The digested sample was washed into 100 mL volumetric flask and mixture obtained was cooled down to room temperature.

3.5.2 Analysis of Heavy Metals with AAS

After performing digestion procedure, metal element concentrations in this digested solution were determined using atomic absorption spectroscopy (AAS) and the amount of each heavy metal was deduced from the calibration graph. The concentration of the heavy metals of Pb, Cu, Ni, Zn, Co, Cd, As, Sc, Hg, Mn, Cr, Ti, Sb, Sr, V and Ba in mg/kg were measured in the laboratory.

3.6 Statistical Analysis

The statistical analysis of all studied heavy metal concentrations was performed using MS Excel to know the relative state of the concentrations of heavy metals in soils. In this study, the

statistical analysis in terms of mean, median, maximum, minimum, standard deviation (SD), skewness and kurtosis was described.

3.7 Assessment of Model Performance

The model performance refers the accuracy of prediction. Prediction accuracy means how close predicted concentration to the measured heavy metal concentrations in soils. It can be measured by some prediction parameters like correlation coefficient (R), root mean square error (RMSE), mean absolute percentage error (MAPE), geometric reliability index (GRI) and percentage recovery.

3.7.1 Correlation Coefficient

A research conducted by Smith (1986) suggested that the value of correlation coefficient (R), lies between 0 to 1. It is also suggested some guidelines for deciding the performance of the model. If $|R| \geq 0.8$: a strong correlation exists, $0.2 < |R| < 0.8$: correlation exists and $|R| \leq 0.2$: a weak correlation exists. When the value of $|R|$ is greater than 0.9, then a very strong correlation exists between the variables.

3.7.2 Root Mean Square Error

For the root mean square error (RMSE), lower values of RMSE indicate better fit of model and zero means no error (Schweizer, 2010). It is the most important criterion for fit, which indicates the goodness of prediction model.

3.7.3 Mean Absolute Percentage Error

For the mean absolute percentage error (MAPE), extreme values occur only at the high end because it is typically based on a right-skewed distribution of absolute percent errors (APE) bounded on the left by zero and unbounded on the right. In a comprehensive analysis of county-level projections, the MAPE was on average higher by about 30–40% than robust measures of central tendency for most methods and projection horizons (Rayer, 2007).

3.7.4 Geometric Reliability Index

According to Leggett and Williams (1981), geometric reliability index (GRI) is a statistical method to determine the reliability of a model. The index is a number GRI ≥ 1.

3.7.5 Percentage recovery

The importance of recovery can be found in the International Council for Harmonisation (ICH) guidelines as well as in the Good Laboratory Practice (GLP) guidelines for analytical method validation (Branch, 2005). According to Walfish (2006), recoveries in the range of 20-200% for internal standard are considered 'acceptable' (depending on the jurisdiction). Food and Drug Administration (FDA), Investopedia declares that recovery should not need to be 100% but should be reproducible. FDA approved variability limit for Lower Limit of Qualification (LLOQ) is +/- 20%. Therefore, the ideal frame of recovery is 80-120% that represents the robustness of the model.

3.8 Artificial Intelligence Techniques

In this study, Artificial Intelligence (AI) techniques such as adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and artificial neural network (ANN) were performed through MATLAB to analysis heavy metal concentrations in soils of waste disposal site. The procedure implemented for the analysis of AI techniques are discussed successively in the following articles.

3.8.1 Adaptive Neuro-Fuzzy Inference System

Various models for different heavy metal concentrations was analysed using "Fuzzy Logic Toolbox 2.2.25" named Neuro-Fuzzy Designer. For predicting the heavy metal concentrations in soils through ANFIS, the values of prediction parameters like root mean square error (RMSE) and regression coefficient (R) were considered to check the validity or accuracy of obtained results from laboratory. The operation of ANFIS in details through MATLAB are also discussed in the followings articles.

3.8.1.1 Data Processing

To train ANFIS, a Training data set has been loaded that contained desired input/target data of the system to be modelled. In this study, latitude and longitude were considered as input which placed in first two columns as well as heavy metal concentrations as target which placed at last column for modelling a Training data set. Testing data set also be arranged similarly for the prediction of heavy metals for unknown soils sampling points from the selected waste disposal site. These data sets are saved in the MATLAB workspace that are ready for the ANFIS operation. The rearranged data sets in MATLAB workspace is shown in Figure 3.4.
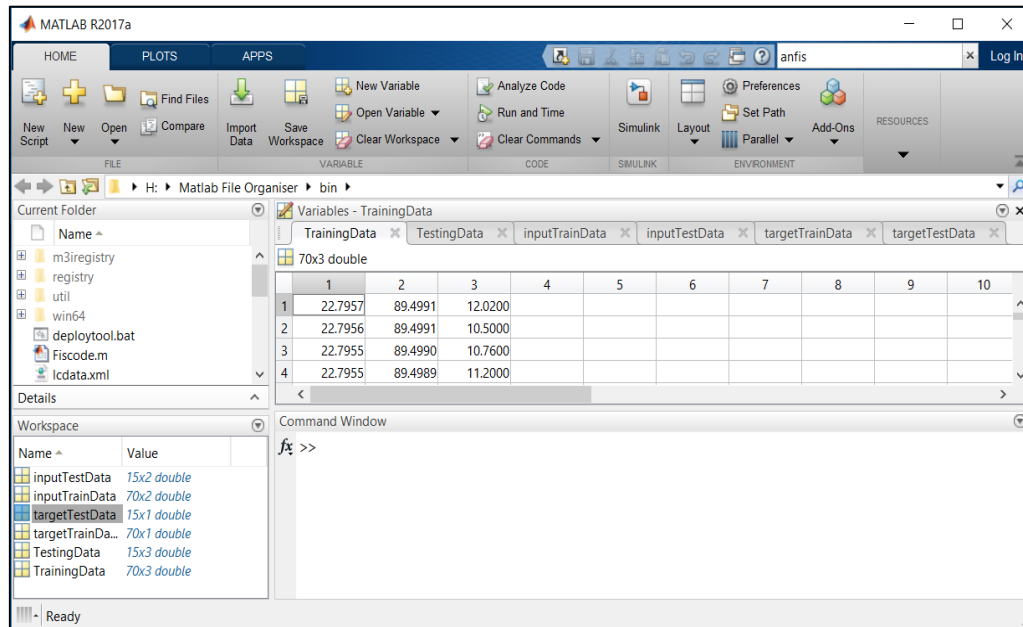

Figure 3.4: Arrangement of data sets in MATLAB workspace.

3.8.1.2 Neuro-Fuzzy Designer App

The following both commands were used to open Neuro-Fuzzy Designer App shown in Figure 3.5 at the MATLAB® prompt:

anfisedit or neuroFuzzyDesigner

In addition, the Neuro-Fuzzy Designer includes four distinct areas to support a typical workflow. The app performs the following tasks:

49

1. Loading, Plotting, and Clearing the Data

2. Generating or Loading the Initial FIS Structure

3. Training the FIS and

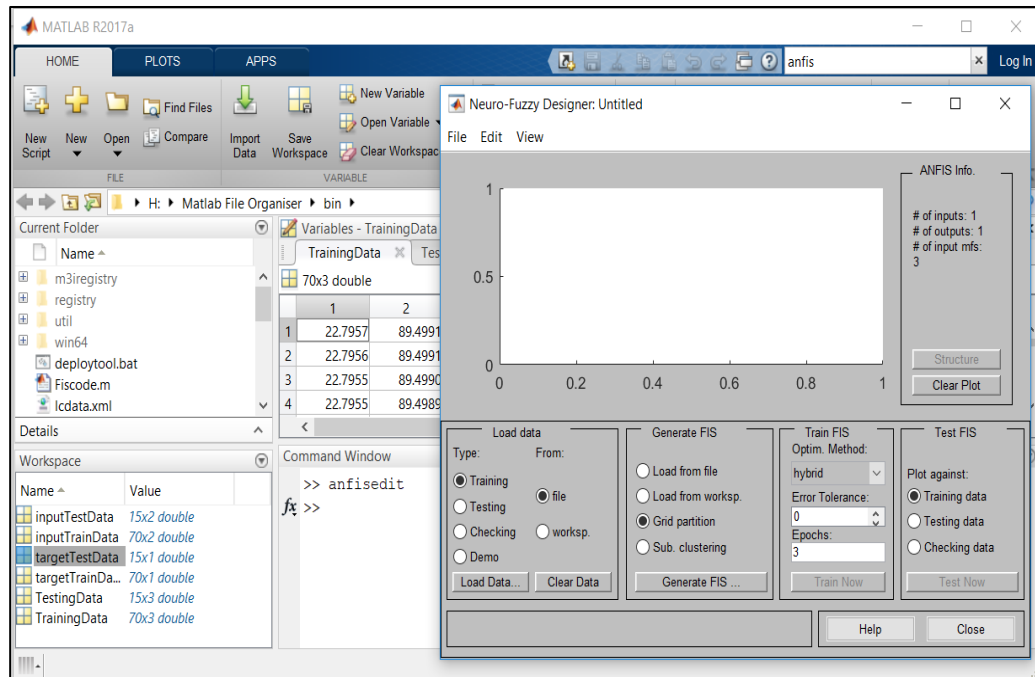4. Validating the Trained FIS or Testing the FIS

Figure 3.5: Commands used to open neuro-fuzzy designer app.

3.8.1.3 Workspace Load Data Portion

For loading training data set from the workspace, Load data portion of the designer was used to select the following options:

Type: Training

From: worksp.

By clicking Load Data option, the Load from workspace dialog box was opened shown in Figure 3.6.
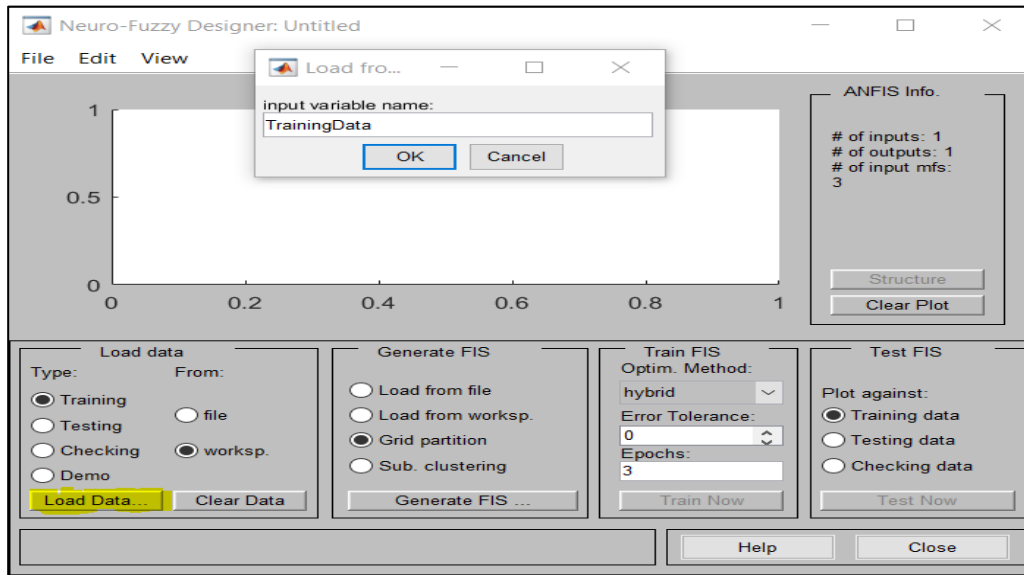
Figure 3.6: Dialog box of load data.

3.8.1.4 Training Data

'TrainingData' is typed in the dialog box and then OK was clicked. The training data set is ready to train a fuzzy system by adjusting the membership function parameters and appears in the plot in the centre of the app as a set of circles shown in Figure 3.7. In Figure 3.7, the horizontal axis was marked as data set index indicating the row from which input data value was obtained (whether or not the input is a vector or a scalar) as well as vertical axis marked as output.
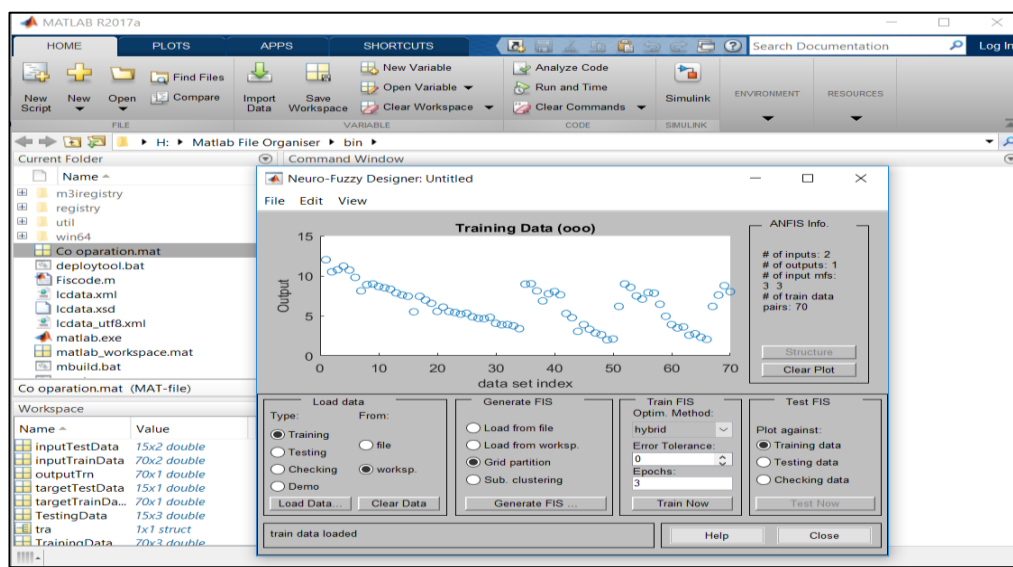

Figure 3.7: Plot of training data.

3.8.1.5 Generation of FIS

For initializing and generating FIS, two partition methods Grid Partitioning of ANFIS (ANFIS-GP) and Subtractive Clustering Method of ANFIS (ANFIS-SCM) can be used. By choosing Grid partition and clicking Generate FIS a menu is displayed from which the number of membership functions (MFs), and the type of input and output membership functions can be chosen. There are several choices for input MF like trainmf, trapmf, gbellmf, gaussmf, gauss2mf, pimf, dsigmf and psigmf as well as only two choices for the output MF: constant and linear shown in Figure 3.8.
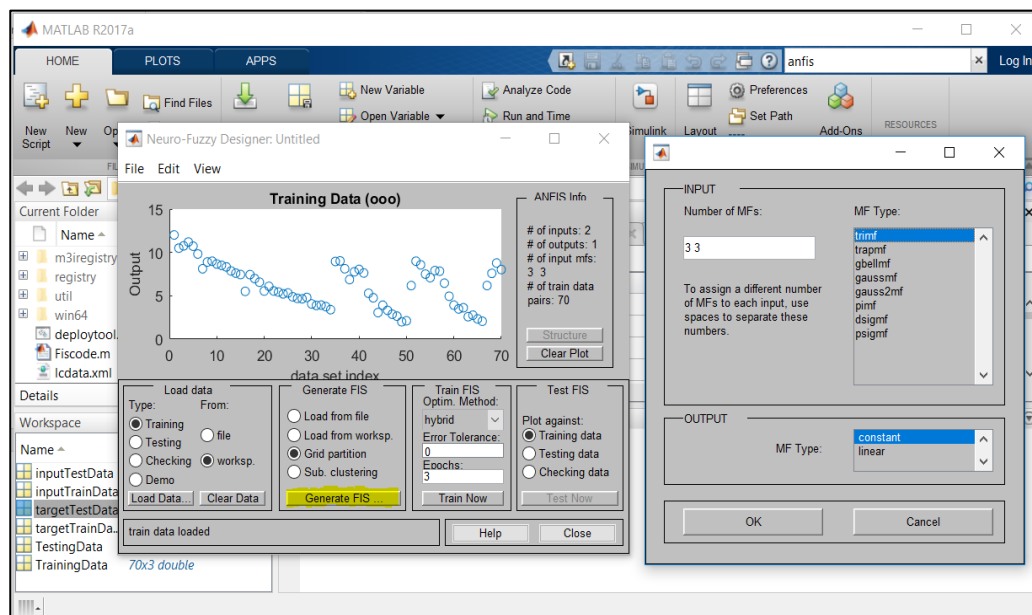

Figure 3.8: FIS generate with ANFIS-GP.

In this study, all the input and output membership functions and the optimization methods were verified to create about twenty ANFIS-GP model and choose a best one in which input membership function was gaussmf, output membership function was linear and optimization method was hybrid. After verifying the ANFIS-GP, the ANFIS-SCM (in which including input MF gaussmf, output MF linear) was used. By selecting Sub. Clustering and clicking Generate FIS a new window was opened shown in Figure 3.9.
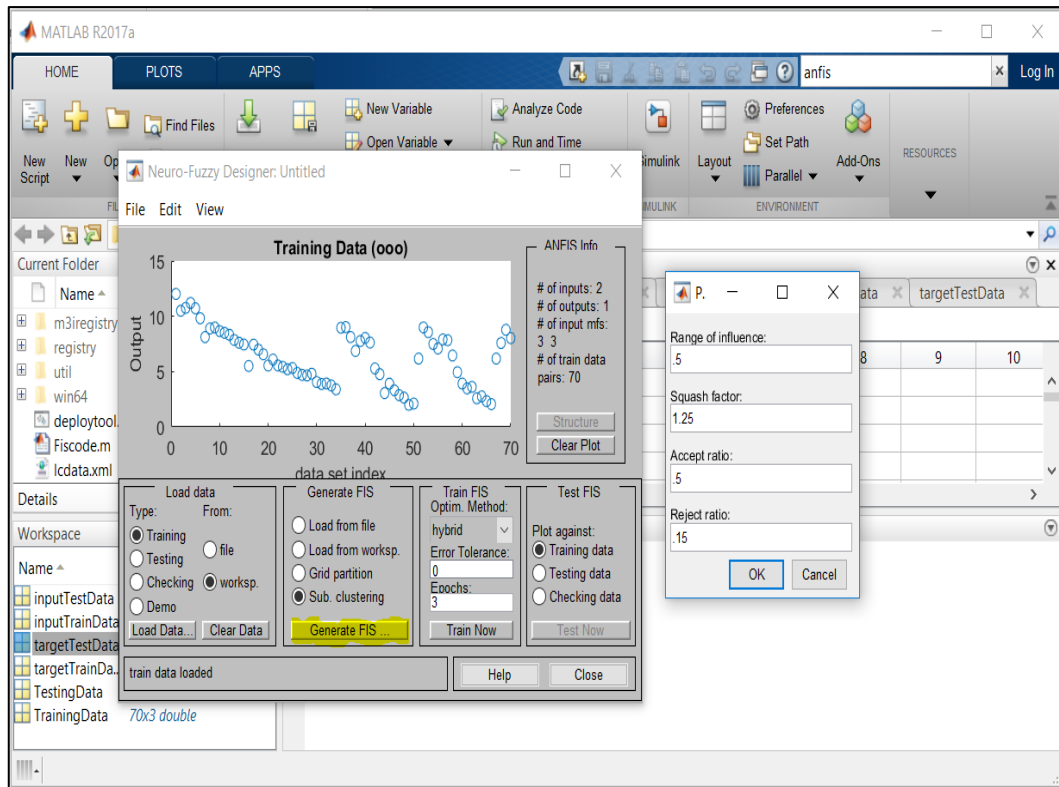
Figure 3.9: FIS generate with ANFIS-SCP.

The range of Influence, squash factor, accept ratio and reject ratio were assumed to fix for the model used in this study. By clicking Ok, the FIS is created.

3.8.1.6 Train of Model

For training FIS, the **optimization method** hybrid was chosen for the proposed model and the number of training **Epochs** was entered to 100 to set the stopping criteria for training. Then the **Train Now** option was clicked to train the FIS. This action adjusts the membership function parameters and displays the error plots, which shows the training error (RMSE value) in the bottom portion of the window shown in Figure 3.10.
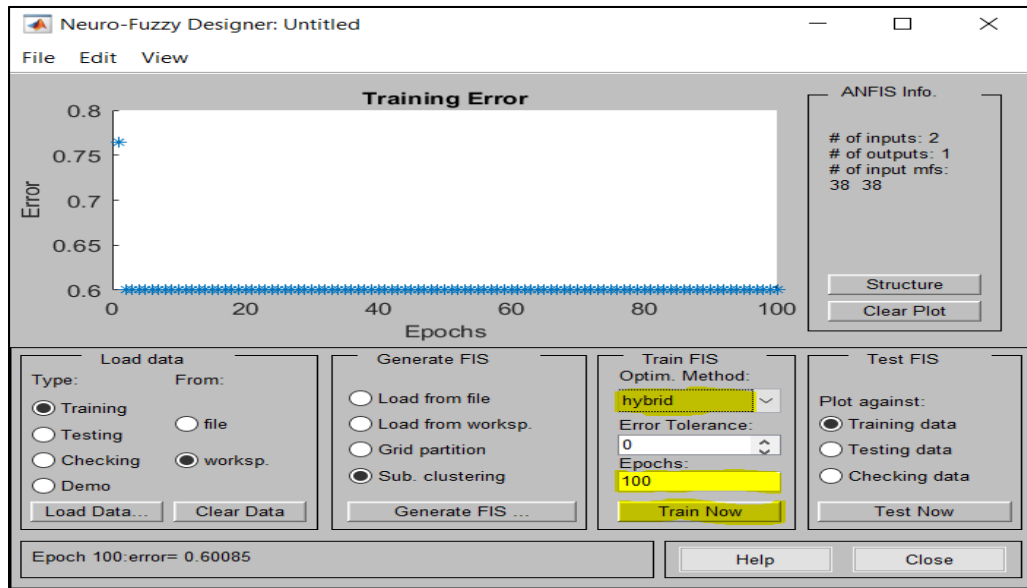
Figure 3.10: Error of ANFIS.

3.8.1.7 FIS Output for Training

A FIS output for training data was displayed by clicking **Checking** and **Test Now** option marked as yellow shown in Figure 3.11. In the centre of the app the plot was represented the output against training data (Figure 3.11).
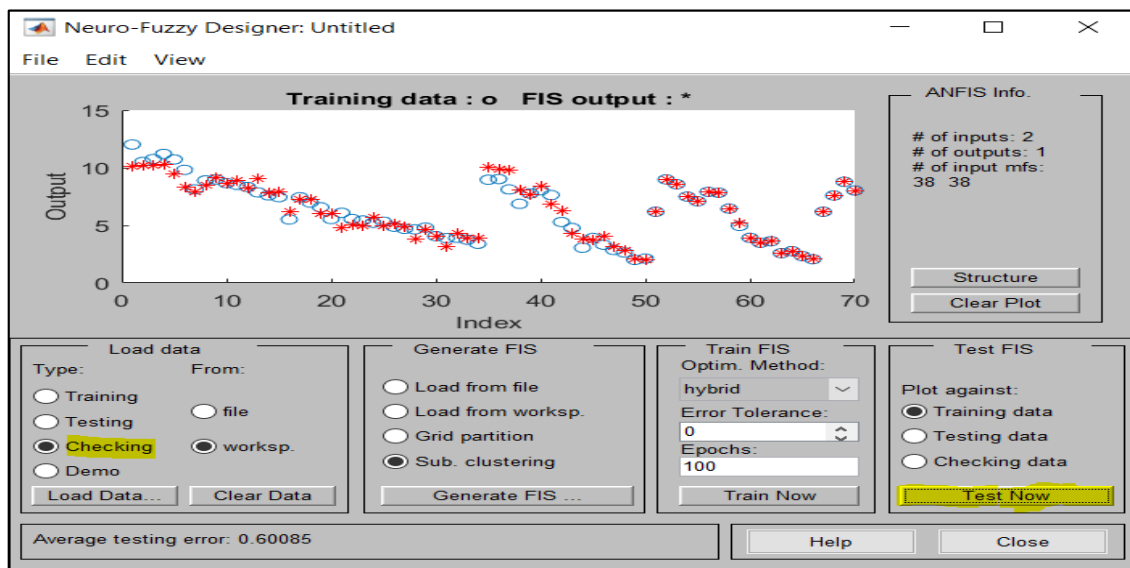

Figure 3.11: FIS output for training data.

3.8.1.8 Load of Testing Data

For validating the trained FIS, the testing data was loaded by typing the TestingData in the dialog box and then **OK** is clicked. TestingData was displayed in Figure 3.12 that was different from the TrainingData.
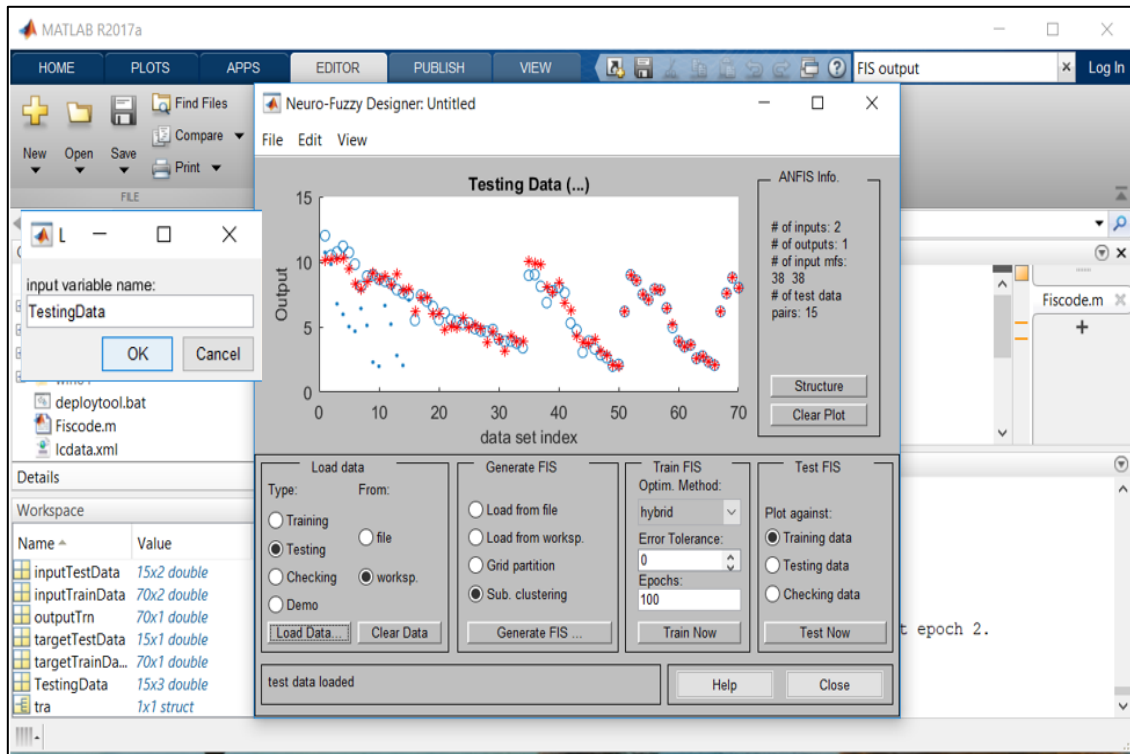

Figure 3.12: Plot of testing data.

3.8.1.9 FIS Output for Testing

After loading testing data a FIS output was displayed by clicking Checking option in the Load data portion as well as Testing data and Test Now option in the Test FIS portion of the Neuro-Fuzzy Designer, which was marked by the yellow colour in Figure 3.13. The following plot represents the predicted heavy metal concentration for 15 unknown points against the TestingData. The red star (*) mark represents the testing measured data whereas blue point (•) mark represents ANFIS output with average testing error (RMSE value) which was displayed in the bottom of window of Neuro-Fuzzy Designer App.
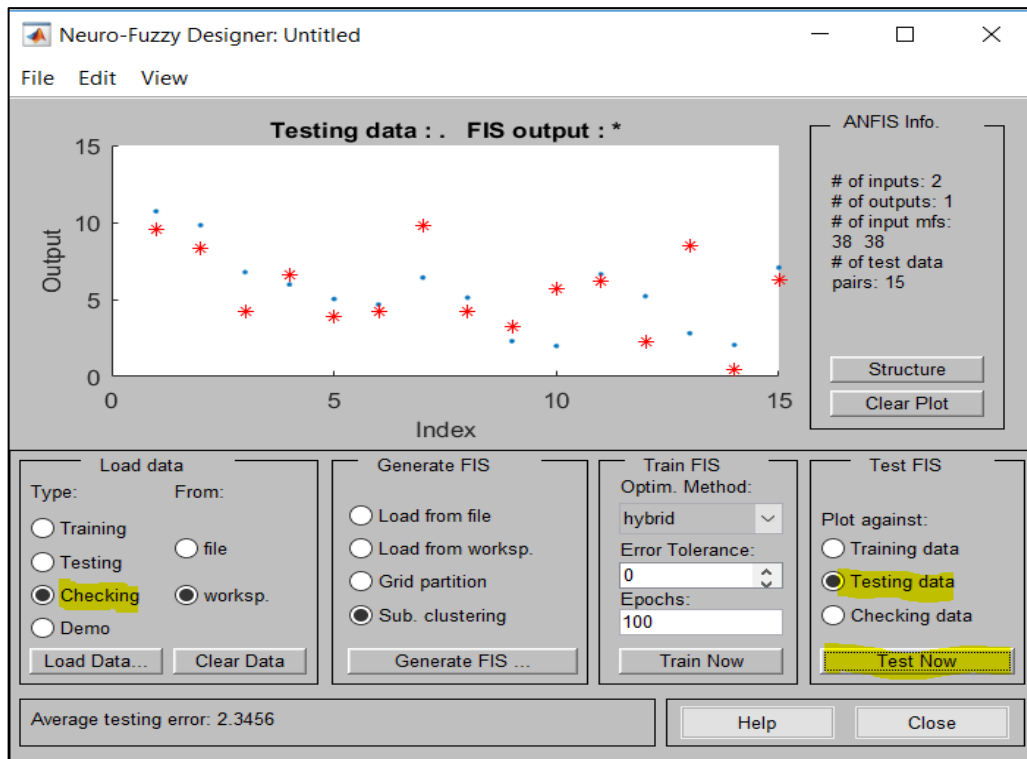
Figure 3.13: FIS output for testing data.

3.8.1.10 ANFIS Structure

The option Structure was clicked for presenting the ANFIS Structure shown in Figure 3.14. These structures were mapped with input characteristics to input membership functions, input membership functions to rules, rules to output membership functions, and the output membership functions to a single-valued output or a decision associated with the output. Such a system was used for the fixed membership functions that were chosen arbitrarily. The shape of the membership functions (must be equal to the no. of rules) depends on the type and intensity of the heavy metal concentrations in soils.
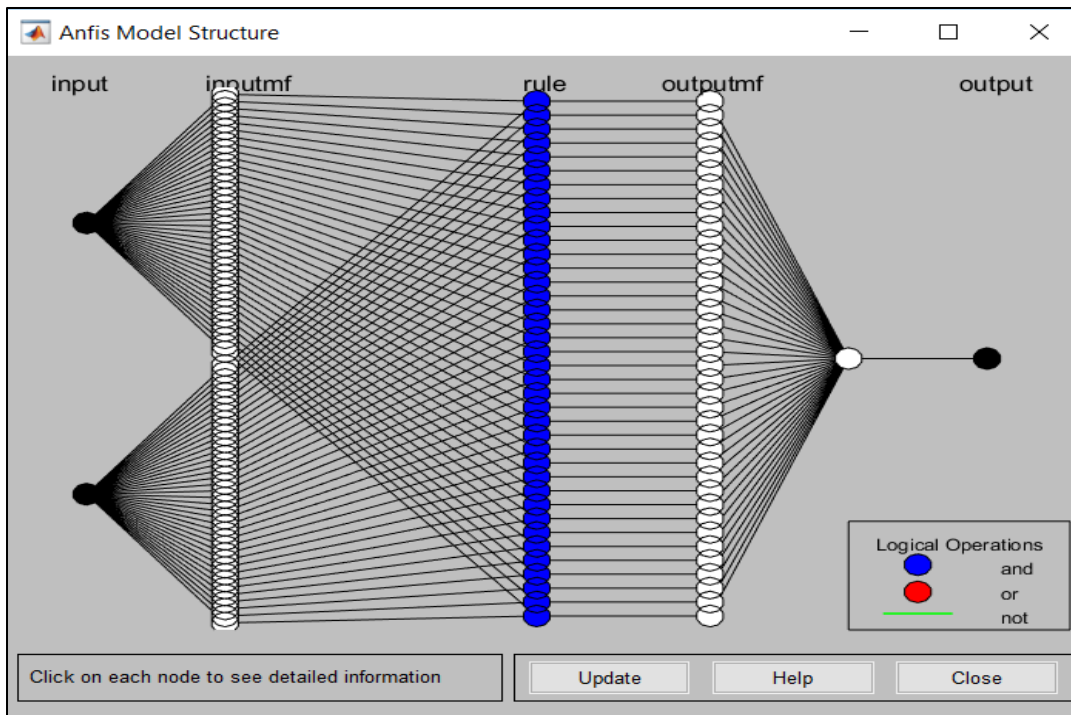
Figure 3.14: ANFIS structure.

3.8.1.11 ANFIS Rule Viewer

The Rule Viewer is a sort of micro view of the fuzzy inference system, which shows one calculation at a time and in details. From the view menu, selecting Rules option rule viewer is opened shown in Figure 3.15. In this rule viewer, two inputs and one output were represented at the top of each column for a particular input value. Each row was indicated the rules for the certain membership functions influencing overall results. Two tasks can be performed in this section. For two inputs, by inserting input vector such as [22.8;89.5] was calculated the output which showing at the top of third column as well as several number of rules was created row wise at randomly. Besides, the red index line was moved horizontally for showing the changes of input values. By clicking the options left, right, down and up, all the rules of ANFIS can be visible.
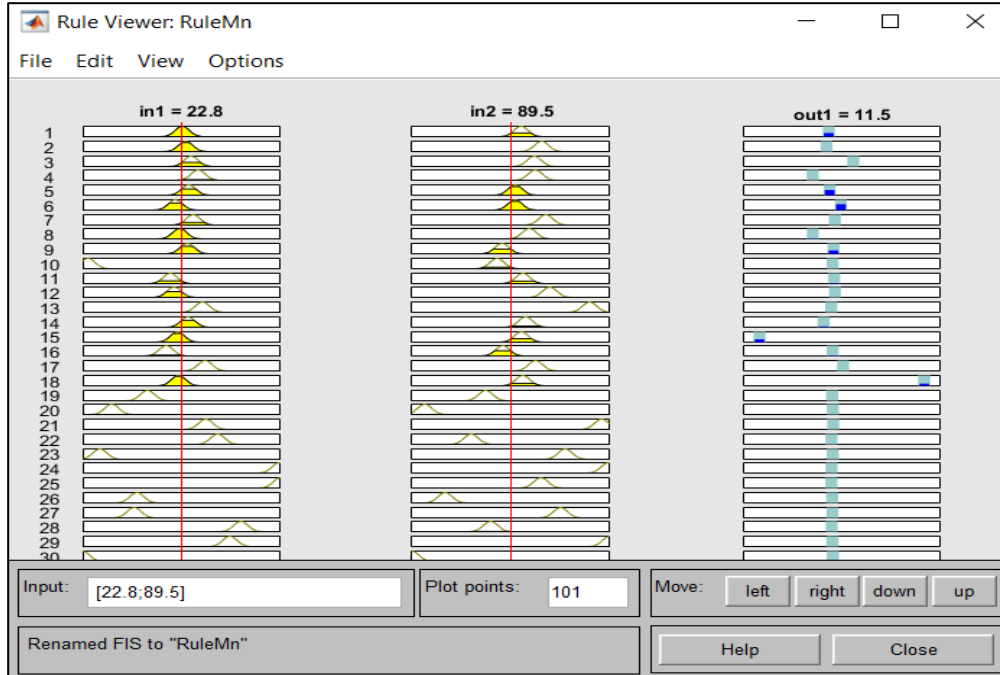
Figure 3.15: ANFIS rule viewer.

3.8.1.12 ANFIS Surface Viewer

The following steps were performed in the surface viewer section.

1. To open the Surface Viewer, **Surface** option is selected from the **View** menu.
2. Opening the Surface Viewer, a three-dimensional curve is shown that represents the mapping from latitude and longitude quality to output heavy metal concentrations shown in Figure 3.16.
3. The grid lines of x-axis and y-axis can be changed with drop-down menus X grids and Y grids.
4. By writing a code "colorbar" at the MATLAB® prompt, which described the intensity of predicted heavy metal concentrations against the two inputs by the variation of color from blue to yellow (Figure 3.16).
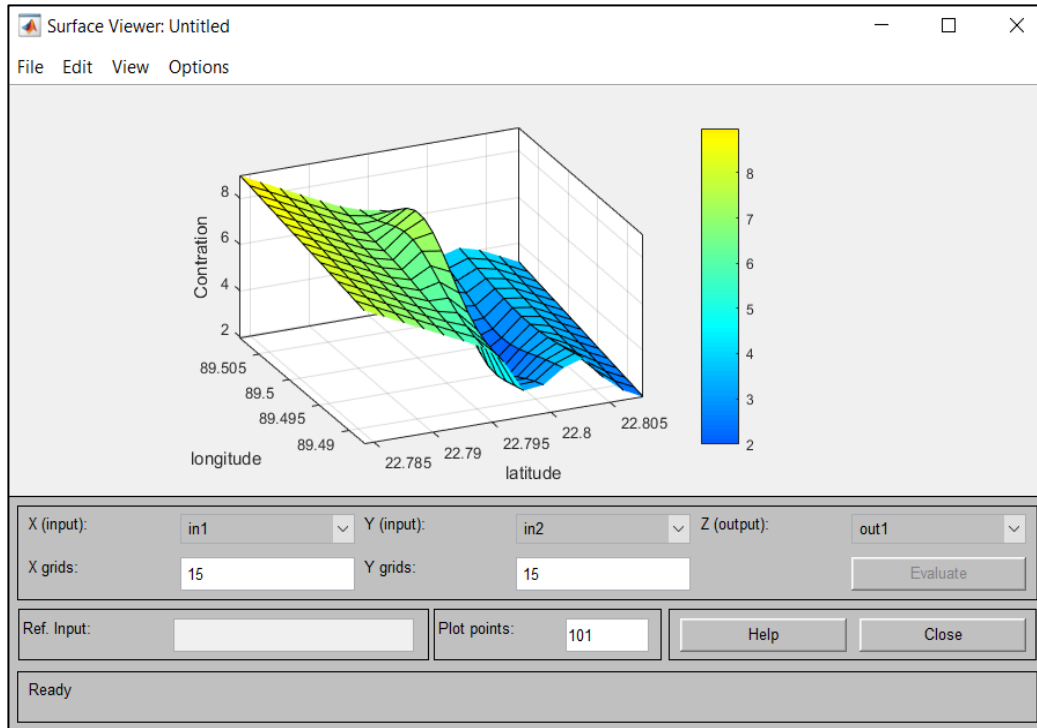
Figure 3.16: ANFIS surface viewer.

3.8.1.13 Output from Training and Testing in ANFIS

There were several tasks for determining the output or predicted concentration of training data set. The successive steps were mentioned in the following.

1. By clicking 'File' option of neuro fuzzy designer app menu, it was exported a dialog box.
2. Then 'tra' was typed to create a train FIS model as well as 'tes' was typed to create a test FIS model and clicked **OK** to save the models to the workspace (Figure 3.17).
3. Finally, the following code was typed to get the predicted concentration for training and testing, respectively.

```
Prediction1= evalfis(inputTrainData,tra)

Prediction2= evalfis(inputTestData,tes)
```
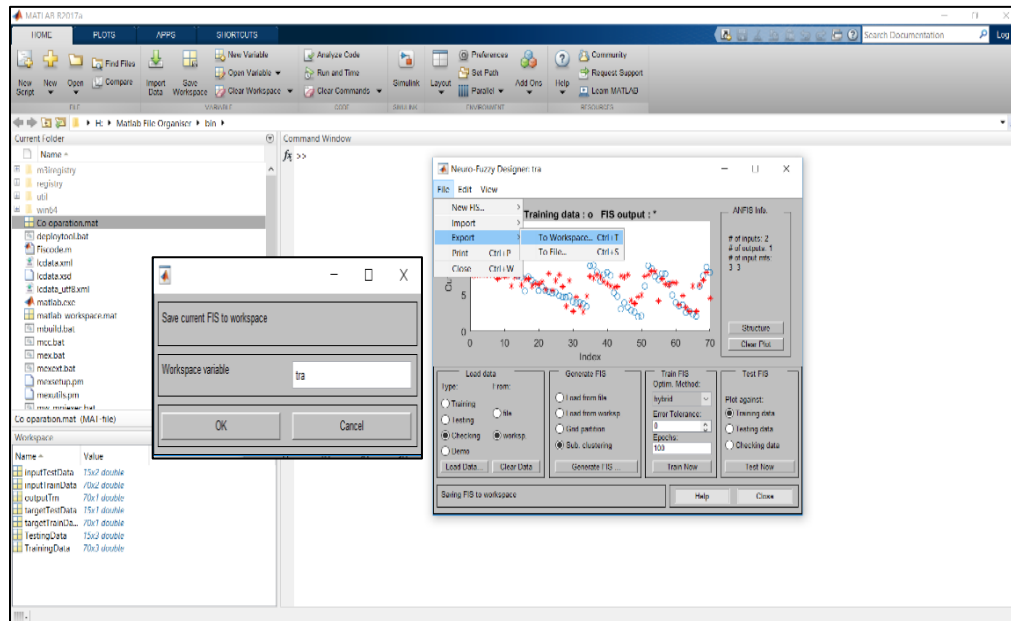
59

Figure 3.17: Saving of train and test FIS model.

3.8.2 Support Vector Machine

The models were formed with different fold numbers and various kernel functions of SVM and using "Regression Learner App". The selected model also performed for the prediction of heavy metal concentrations for unknown soils sampling points. The operation of SVM in details through MATLAB are hence discussed in the following articles.

3.8.2.1 Data Processing for SVM Analysis

To train SVM, a Training and Testing data set have been loaded that contained desired input/target data of the system to be modelled. In this study, latitude and longitude were considered as input or predictor, which placed in first two columns as well as heavy metal concentrations as target or response that placed at last column for modelling a Training and Testing data set. Among the 85 sampling points of soils from the selected waste disposal site, 83% (70) for training and 17% (15) for testing was considered for the prediction analysis. These data sets are saved in the MATLAB workspace that are ready for the SVM operation. The rearranged data sets in MATLAB workspace.

60

3.8.2.2 Regression Learner App Opening

The following command was used to open Regression Learner App shown in Figure 3.18 at MATLAB® prompt:
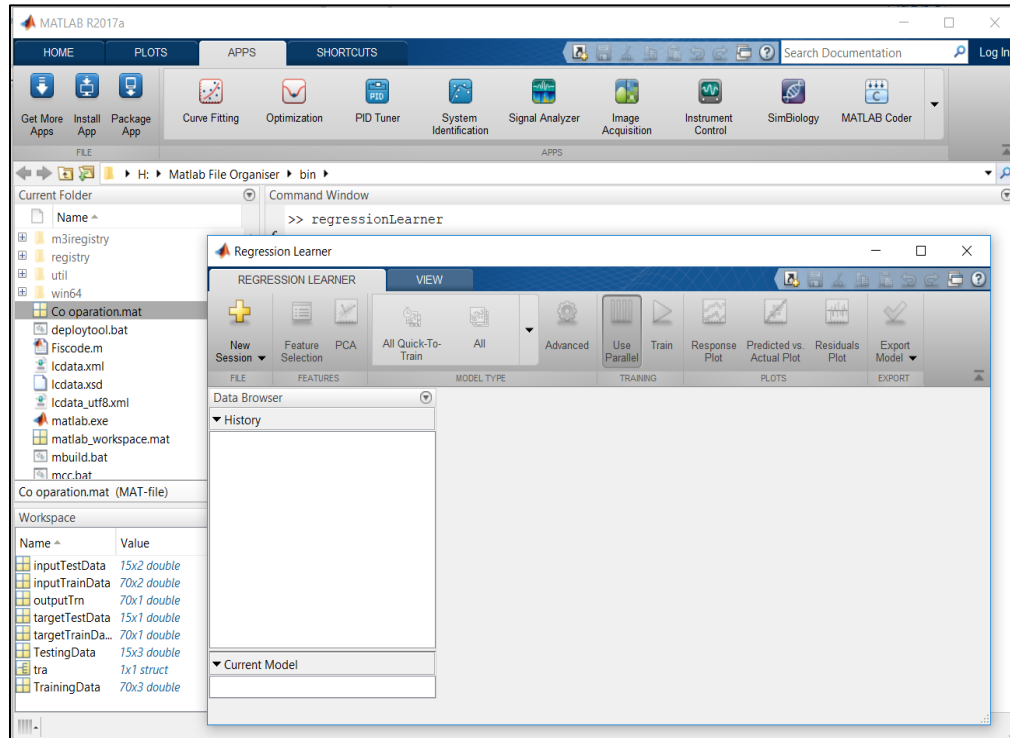
regressionLearner

Figure 3.18: Commands used to open regression learner app.

3.8.2.3 Opening New Session in MATLAB

By clicking "New Session" from the regression learner, a new window has been opened for selecting data sets from workspace shown in Figure 3.19.
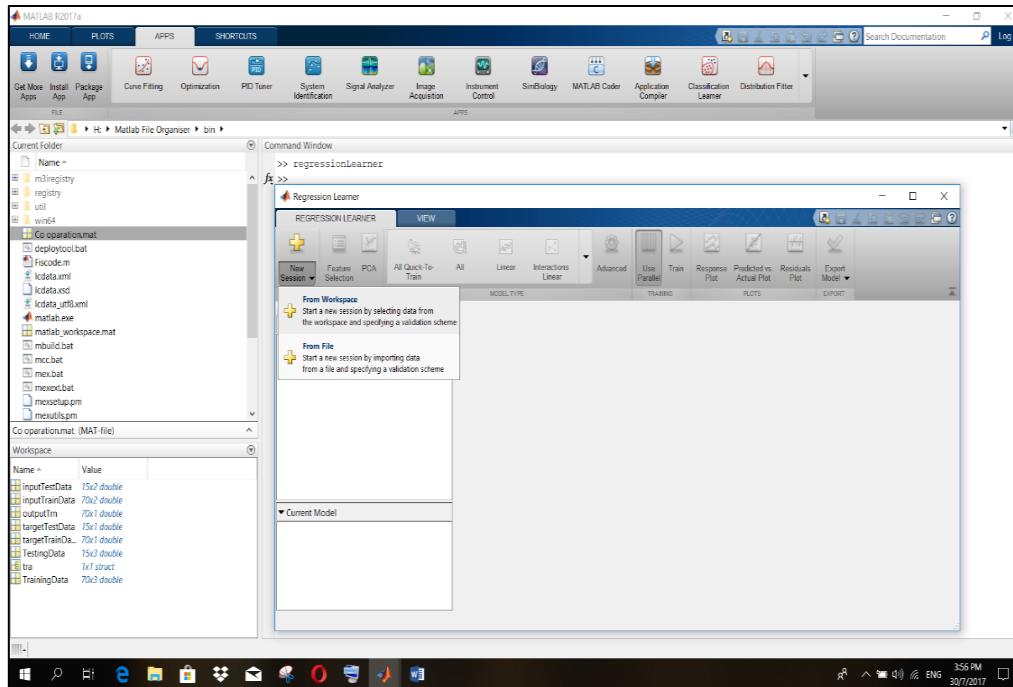
Figure 3.19: New session window open from regression learner app.

3.8.2.4 Importing Data and Validation Method Selection

In New Session window includes three distinct areas to support a typical workflow. The window performs the following tasks:

1. A table of training data sets named "TrainingData" was selected for the next operation shown in Figure 3.20.
2. Column 1 and column 2 which containing latitude and longitude, respectively, were selected as predictor as well as column 3 containing the measured heavy metal concentrations was selected as response.
3. In this study, Cross-Validation method was selected for controlling the overfitting of data set by changing the different fold number of the cross-validation.
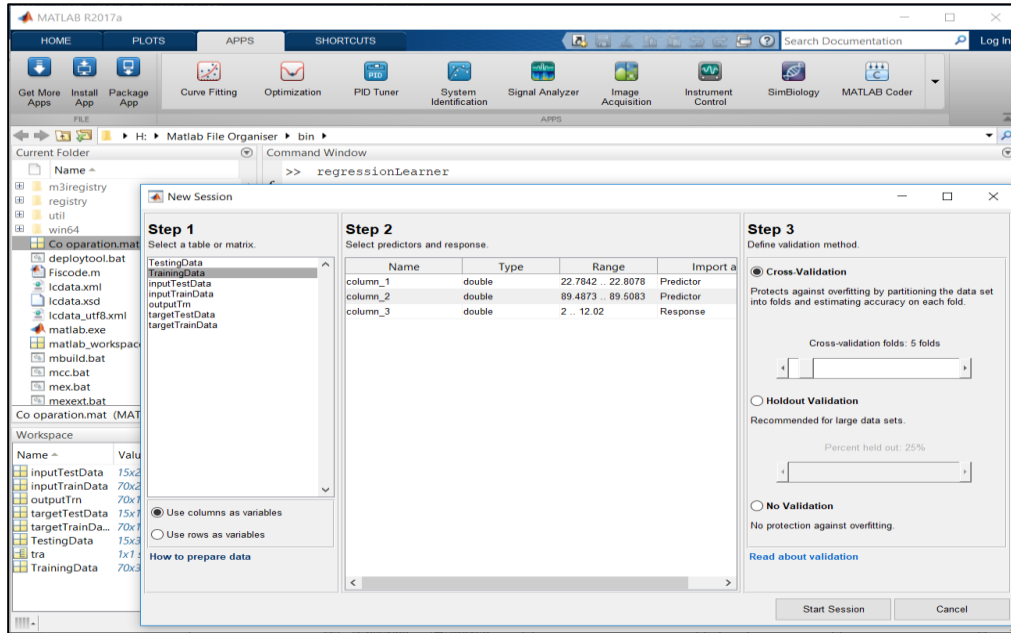
Figure 3.20: Importing data and validation method selection.

3.8.2.5 Introduction of Response Plot

The "Start Session" option was clicked to introduce the response plot which shown in Figure 3.21. This plot shows the true or measured concentration (response) of heavy metals in soils against the 83 % (70) soils sampling points for training.

3.8.2.6 Selection of Model

By clicking the "Model Type" options, various types of model can be selected which shown in Figure 3.22. In this study, "All SVMs" option was selected for training all the kernel functions and then choosing best of them.
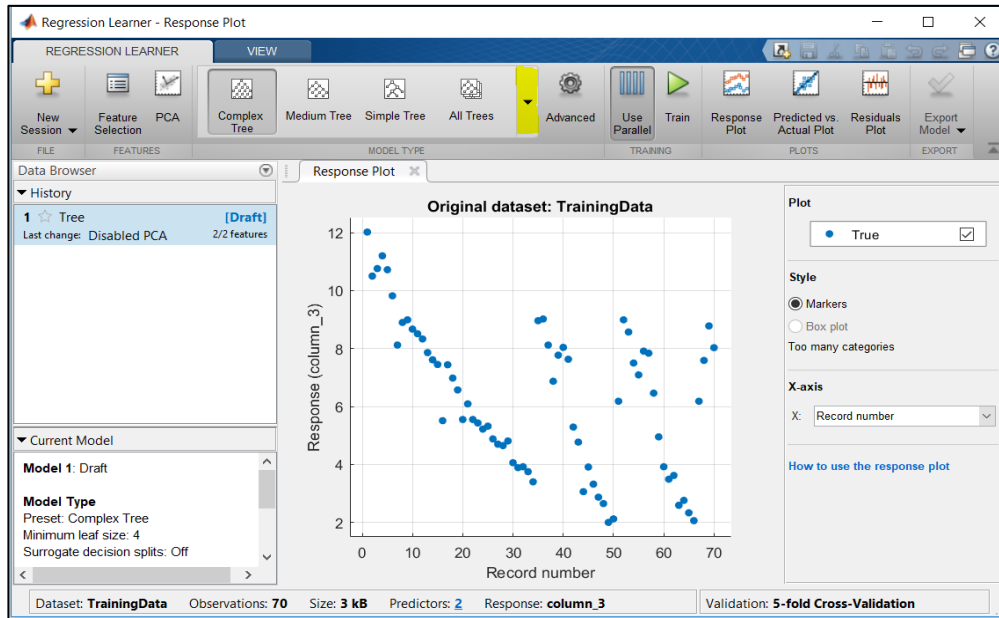
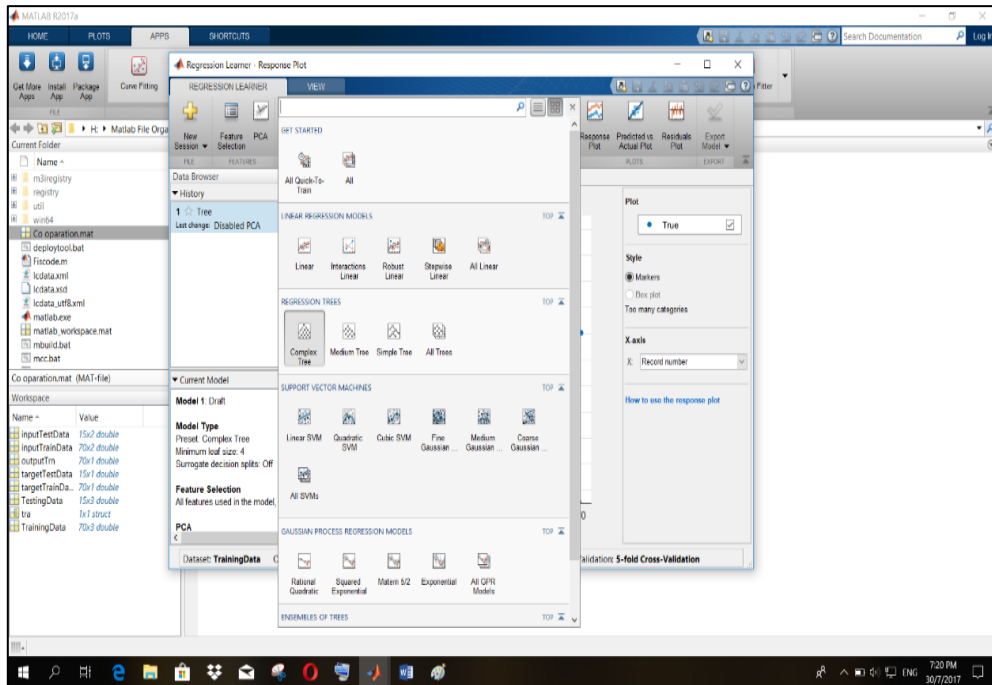Figure 3.21: Response plot of training data set.

Figure 3.22: Model selection for SVM training.

3.8.2.7 Train of Model

After selection of model type, the model was trained and then it provided RMSE (along with R^2, MSE, MAE etc.) value for all the model of various kernel functions. Figure 3.23a shows the prediction model of SVM with fine gaussian kernel. This plot of prediction model represents true (measured) and predicted value of heavy metal concentrations against the soils sampling points of waste disposal site. It also shows the variation of predicted and measured values by selecting error option (Figure 3.23b). In addition, by clicking the "Predicted vs Actual Plot" option, it was provided the plot of predicted and measured heavy metal concentrations with perfect prediction line shown in Figure 3.24. In this figure, observations (measured values) close to the perfect prediction line represents the robustness of the model.

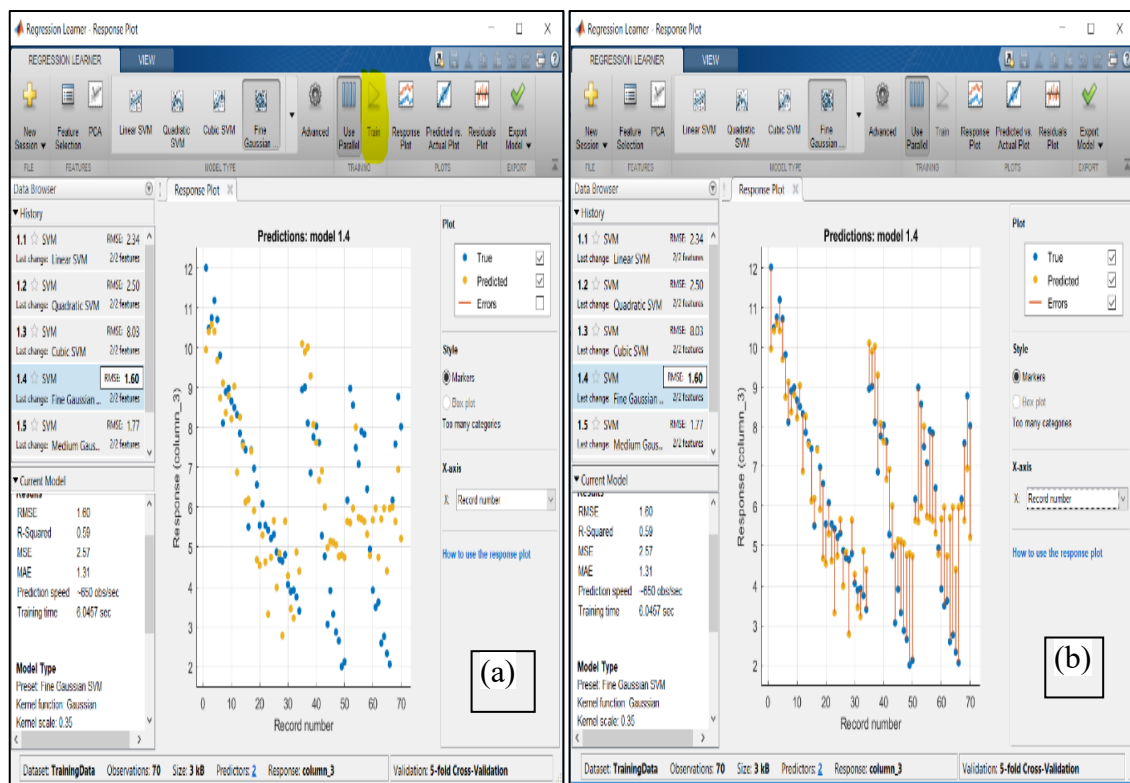Figure 3.23: (a) Prediction model of SVM and (b) Error histogram of true and predicted data.
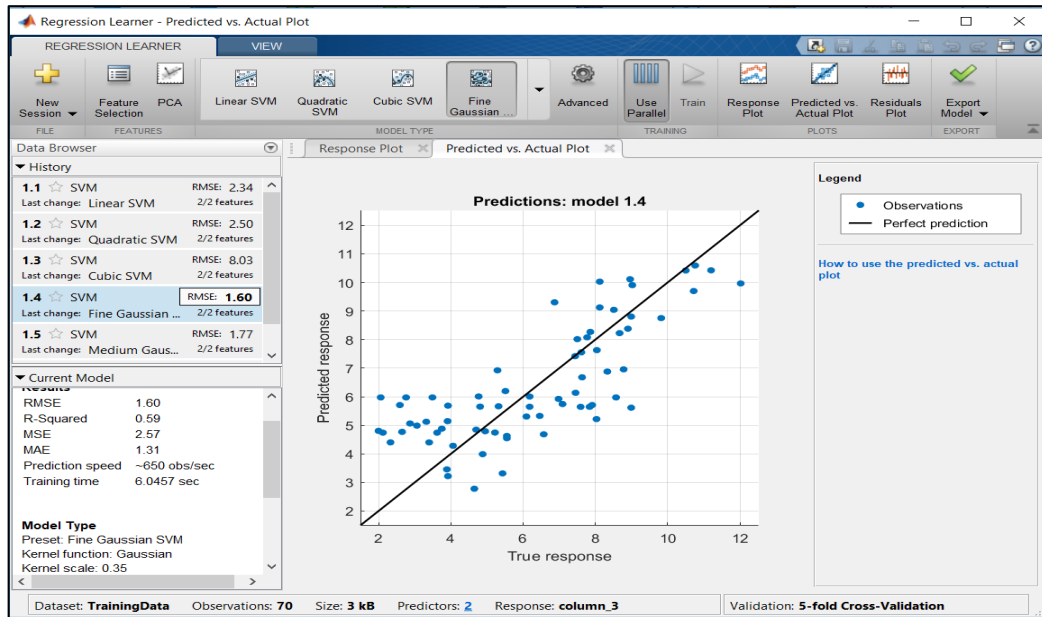
Figure 3.24: Relationship of predicted and measured values for selected kernel function of SVM.

3.8.2.8 Export of Model

By clicking option "Export Model", a window of "Export Model" was opened that was shown in Figure 3.25. Then the trainedModel was saved by clicking Ok.
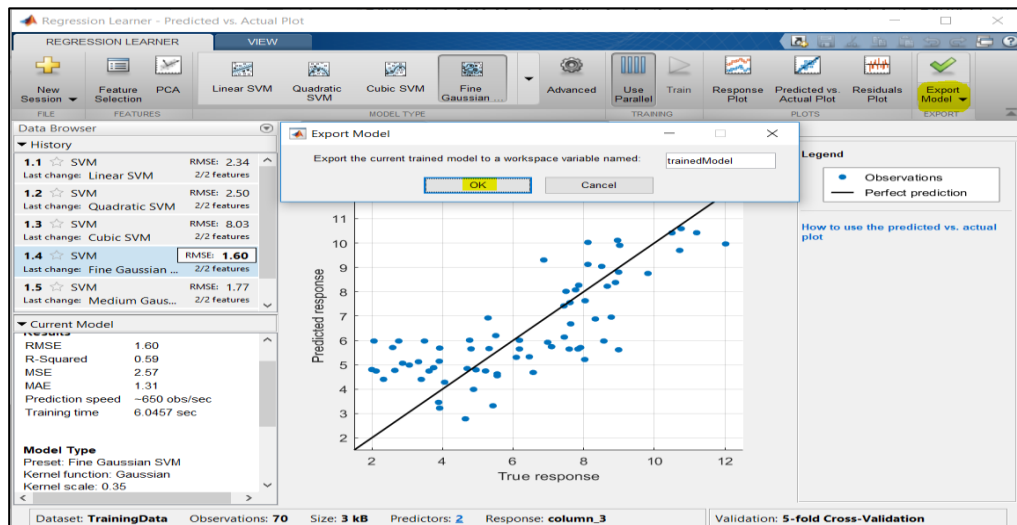

Figure 3.25: Exporting of trained model.

3.8.2.9 Coding of SVM for Training and Testing Output

Finally, the following code was typed to get the predicted concentration for training and testing, respectively.

```
Prediction1= trainedModel.predictFcn(inputTrainData)

Prediction2= trainedModel.predictFcn(inputTestData)
```

3.8.2.10 Coding for Representing the Outputs of SVM-RBF

The outputs of SVM-RBF model for all heavy metals were performed in "curve fitting tool" for the representation of the predicted concentration in terms of contour map, surface viewer and residuals. The following code was used in MATLAB for the representation of outputs of Pb (similarly others heavy metal).

```
function [fitresult, gof] = createFit(Latitude, Longitude, Pb,
Distance)

%% Fit: 'untitled fit 1'.

[xData, yData, zData, weights] = prepareSurfaceData( Latitude,
Longitude, Pb, Distance );

% Set up fittype and options.

ft = 'biharmonicinterp';

opts = fitoptions( 'Method', 'BiharmonicInterpolant' );

opts.Normalize = 'on';

opts.Weights = weights;

% Fit model to data.

[fitresult, gof] = fit( [xData, yData], zData, ft, opts );

% Create a figure for the plots.

figure( 'Name', 'untitled fit 1' );
```

```matlab
% Plot fit with data.
subplot( 2, 2, 2 );
plot( fitresult, [xData, yData], zData );
xlabel Latitude
ylabel Longitude
zlabel Pb
grid on
view( -38.5, 31.9 );
colorbar
% Plot residuals.
subplot( 2, 2, 4 );
plot( fitresult, [xData, yData], zData, 'Style', 'Residual' );
xlabel Latitude
ylabel Longitude
zlabel Pb
grid on
view( -38.5, 31.9 );
% Make contour plot.
subplot( 1, 2, 1 );
plot( fitresult, [xData, yData], zData, 'Style', 'Contour' );
xlabel Latitude
ylabel Longitude
grid on
colorbar
```

3.8.3 Artificial Neural Network

The heavy metal concentrations were analysed by the various models for different neuron numbers, training functions and transfer functions using "nntool" through ANN. Prediction of heavy metal concentrations for unknown soils sampling points were also performed by the selected model. The operation of ANN in details through MATLAB are hence discussed in the following articles.

3.8.3.1 Data Processing for ANN Analysis

To train ANN, a data set has been loaded for Training that contained desired input and target data of the system to be modelled. In this study, latitude and longitude were considered as input as well as heavy metal concentrations considered as target data. Among the 85 sampling points of soils from the selected waste disposal site, 83% (70) for training and 17% (15) for testing was considered for the prediction analysis. These data sets are saved in the MATLAB workspace that are ready for the ANN operation. The rearranged data sets in MATLAB workspace is shown in Figure 3.26.
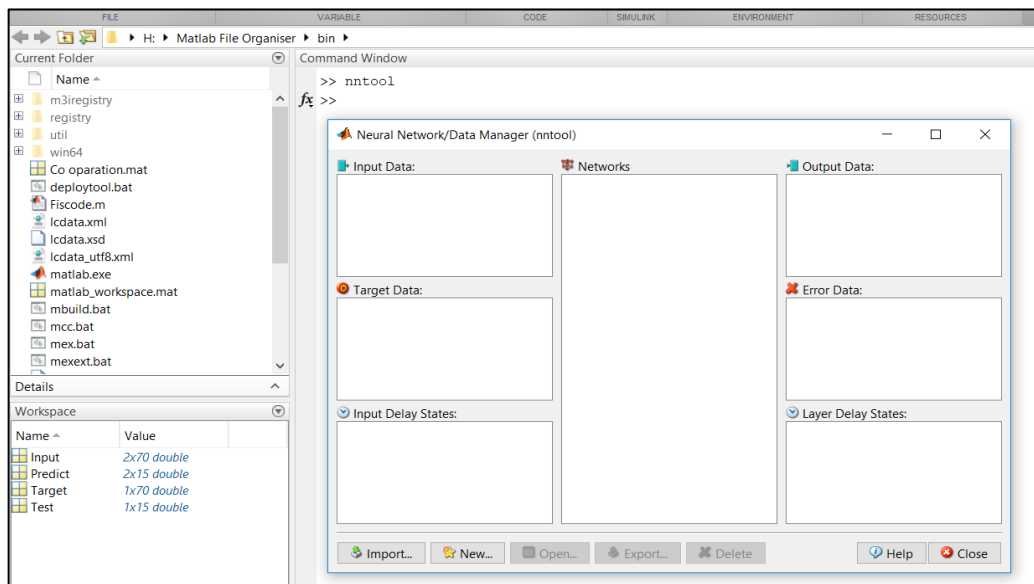

Figure 3.26: Rearranged data for ANN.

3.8.3.2 Neural Network Opening

The neural network or data manager window was opened by coding "nntool" at the MATLAB® prompt, which was shown in Figure 3.27.
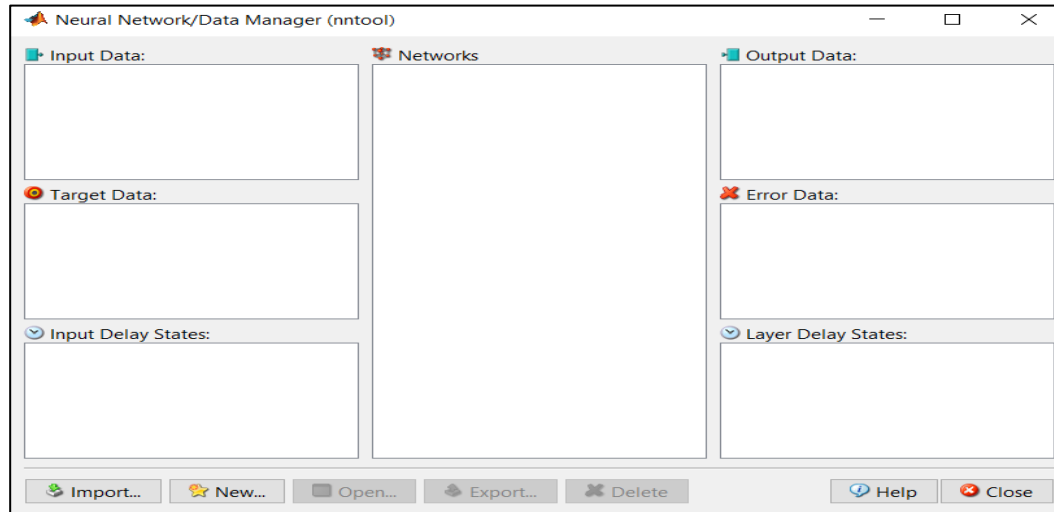

Figure 3.27: Data Manager of ANN.

3.8.3.3 Importation of Data

By clicking "import" option from the data manager, a new window has been opened for selecting data from workspace shown in Figure 3.28.
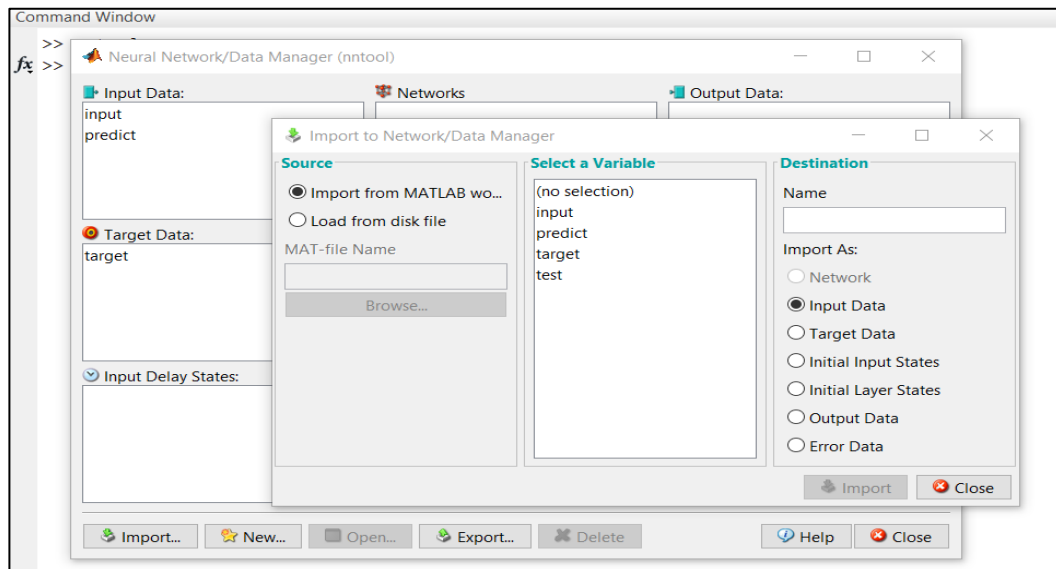

Figure 3.28: Importing of training, testing and target data.

70

3.8.3.4 Selection of Neural Network Model

The "Create Network or Data" window was opened by clicking "New" option that shown in Figure 3.29. From this window, various model of neural network was created by changing following categories.

1. Network Type
2. Training function
3. Number of neurons and
4. Transfer function

The model of neural network was assessed by the prediction parameters for selecting a best model among them. Finally, a model with network type: feed forward back propagation, training function: TRAINLM (Levenberg-Marquardt), number of neuron: 10 and transfer function: TANSIG (hyperbolic tangent sigmoid) was selected for the prediction of concentration of all studied heavy metals in soils.
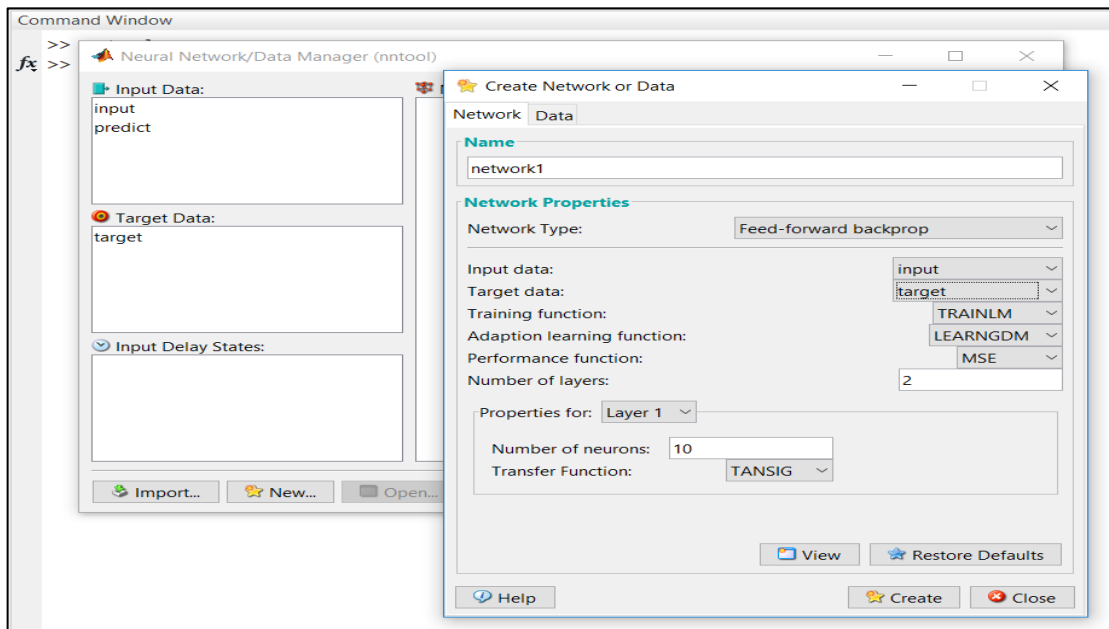
Figure 3.29: Neural network model selection.

3.8.3.5 Opening Selected Neural Network

By clicking "network1", selected neural network was opened for the training, which was shown in Figure 3.30.
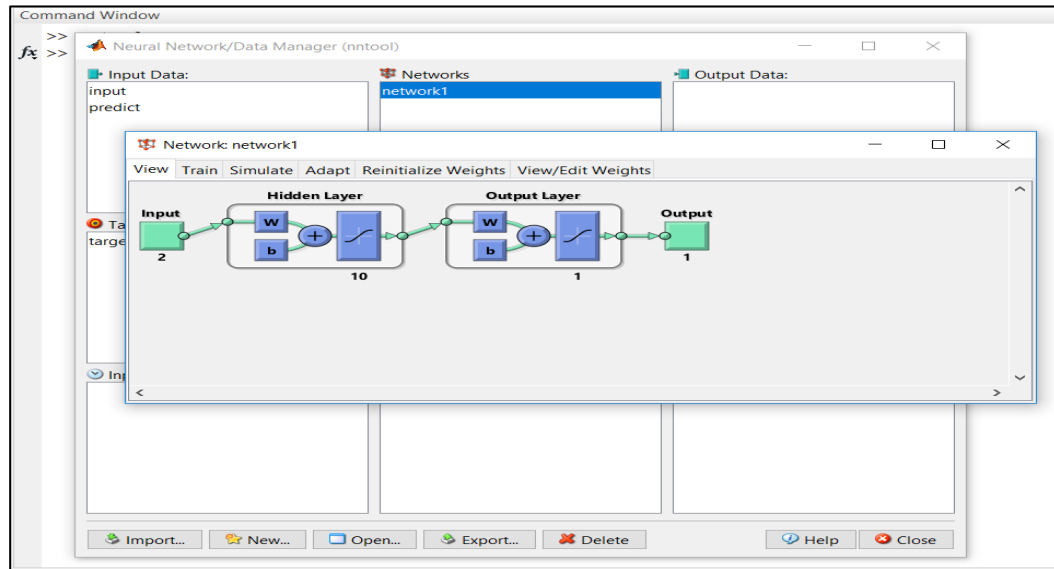

Figure 3.30: Opening of network window.

3.8.3.6 Train of ANN Model

Inserting input and target values, "Train Network" option (shown in Figure 3.31) was clicked for training the selected model.
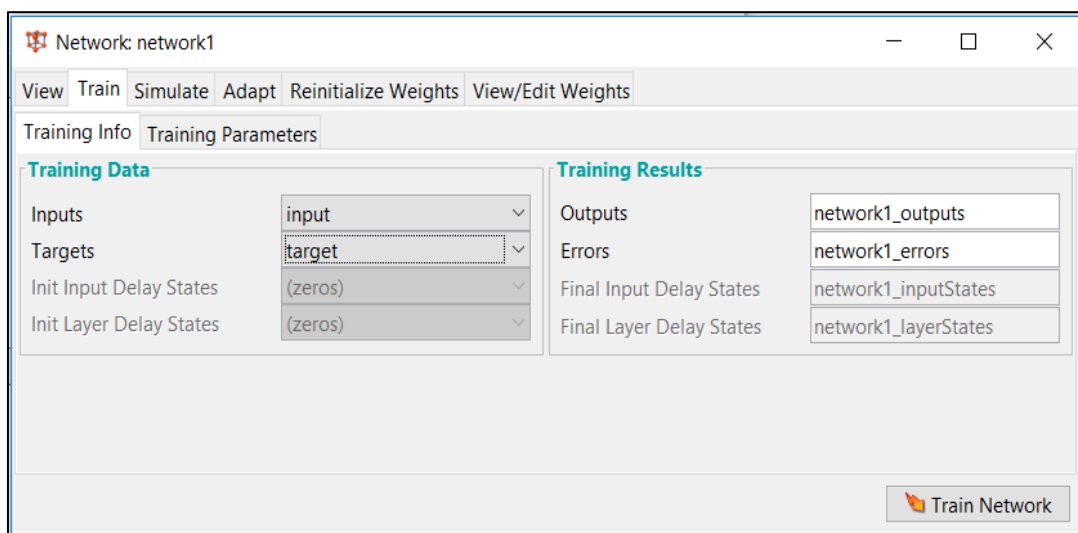
Figure 3.31: Train of neural network.

3.8.3.7 Simulation of Test Data

By inserting the inputs of 15 unknown sampling points, the network was simulated for getting the predicted concentration. Figure 3.32 shows the application of simulation.
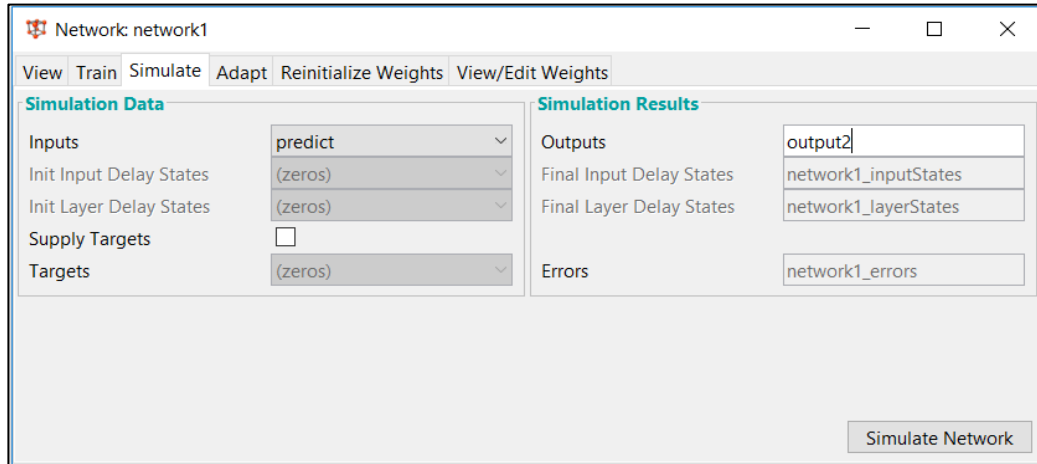
Figure 3.32: Simulation of test data.

3.8.3.8 Export of All Outputs

After training and simulation the model, outputs were exported to save in workspace, which shown in Figure 3.33.
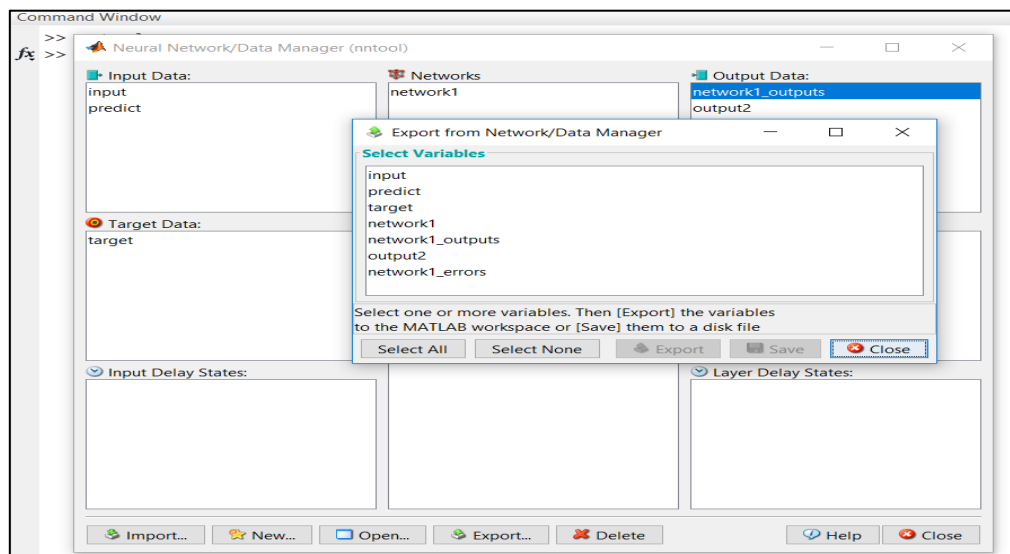
Figure 3.33: Exporting output data.

3.8.3.9 Regression and RMSE Value

The correlation coefficient, R and RMSE were determined for training and test by following coding, which shown in Figure 3.34a and 3.34b, respectively.
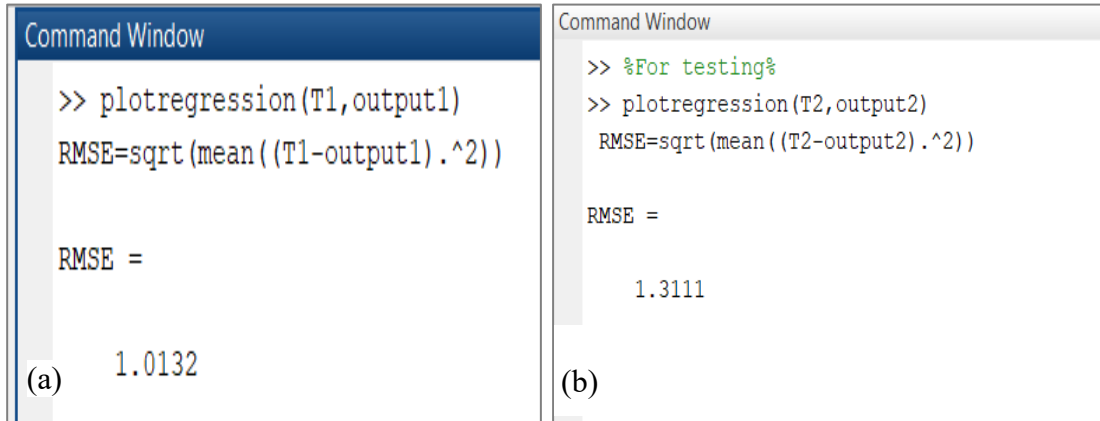
Figure 3.34: Coding of regression plot and RMSE for (a) training and (b) testing.

The correlation coefficient, R for training and testing were found from regression plot, which shown in Figure 3.35a and 3.35b, respectively.

Figure 3.35: Regression plot As for (a) training and (b) testing.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 General

This chapter deals with the selection of model from AI techniques like adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and artificial neuron network (ANN) using MATLAB. The prediction techniques of heavy metal concentrations such as Pb, Cu, Ni, Zn, 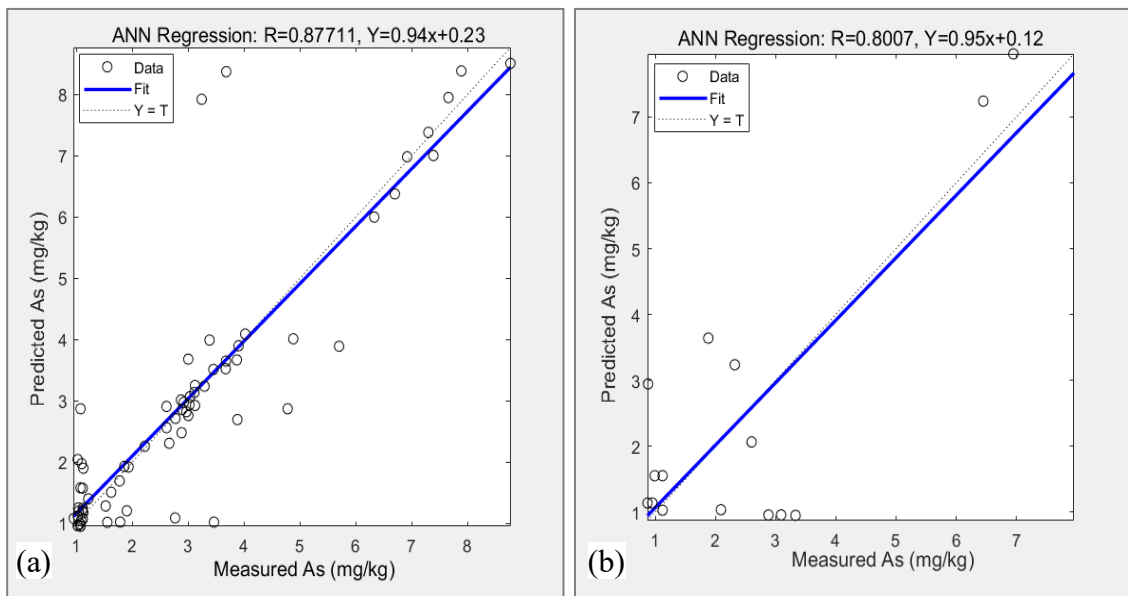Co, Cd, As, Sc, Hg, Mn, Cr, Ti, Sb, Sr, V and Ba in soils of waste disposal site for the unknown sampling points are also highlighted in this chapter. This chapter also depicts the performance analysis of AI techniques based on correlation coefficient (R) and root mean square error (RMSE). In addition, for the performance of various AI techniques were assessed by ascertaining the error in the predictions based on mean absolute percentage error (MAPE). For checking the accuracy level of prediction as well as reliability of selected models from AI techniques, geometric reliability index (GRI) and percentage recovery were also used and hence discussed in this chapter.

4.2 Basic Statistical Analysis

Statistical analysis through XLSTAT was performed to assist the interpretation of heavy metals presence in soils of waste disposal site. The descriptive statistics were represented in terms of mean, median, maximum, minimum, standard deviation (SD), skewness and kurtosis. The SD is the measure that is used to quantify the amount of variation or dispersion of a set of data or in a distribution of data. Besides, the value of skewness indicated the degree to which the analysed data are not symmetrical and kurtosis specified exactly how the peak and tails of a distribution fluctuated from the normal distribution (George and Mallery, 2010). Hence, skewness and kurtosis supports by giving primarily recognition of overall characteristics about the distribution of the heavy metal concentrations in soils. According to George and Mallery (2010), the values for skewness and kurtosis between -2 and +2 are considered acceptable in order to prove normally distributed data. The descriptive statistical data of heavy metals in soils of waste disposal site is provided in Table 4.1. Table 4.1 illustrates the preponderance of heavy

Table 4.1: Descriptive statistical data of heavy metals in soils of waste disposal site

Heavy metal elements	Mean	Median	Maximum	Minimum	SD	Skew	Kurt
Pb	27.367	23.386	90.550	10.880	13.739	1.863	5.103
Cu	4.391	3.690	16.450	0.730	3.185	2.111	4.675
Ni	3.728	3.450	8.060	1.080	1.709	0.628	-0.069
Zn	25.932	23.075	50.760	11.820	10.003	0.783	-0.162
Co	6.152	6.180	12.020	1.980	2.539	0.140	-0.796
Cd	3.309	3.108	7.030	1.200	1.529	0.496	-0.580
As	2.862	2.660	8.770	0.870	1.957	1.259	0.972
Sc	10.013	9.712	20.411	3.020	3.534	0.418	0.207
Hg	3.210	2.946	9.200	0.720	2.166	0.975	0.448
Mn	10.700	11.760	30.760	1.020	6.536	0.428	0.453
Cr	4.014	4.345	9.820	0.770	2.370	0.260	-0.834
Ti	973.73	916.17	1937.36	243.88	419.50	0.517	-0.421
Sb	4.790	4.710	12.549	0.980	2.257	0.678	0.961
Sr	22.393	21.710	54.120	8.880	8.540	0.889	1.254
V	32.563	30.855	83.351	6.920	15.803	0.785	0.724
Ba	52.890	47.220	121.902	18.200	23.175	0.879	0.637

metals in soils of disposal site very clearly. The SD for the heavy metal of Ti was found to be greater than that of other heavy metals which indicating the highest dispersion range of Ti within the soils sampling area. In addition, the values of skewness and kurtosis of the heavy metals in soils indicated that the data was distributed normally excepting for Pb and Cu.

4.3 Soil Quality Parameters and Allowable Limit

In this study, the arithmetic mean of heavy metal concentration in soils were compared with the maximum allowable limits for soils published by different countries available in literatures (Table 4.2). In this analysis, the mean was considered to avoid the uncertainty associated with estimating the true average concentration at a site stated in USEPA (1989). Different organizations and countries provide the maximum allowable limits of heavy metal concentrations in soil. Because, quality guidelines depend on this allowable limits for the protection of environment and human health from contaminated site. In the present study, the soil samples showed the comparatively lower concentration of Cu, Ni, Zn, Co, As, Hg, Mn, Cr,

Sb, V and Ba than that of maximum allowable limits proposed by different countries (Table 4.2). The average concentration of Pb exceed the WHO standard maximum allowable limits. Besides, the concentration of Cd exceeds the allowable limits stated by CCME, Poland, U.K. and U.S.A. except WHO standard, Austria and Canada (Table 4.2).

Table 4. 2: Comparison of mean concentration of heavy metals (mg/kg) in soils with maximum allowable limits in present study

Heavy Metal	WHO standard	CCME	Austria	Canada	Poland	Japan	UK	Germany	U.S.A	Present study
Pb	15-20	70	100	200	100	400	100	500	200	27.37
Cu	-	63	100	100	100	125	100	50	100	4.40
Ni	0-100	45	100	100	100	100	50	100	500	3.73
Zn	20-300	200	300	400	300	250	300	300	300	25.93
Co	-	40	50	25	50	50	-	-	40	6.15
Cd	0-30	1.4	5	8	3	-	3	-	0.7	3.31
As	-	12								2.86
Hg	-	6.6								3.21
Mn	200-9000	-								10.70
Cr	0-85	64	100	75	100		50	200	1000	4.00
Sb	-	20								4.80
V	-	130								32.56
Ba	-	750								52.89

CCME=Canadian Council of Ministers of the Environment (Source: Fahmida and Rafizul, 2017).

The concentration of heavy metals can easily accumulate in the bodies of soil organisms. In fact, they are particularly very dangerous chemical, as they can accumulate in individual organisms but also in entire food chain (Shingh et al., 2010). Moreover, soil samples are disturbed by Pb and Cd intervention, where extreme concentrations are present. The main sources of Pb and Cd in municipal solid waste are lead-acid batteries, household batteries, consumer electronics, glass and ceramics, plastics, soldered cans, pigments etc. (Korzun and Heck, 1990). The concentration

of heavy metals that have the most damaging effects on human health like anaemia, high blood pressure, kidney damage, brain damage, etc.

4.4 Artificial Intelligence Techniques

In this study, Artificial Intelligence (AI) techniques such as adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and artificial neural network (ANN) were performed through MATLAB to analysis heavy metal concentrations in soils of waste disposal site. The latitude and longitude of soil sampling points of the selected waste disposal site were used as inputs, while, heavy metal concentrations were considered as outputs in AI techniques. In this study, total 85 sampling point's data were considered among which training data 83% (70) and testing data 17% (15) were assigned in AI techniques. The adopted AI techniques are hence discussed in the following articles.

4.4.1 Adaptive Neuro-Fuzzy Inference System

In this study, for the validation of models from various functions of ANFIS, the performance of prediction model with acceptable limits of prediction parameters and FIS outputs were deliberated successively. The results for the heavy metals such as Pb, Hg, Sb and Sc in soils for ANFIS were discussed in details here. However, the results for other heavy metals such as Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Sb, Sr, V and Ba in soils from ANFIS were reported in Appendix-A (Figures) and Appendix-B (Tables).

4.4.1.1 Validation of Models in ANFIS

In this study, to validate the models of ANFIS, twenty models symbolized A to T were formed considering sub-clustering partitioning (SCP); different input membership function (MF) like gaussmf, trimf, trapmf, psigmf, gbellmf; output MF like linear and constant; optimization method such as hybrid or back-propagation (BP) as well as number of epochs. The performance of different models (A to T) of ANFIS were examined based on the satisfactory limits of the prediction parameters such as R and RMSE. The results of twenty models (A to T) in ANFIS at different functions for Co with the values of R and RMSE provided in Table 4.3.

Table 4.3: Validation of different models in ANFIS for Co

Model name	GENFIS	Input membership function	Output membership function	Optimization method	Epochs	Co	
						R	RMSE
A	Sub Clustering	gaussmf	linear	Hybrid	100	**0.80**	**1.52**
B	Sub Clustering	gaussmf	Constant	Hybrid	100	0.64	1.92
C	Sub Clustering	gaussmf	linear	Back Propagation	100	-0.28	2.53
D	Sub Clustering	gaussmf	Constant	Back Propagation	100	-0.13	4.98
E	Sub Clustering	trimf	linear	Hybrid	100	0.79	1.53
F	Sub Clustering	trimf	Constant	Hybrid	100	0.62	1.97
G	Sub Clustering	trimf	linear	Back Propagation	100	0.39	2.66
H	Sub Clustering	trimf	Constant	Back Propagation	100	0.03	6.71
I	Sub Clustering	trapmf	linear	Hybrid	100	0.67	1.87
J	Sub Clustering	trapmf	Constant	Hybrid	100	0.50	2.16
K	Sub Clustering	trapmf	linear	Back Propagation	100	0.32	2.40
L	Sub Clustering	trapmf	Constant	Back Propagation	100	0.04	5.30
M	Sub Clustering	psigmf	linear	Hybrid	100	0.62	1.96
N	Sub Clustering	psigmf	Constant	Hybrid	100	0.61	1.98
O	Sub Clustering	psigmf	linear	Back Propagation	100	0.17	2.45
P	Sub Clustering	psigmf	Constant	Back Propagation	100	0.19	4.98
Q	Sub Clustering	gbellmf	linear	Hybrid	100	0.66	1.88
R	Sub Clustering	gbellmf	Constant	Hybrid	100	0.53	2.13
S	Sub Clustering	gbellmf	linear	Back Propagation	100	0.47	2.35
T	Sub Clustering	gbellmf	Constant	Back Propagation	100	0.13	4.98

For model A (SCP, gaussmf, linear and hybrid) the value of |R| was found to be 0.80 indicating the strong correlation between input and output variables in ANFIS analysis. In this study, the models B, E, F, G, I, J, K, M, N, Q, R, and S provided the R-value within a range of 0.2 < |R| < 0.8 indicating correlations between input and output variables. However, R-values for rest models (C, D, H, L, O, P and T) were found below 0.2 with weak correlations. Figure 4.1 shows the variation of R and RMSE in case of Co for different models performed in ANFIS analysis. The model H (SCP, trimf, constant and BP) shows comparatively the lower R-value (0.03) with higher RMSE value (6.71) than that of other models.
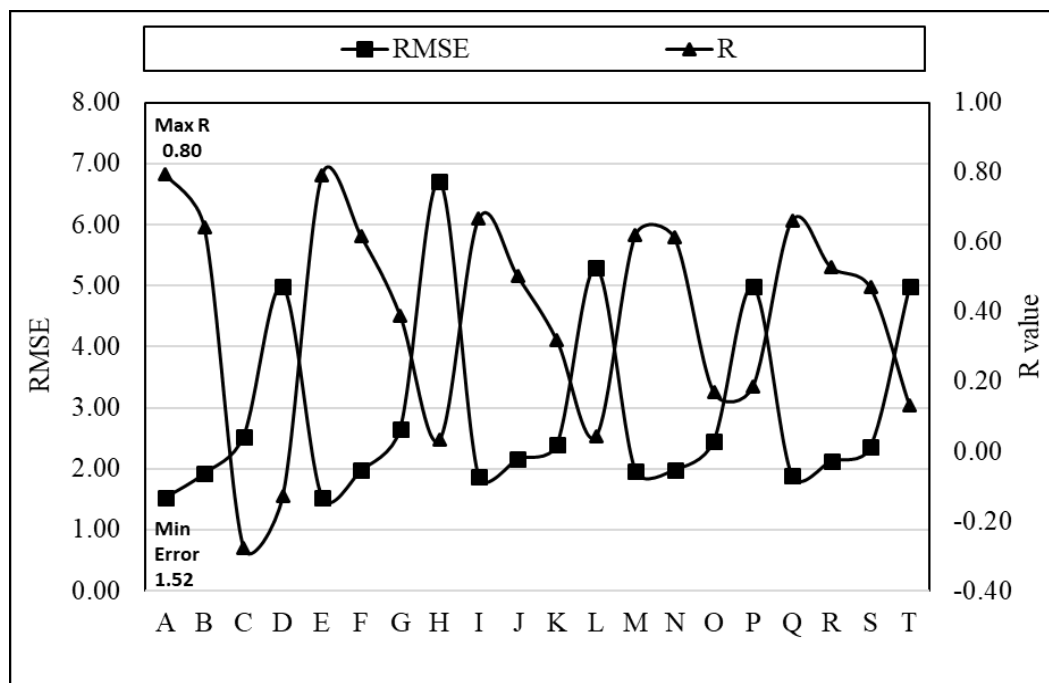
Figure 4.1: Variation of R and RMSE of Co for different ANFIS models.

In addition, two models such as C and D shows R and RMSE values of -0.28 and -0.13 as well as 2.53 and 4.98, respectively, these two models indicated the performance of weak downhill correlations between input and output variables. Besides, the model A shows the maximum value of R 0.80 and minimum value of RMSE 1.52 (Figure 4.1). Based on aforementioned results of R and RMSE, model A with SCP, gaussmf, linear and hybrid can be considered as fitted model of ANFIS for the prediction of all studied heavy metal concentrations in soils of waste disposal site.

80

4.4.1.2 Representation of ANFIS

The results of ANFIS model in terms of model structure, rules viewer, outputs for training data, outputs of testing data and surface viewer were analysed and hence discussed in the following articles.

4.4.1.2.1 Structures of ANFIS

The model A with SCP of gaussmf (input MF), linear (output MF) and hybrid algorithm (optimisation method) were selected for the prediction of heavy metal concentrations in soils. The structures of ANFIS model for Pb and Hg were shown in Figure 4.2 (a) and Figure 4.2 (b), respectively, for training and testing. In this study, ANFIS structure of Pb contained 66 rules (Figure 4.2a) and Hg contained 11 rules (Figure 4.2b) with two inputs and one single output for the heavy metal concentrations in soils. ANFIS model structure for Pb shows comparatively more rules due to the higher intensity of the concentration of Pb (Figure 4.2a) than that of Hg (Figure 4.2b) in soils. Result reveals the shape of MF changes for the different number of rules of these two heavy metals in soils.
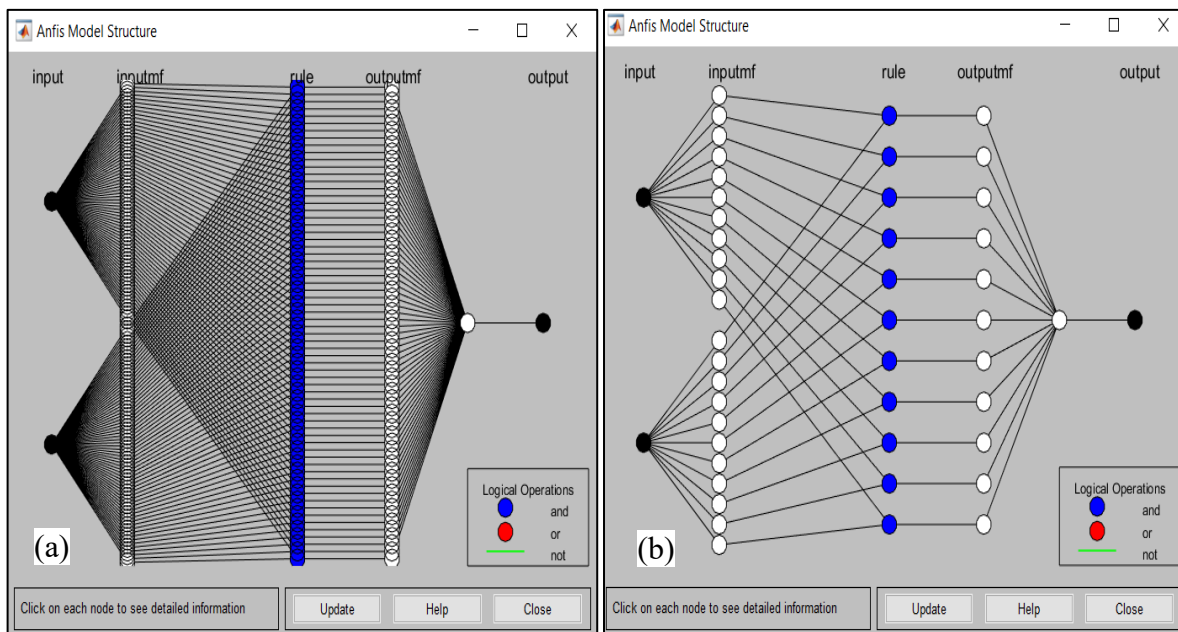
Figure 4.2: ANFIS structure for (a) Pb and (b) Hg.

4.4.1.2.2 Rule Viewer of ANFIS

In this analysis, the variables such as latitude, longitude and measured concentrations of heavy metal were considered to develop models for predicting heavy metal concentrations of unknown soil sampling points within the selected waste disposal site. At the top of each column from left to right of these rule viewer represents latitude, longitude and the predicted heavy metal concentrations for a particular input value (latitude and longitude) shown in Figure 4.3.
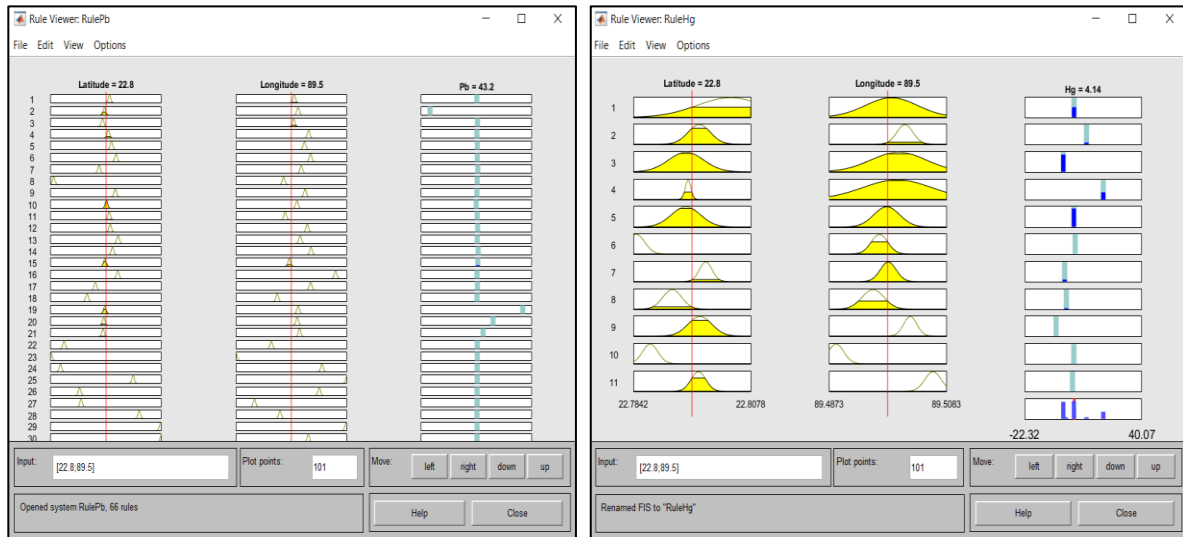
(a) 66 Rules   (b) 11 Rules
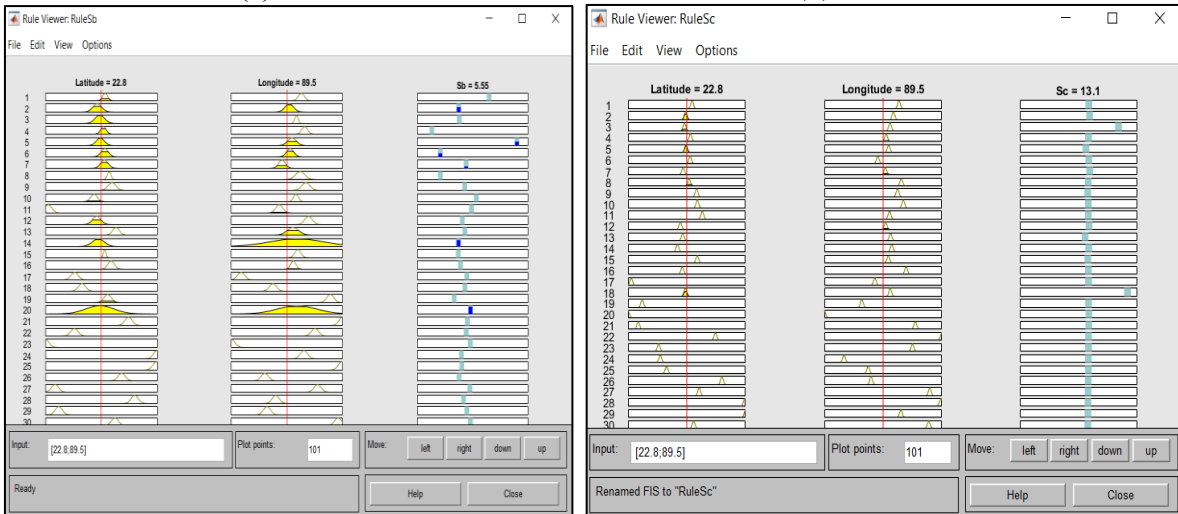
(c) 38 Rules   (d) 65 Rules

Figure 4.3: ANFIS rule viewer for (a) Pb (b) Hg (c) Sb and (d) Sc.

In this analysis, the developed each row represents the rules for a particular heavy metals in soils. The model for each heavy metal, different number of rules like (66 rules for Pb, 11 rules for Hg, 38 rules for Sb and 65 rules for Sc) were established shown in Figures 4.3 (a), 4.3 (b), 4.3 (c) and 4.3 (d), respectively. From this analysis, it can be seen that when the values of latitude (22.796991) and longitude (89.500559) were considered as input in the developed model (Figure 4.3a) for Pb, the output/predicted value of Pb was obtained as 43.2.

Similarly, Figure 4.3b, 4.3c and 4.3d were described respectively in such a manner of:

If latitude = 22.794104 and longitude = 89.499353 then the output of Hg = 4.14.

If latitude = 22.796991 and longitude = 89.500559 then the output of Sb = 5.55.

If latitude = 22.794687 and longitude = 89.498253 then the output of Sc = 13.1.

The range of the variables were shows in bottom of each column. In this study, the predicted concentrations were found 12.11 to 85.72 for Pb; 0.57 to 6.77 for Hg; 1.12 to 8.89 for Sb and 3.75 to 20.22 for Sc; whereas the measured concentrations were 12.11 to 90.55 for Pb; 0.72 to 9.2 for Hg; 1.01 to 12.55 for Sb and 3.75 to 20.41 for Sc. The higher intensity of concentration for Pb was shows the higher number of rules (Figure 4.3a) than that of others studied heavy metals in soils. On the contrary, Figure 4.3b illustrates only 11 rules due to lower intensity of concentration of Hg in soils. Furthermore, some unknown soil sampling points were considered as checking points (C) for the validation of predicted concentration with nearest soil sampling points (Table 4.4).

Table 4.4: Validation of predicted concentration of checking points with nearest soil sampling points

| Checking points (C)	Latitude	Longitude	Predicted Concentration		Nearest points of C	Latitude	Longitude	Measured Concentration	
			Pb (mg/kg)	Sb (mg/kg)				Pb (mg/kg)	Sb (mg/kg)
C1	22.79606	89.49891	41.3	8.28	S 9	22.79599	89.49835	40.76	7.77
C2	22.79610	89.49975	21.8	7.32	SS 44	22.79615	89.499611	23.16	5.33
C3	22.79555	89.49888	70.2	8.91	SS 4	22.79548	89.498896	70.87	8.87
C4	22.79515	89.50037	23.4	6.11	SS 27	22.79537	89.501574	26.65	4.88
C5	22.79709	89.50048	42.9	5.62	SS 21	22.79712	89.500262	40.77	5.61
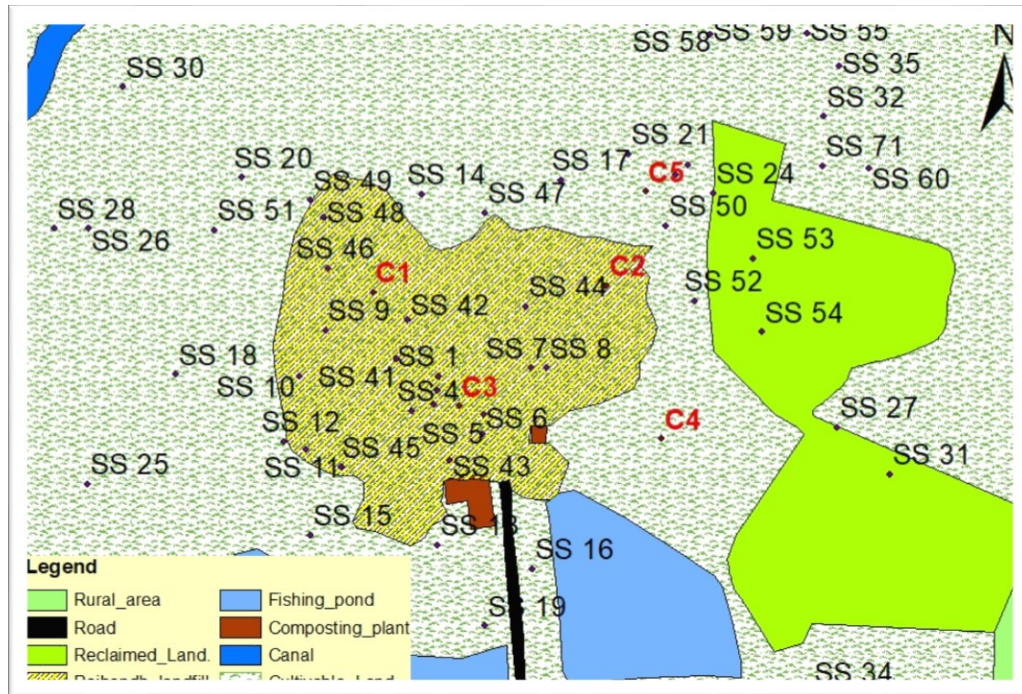
Figure 4.4: Location of some checking points in selected waste disposal site.

The location of some checking points in selected waste disposal site were shown in Figure 4.4. In Table 4.4, the predicted concentration and measured concentration of Pb for checking point C1 (41.3, 40.76); C2 (21.8, 23.16); C3 (70.2, 70.87); C4 (23.4, 26.65) and C5 (42.9, 40.77), respectively, were showed comparatively less variation of predicted results. Similarly, rule viewer provided the predicted concentration of Sb and showed very close value of measured concentration for the checking points. Here, it can be noted that one can easily be computed the concentration of particular heavy metals in soils of the selected waste disposal site by inserting only GPS values (latitude and longitude) in this developed model. Moreover, the rule viewer for other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, , Sr, V and Ba were reported in Figure A-1 to Figure A-12 in Appendix-A.

### 4.4.1.2.3 FIS Outputs of ANFIS

The Figures 4.5 (a), 4.5 (b), 4.5 (c) and 4.5 (d) represents the FIS output of training data set for the heavy metals of Pb, Hg, Sb and Sc in soils, respectively. The FIS result represents the closeness of predicted values with observed one with RMSE for the heavy metals of Pb, Hg, Sb and Sc in soils (Figure 4.5). In case of training data, RMSE values were found 1.38, 1.42, 0.91

and 0.12 for Pb, Hg, Sb and Sc, respectively. The minimum value of RMSE was found 0.12 for Sc (Figure 4.5d) indicating the predicted value from model was very close to measured concentration of Sc in the laboratory. On the other hand, the maximum RMSE (1.42) indicated the more variation of predicted and measured concentrations of Hg in soils. Moreover, the FIS output of training data for other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure A-13 to Figure A-24 in Appendix-A.
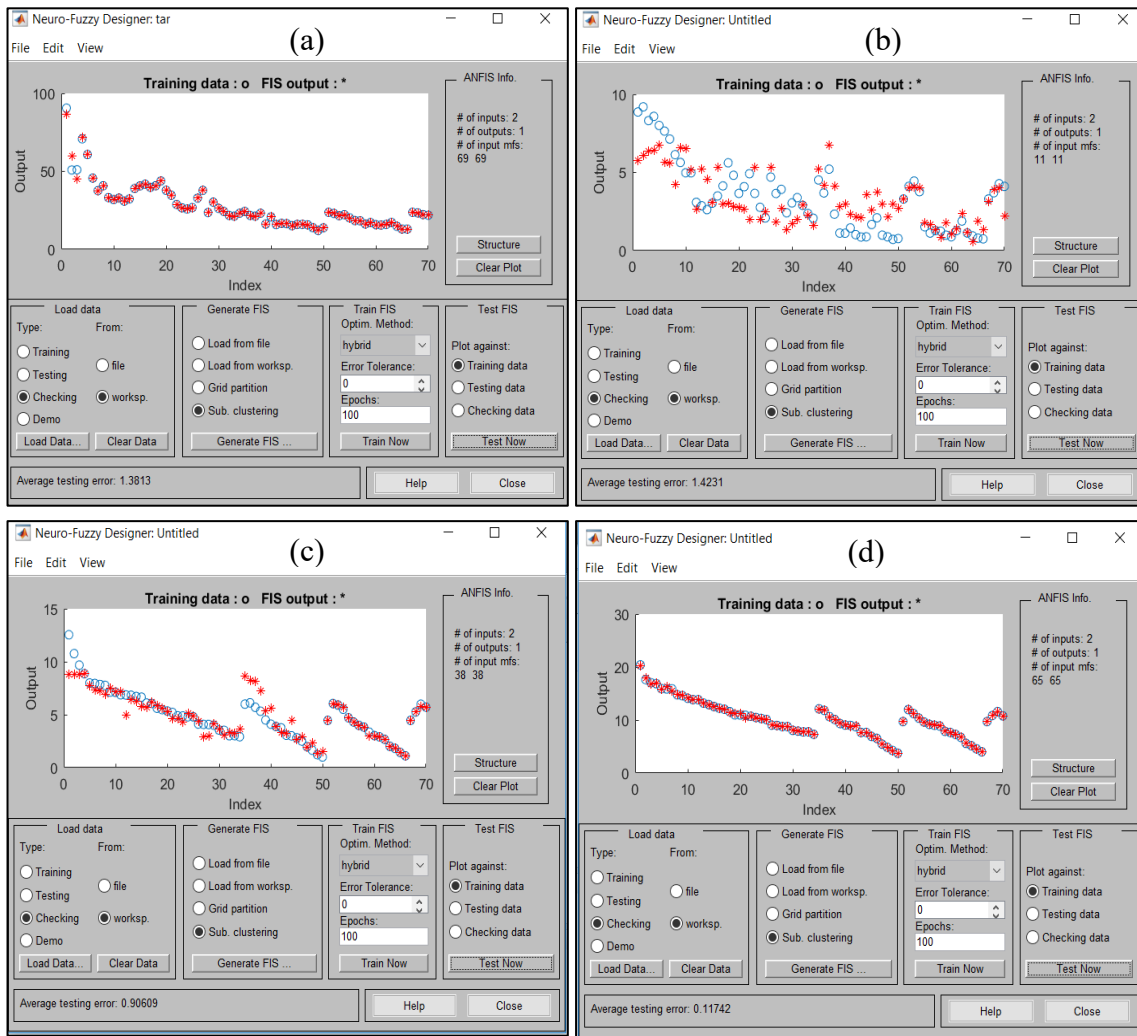
Figure 4.5: FIS output of training data for (a) Pb, (b) Hg, (c) Sb and (d) Sc.

Figure 4.6 (a), 4.6 (b), 4.6 (c) and 4.6 (d) represents the FIS output of testing data set for Pb, Hg, Sb and Sc, respectively. In case of testing data, RMSE values were found 15.25, 1.95, 2.29 and 2.97 for Pb, Hg, Sb and Sc, respectively. In testing, the minimum value of RMSE was found

1.95 (Figure 4.6b) for Hg showing insignificant variation of predicted values with measured one as well as maximum RMSE value was found 15.25 (Figure 4.6a) for Pb showing more variation between the predicted and measured values. Moreover, the FIS output of testing data for other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure A-25 to Figure A-36 in Appendix-A.
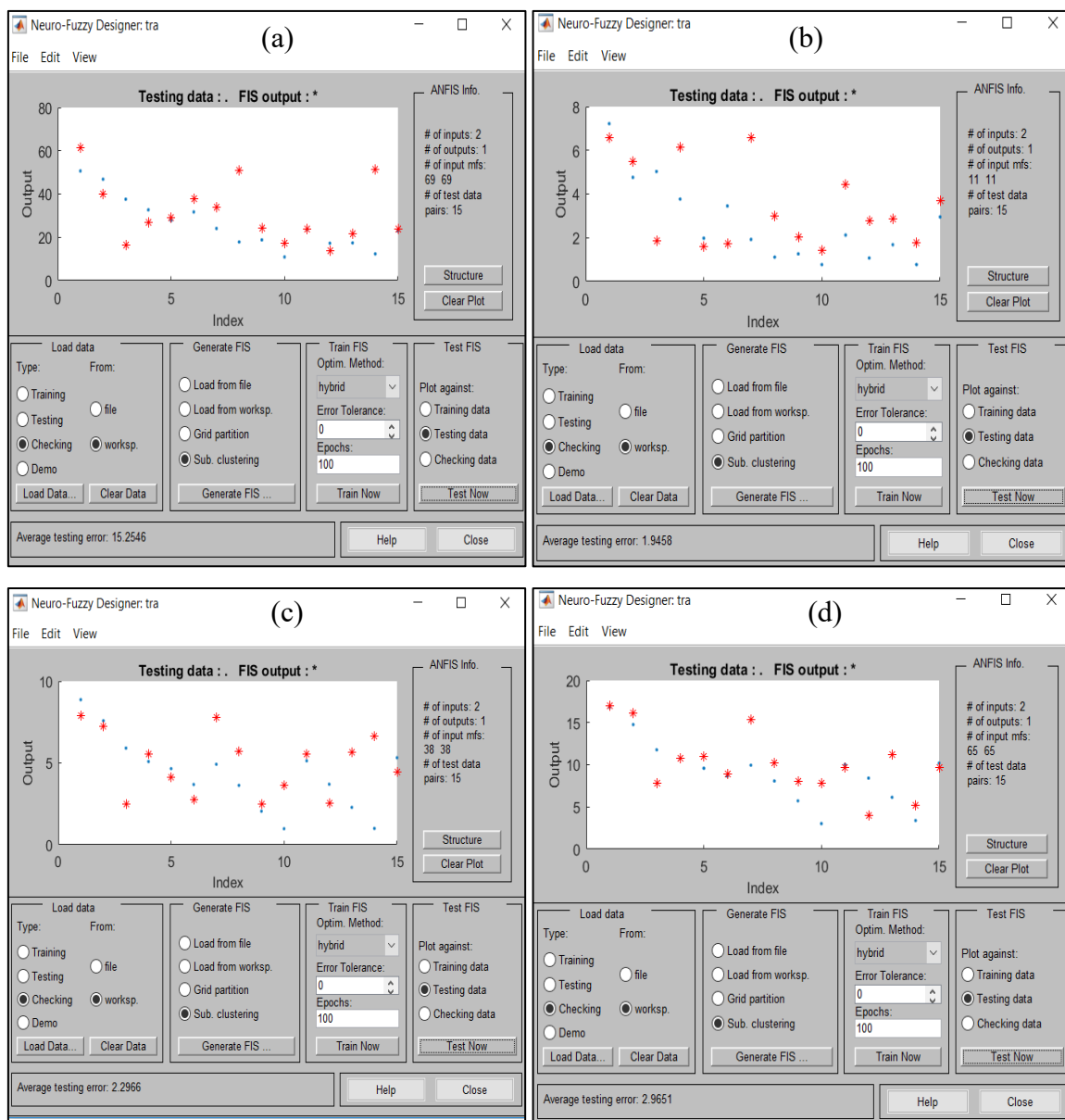
Figure 4.6: FIS output of testing data for (a) Pb, (b) Hg, (c) Sb and (d) Sc.

4.4.1.2.4 Surface Viewer of ANFIS

In ANFIS model, surface viewers represent the predicted values of heavy metal in soils. In this study, the predicted concentrations of heavy metal such as Pb, Hg, Sb and Sc in soils was presented by surface viewer shown in Figure 4.7a to Figure 4.7d, respectively. In these figures, the intensity of heavy metal concentrations represents by different colours.
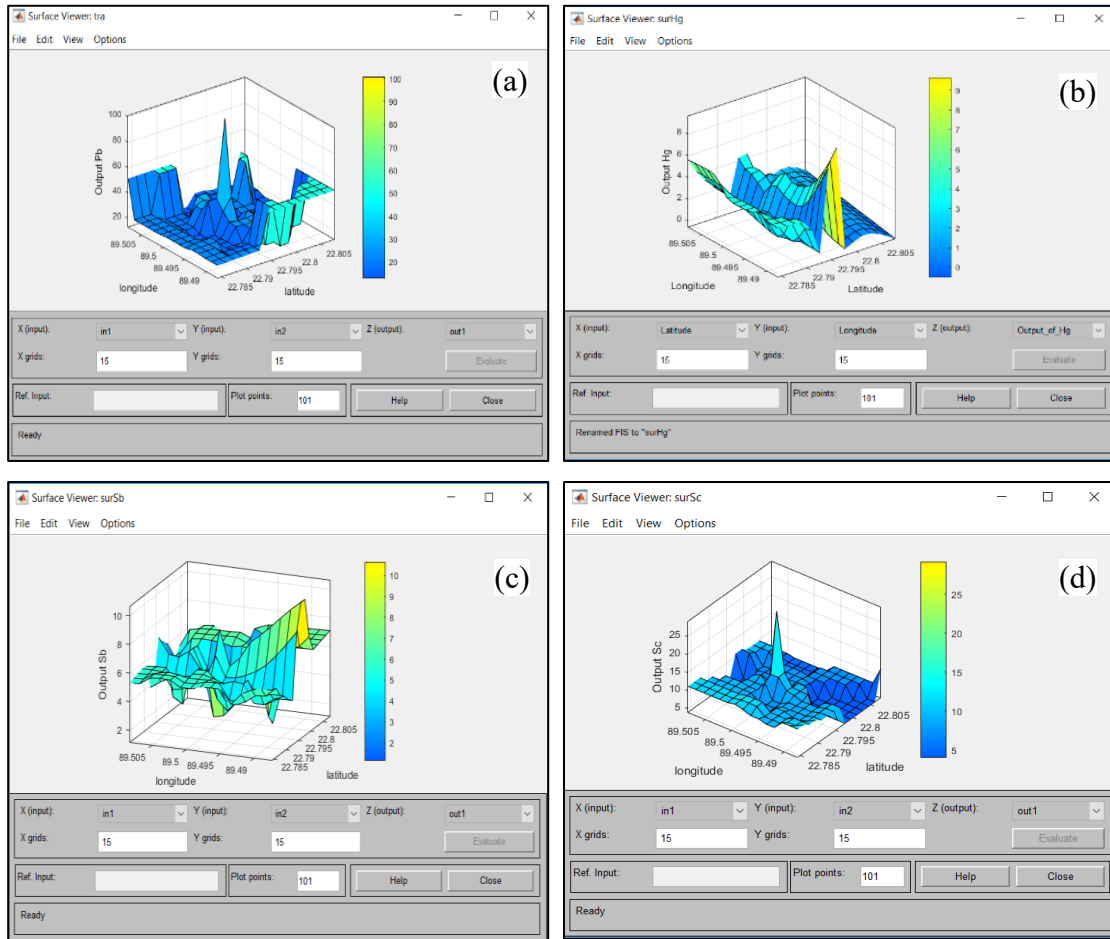
Figure 4.7: Surface viewer of the outputs for (a) Pb, (b) Hg, (c) Sb and (d) Sc.

In Figure 4.7a, the concentration of Pb varies from 12.11 to 85.72 mg/kg and the centre of the plot represents comparatively the higher values of Pb than that of the other soils sampling points. The yellow color in the colorbar represents comparatively the higher intensity of concentration than that of other colours (Figure 4.7). Figure 4.7b illustrates the concentration of Hg ranges from 0.57 to 6.77 mg/kg. Similarly, Figure 4.6c and Figure 4.6d represents the concentrations of Sb and Sc ranges 1.12 to 8.89 mg/kg and 3.75 to 20.22 mg/kg, respectively. Moreover, the

surface viewer of output for testing data of other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure A-37 to Figure A-48 in Appendix-A.

4.4.1.3 Assessment of Model Performance in ANFIS

In this study, the model A with SCP of gaussmf (input MF), linear (output MF) and hybrid algorithm (optimisation method) was selected based on the acceptable limits of R and RMSE for the prediction of heavy metal concentrations in soils. Table 4.5 shows the values of R and RMSE for training and testing for all studied heavy metals in soils.

Table 4.5: Results of R and RMSE for heavy metals in soils from ANFIS

Heavy metals	ANFIS (training)		ANFIS (testing)	
	R value	RMSE	R value	RMSE
Pb	0.99	1.38	0.55	15.25
Cu	0.98	0.63	0.86	2.45
Ni	0.81	1.00	0.71	1.17
Zn	0.998	0.67	0.63	9.18
Co	0.97	0.60	0.58	2.35
Cd	0.91	0.65	0.79	1.17
As	0.98	0.36	0.88	1.44
Sc	0.999	0.12	0.70	2.97
Hg	0.76	1.42	0.53	1.95
Mn	0.86	3.33	0.45	6.24
Cr	0.998	0.15	0.72	2.02
Ti	0.997	29.60	0.76	341.76
Sb	0.92	0.91	0.40	2.30
Sr	0.99	0.88	0.75	7.17
V	0.999	0.69	0.72	12.40
Ba	0.999	0.88	0.82	15.52

In training, the value of |R| were found greater than 0.9 for Pb, Cu, Zn, Co, Cd, As, Sc, Cr, Ti, Sb, Sr, V and Ba in soils indicating very strong correlation between input and output variables. In addition, |R| ≥ 0.8 for Ni and Mn indicated strong correlation and R-value was found 0.76 in the ranges of 0.2 < |R| < 0.8 for Hg which specified the existing of correlation between input and output variables. A study conducted by Smith (1986) and stated that when the values of R is greater than 0.90 indicated the very strong correlations. Therefore, the findings for almost all the studied heavy metals of this study are agreed well in the postulation stated by Smith (1986).

Among all the heavy metals, the maximum R-value of Sc was found 0.999 with the minimum RMSE 0.12 indicating the best correlation in prediction of Sc in soils of the selected waste disposal site.
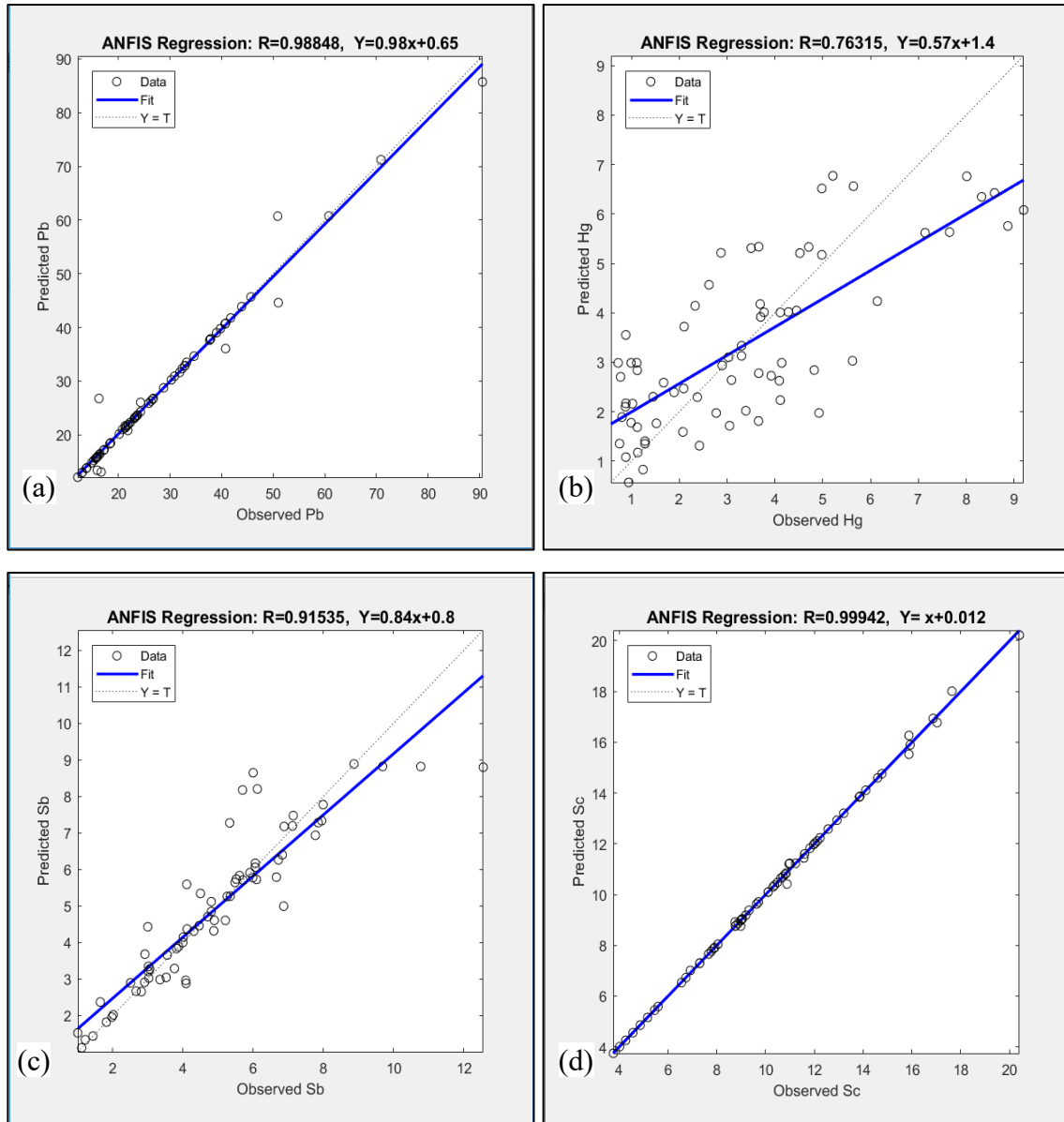
Figure 4.8: Regression analysis of ANFIS (training) for (a) Pb, (b) As, (c) Sb and (d) Sc.

On the other hand, the minimum R-value of Hg was found 0.76 with RMSE of 1.42 indicating the worst correlation for Hg than that of other studied heavy metals. For Ti, it was presented the maximum value of RMSE with 29.60 due to the higher intensity of concentrations than other

studied heavy metals in soils. The relationships of predicted and measured concentrations of Pb, Hg, Sb and Sc in soils in training is shown in Figures 4.8a, 4.8b, 4.8c and 4.8d, respectively. Figures reveal all the predicted concentrations were closed to the fitted line and Sc (Figure 4.8d) shows the best performance. Moreover, the ANFIS regression for training of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure A-49 to Figure A-60 in Appendix-A.
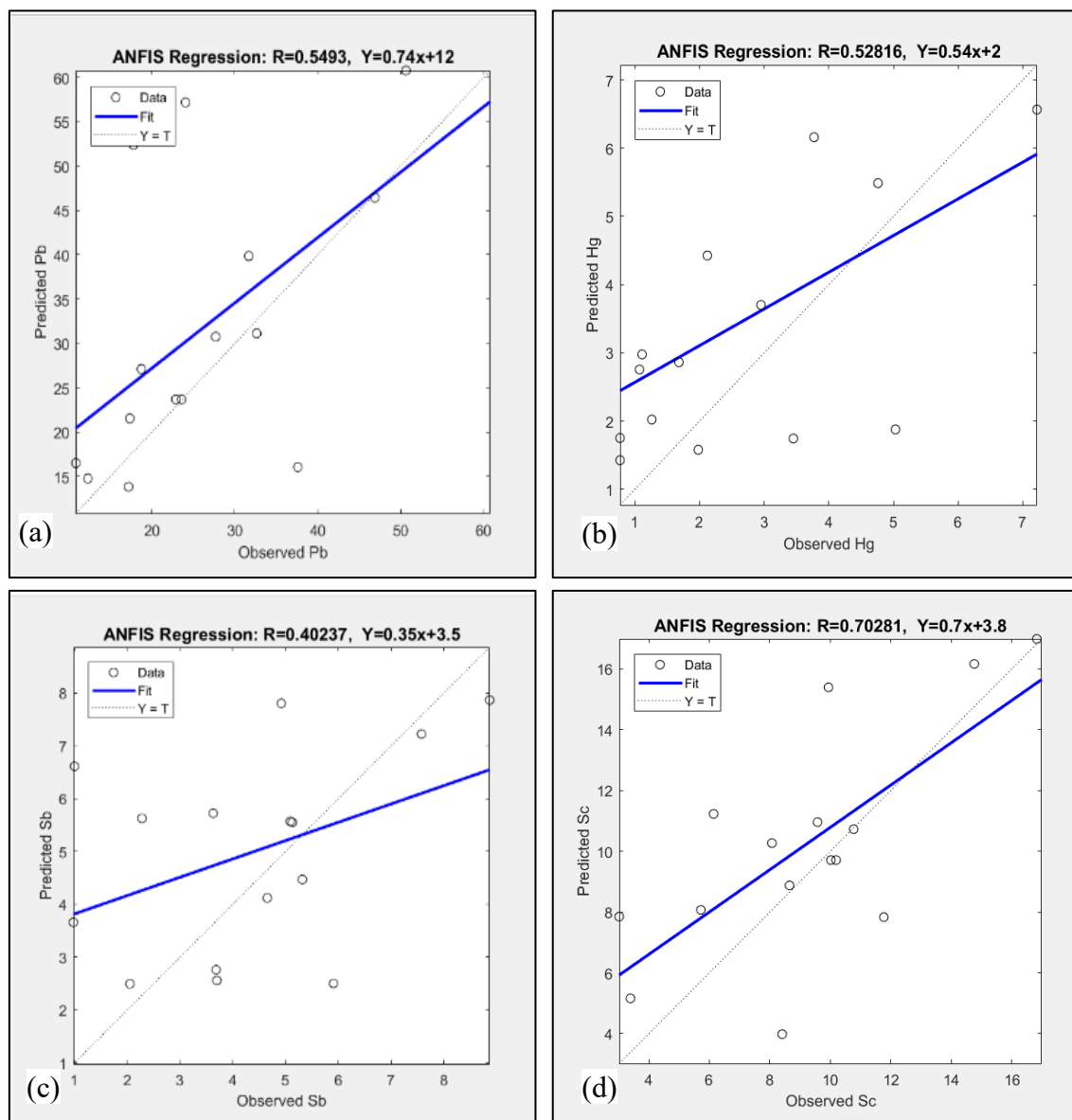
Figure 4.9: Regression analysis of ANFIS (testing) for (a) Pb, (b) As, (c) Sb and (d) Sc.

In addition for testing, the value of |R| for Cu, Hg and Ba were found greater than 0.8 indicating strong correlation. The R-value for Pb, Ni, Zn, Co, Cd, As, Sc, Hg, Mn, Cr, Ti, Sb, Sr, and V was found between the ranges of 0.2 < |R| < 0.8 specifying the existence of correlation between inputs and outputs. Among all the heavy metals in testing, the maximum R-value (0.88) with minimum RMSE (1.44) for As indicating the best correlation in prediction whereas R=0.40 (minimum) and RMSE=2.30 (maximum) for Sb indicating the worst correlation in prediction than that of other studied heavy metals. The relationships of predicted and measured concentrations of Pb, Hg, Sb and Sc in soils were represented in Figures 4.9a, 4.9b, 4.9c and 4.9d, respectively. The plotted data for Sc was closed to the fitted line (Figure 4.9d) whereas for Sb data was scattered from the fitted line (Figure 4.9c). Moreover, ANFIS regression for testing data of other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, , Sr, V and Ba were reported in Figure A-61 to Figure A-72 in Appendix-A.

4.4.1.3.1 Assessment of Pb in ANFIS

In this study, ANFIS model was conducted to predict heavy metal concentrations in soils of 15 unknown sampling points of the selected waste disposal site. The aim of this section was to evaluate the closeness of the predicted and measured concentration of Pb in soils. Table 4.6 represents the predicted and measured concentration of Pb in soils. The variation of predicted and measured concentration of Pb was clearly exposed in Figure 4.10. In this figure, most of the soil sampling points (10, 25, 30, 45, 50, 65, 70, 80 and 85) shows insignificant variation as well as others sampling points shows higher variation of concentrations of Pb in soils. Result reveals the measured values of 46.89 and 23.62, while, predicted values of 46.47 and 23.72 mg/kg, for soil sampling points 10 and 65, respectively, in case of Pb.

Table 4.6: Recovery level of Pb in ANFIS (testing)

Soils sampling points	Recovery level of Pb for testing in ANFIS		
	Measured concentration	Predicted concentration	% recovery
5	50.66	60.78	119.98
10	46.89	46.47	99.1
20	37.62	32.12	85.38
25	32.66	31.15	95.37
30	27.72	30.77	111.01
35	31.705	37.87	119.43
45	24.09	24.75	102.75
50	17.82	20.02	112.35
55	18.77	24.14	128.6
60	10.88	16.53	151.95
65	23.62	23.72	100.42
70	17.24	13.86	80.39
75	17.39	21.56	123.98
80	12.33	14.8	120.03
85	22.93	23.72	103.45
% mean recovery			110.28
RMSE			15.25
MAPE			15.58
GRI			1.2
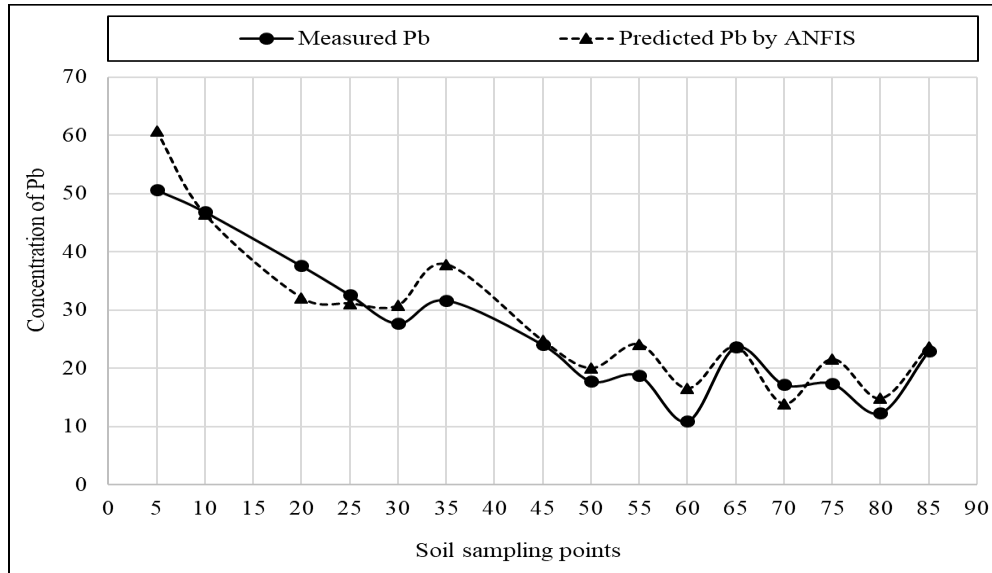
Figure 4.10: Comparison of predicted and measured concentration of Pb from ANFIS (testing).

92

Table 4.6 also demonstrates RMSE, MAPE, GRI and percentage recovery. The percentage recovery means what percentage of measured value is recovered by the predicted value. 20-200% recovery represents the acceptance of model, 100% recovery represents the fit level and 80-120% recovery represents the robustness of the model. The variation of recovery level of Pb for testing in ANFIS model is shown in Figure 4.11.
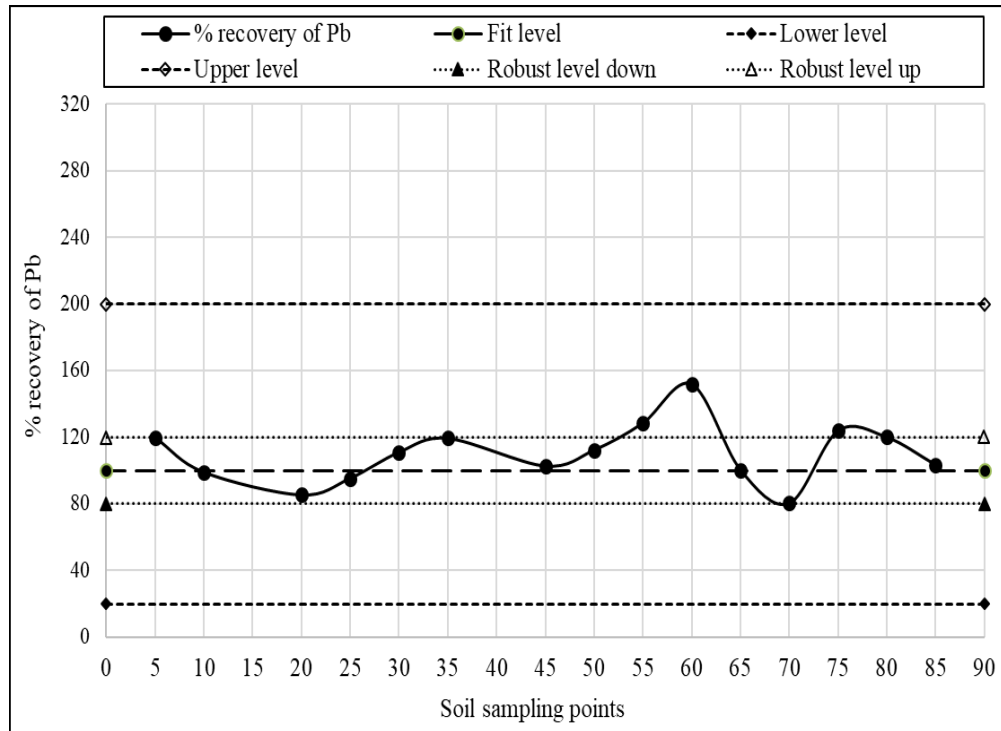
Figure 4.11: Variation of recovery level of Pb for ANFIS (testing).

In this figure, all the soil sampling points were in the acceptable ranges 20-200%. Besides, about 10 soil sampling points (5, 10, 25, 30, 35, 50, 65, 70, 80 and 85) were found very near and between the ranges 80-120% which indicating the robustness of model for Pb. Again, the MAPE value was found 15.58% for Pb that was in the acceptable ranges of 30-40% (Rayer, 2007). Furthermore, the value of GRI was found 1.2 (greater than 1) for Pb which representing the reliability of the model. Therefore, ANFIS model was considered as robust and reliable for predicting of Pb concentration in soils of the selected waste disposal site.

4.4.1.3.2 Assessment of Hg, Sb and Sc in ANFIS

The measured, predicted concentration as well as percentage recovery of the results of Hg, Sb and Sc in soils depicted in Table 4.7. The values of MAPE were found 56.57, 28.89 and 19.74 for Hg, Sb and Sc in soils, respectively. The value of error was minimum for Sc indicating closeness of predicted values with measured one of Sc in soils. The performance of Sb was better than Hg due to less error (28.89 < 56.57).

Table 4.7: Recovery level for Hg, Sb and Sc in ANFIS (testing)

Soils sampling points	Hg			Sb			Sc		
	Measured	Predicted	% recovery	Measured	Predicted	% recovery	Measured	Predicted	% recovery
5	7.22	6.56	90.92	8.87	7.87	88.72	16.83	16.99	100.94
10	4.76	5.49	115.26	7.58	7.23	95.36	14.76	16.16	109.51
20	5.03	4.88	96.96	5.91	4.50	76.24	11.77	10.83	92.03
25	3.77	3.16	83.82	5.09	5.57	109.45	10.77	10.73	99.59
30	1.98	1.58	79.74	4.65	4.12	88.59	9.58	10.96	114.43
35	3.45	1.74	50.54	3.68	3.76	102.12	8.65	8.88	102.62
45	1.92	2.56	133.38	4.92	4.81	97.68	9.94	12.39	124.69
50	1.11	2.38	214.12	3.63	5.73	157.72	8.07	10.27	127.25
55	1.26	2.02	160.37	2.05	2.50	121.76	5.72	8.07	141.12
60	0.77	1.43	185.07	0.98	1.66	169.33	3.02	4.85	160.54
65	2.12	3.43	161.58	5.13	5.55	108.26	10.02	9.71	96.90
70	1.07	2.76	257.51	3.70	2.56	69.19	8.41	5.98	71.12
75	1.68	2.86	170.34	2.28	4.63	203.04	6.14	7.23	117.75
80	0.77	1.75	227.43	1.00	1.61	161.49	3.39	5.16	152.22
85	2.95	3.70	125.44	5.32	4.47	84.07	10.19	9.71	95.29
% mean recovery			143.50	% mean recovery		115.53	% mean recovery		113.73
RMSE			1.95	RMSE		2.30	RMSE		2.97
MAPE			56.57	MAPE		28.89	MAPE		19.74
GRI			1.65	GRI		1.37	GRI		1.26

The variation of predicted and measured concentration of Hg, Sb and Sc in soils were illustrated in Figures 4.12, 4.13 and 4.14, respectively. The perfect prediction was observed in case of Hg in soils for the soil sampling points of 5, 10, 20, 25, 30, 45, 60 and 85 (Figure 4.12). In contrast, 35, 50, 55, 65, 70, 75 and 80 shows the less perfection for the prediction of the concentrations of Hg in soils.
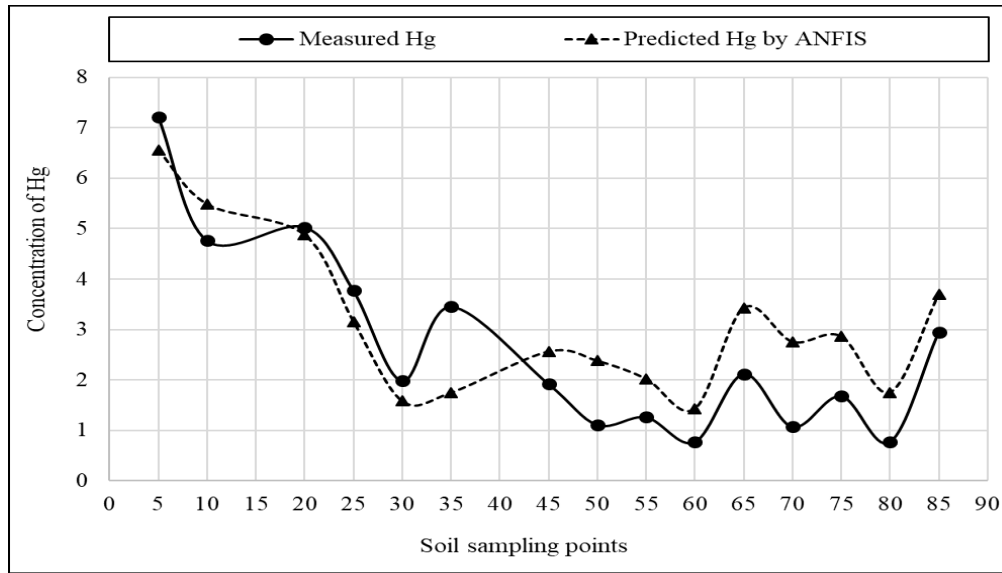
Figure 4.12: Comparison of predicted and measured concentration of Hg
from ANFIS (testing).

Similarly, the perfect prediction was observed for the soil sampling points 10, 25, 30, 35, 45, 55, 60, 65 and 85 for Sb (Figure 4.13), whereas, the soil sampling points 5, 20, 25, 35, 65, 75 and 85 were denoted the perfect prediction for Sc (Figure 4.14).
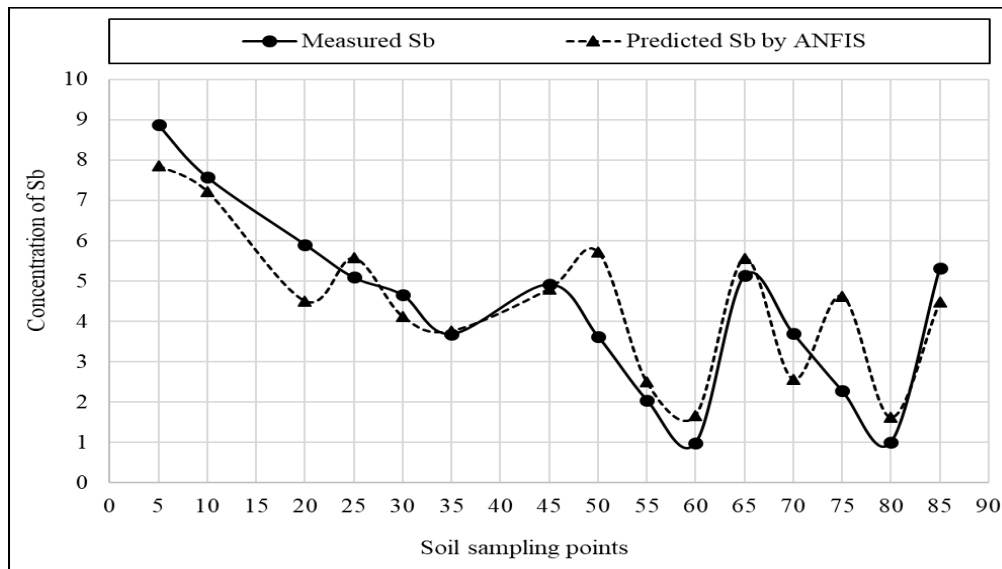
Figure 4.13: Comparison of predicted and measured concentration of Sb from
ANFIS (testing).
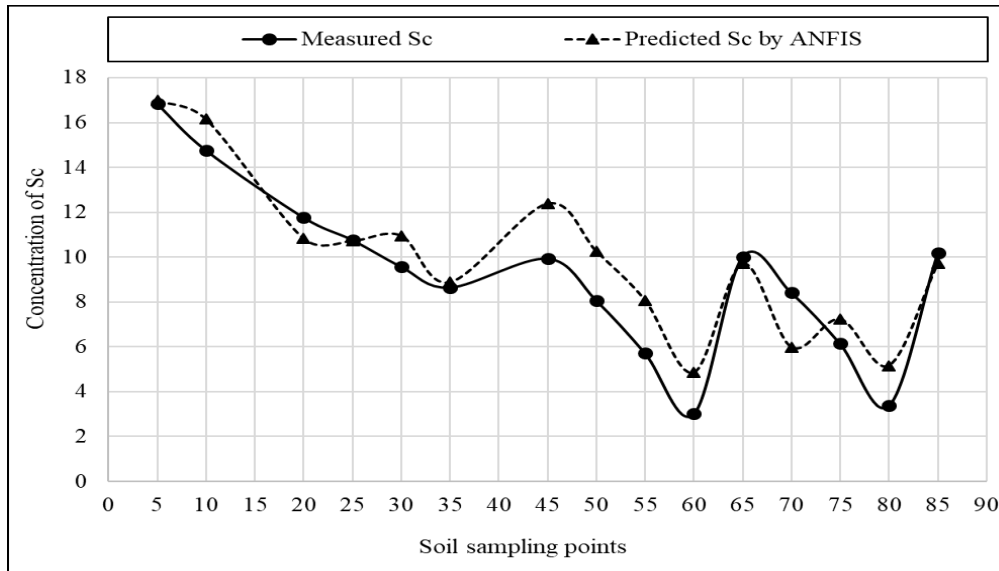
Figure 4.14: Comparison of predicted and measured concentration of Sc from ANFIS (testing).

The values of percentage recovery were found 143.50, 115.53 and 113.73 for Hg, Sb and Sc, respectively (Table 4.7). Among which for Sc, it was very close to fit level 100 than others. The level of percentage recovery for Hg, Sb and Sc were expressed in Figure 4.15, 4.16 and 4.17, respectively.
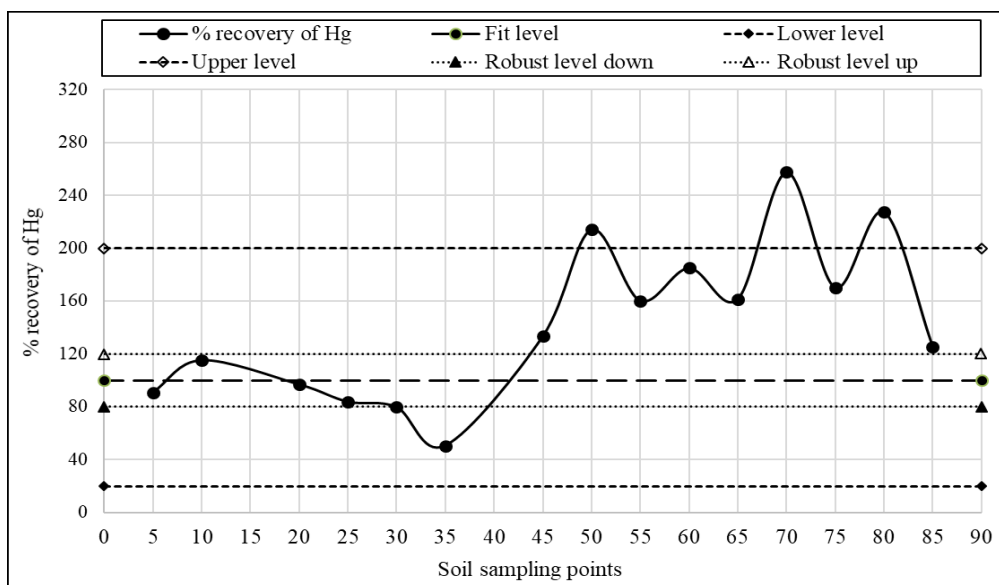
Figure 4.15: Variation of recovery level for Hg in ANFIS.

96

Figure 4.15 shows the percentage recovery ranges from 20-200% for all soil sampling points except 50, 70 and 80 in case Hg. Based on this result, it was recommended that the prediction of Hg in soils was in acceptable range. Among the accepted soil sampling points; 5, 10, 20, 25, 30 and 85 were in robust level (80-120%) (Figure 4.15).
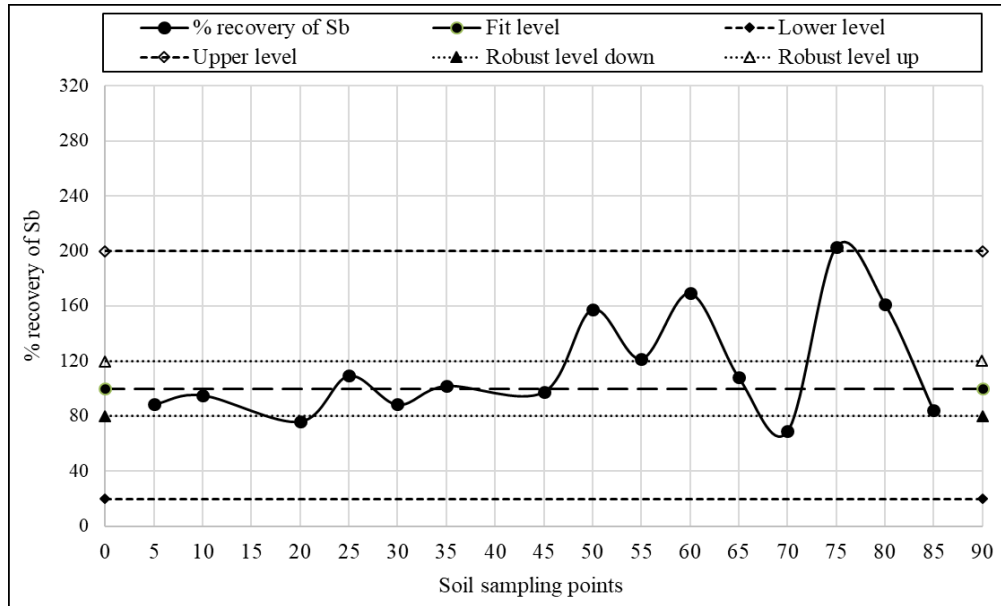
Figure 4.16: Variation of recovery level for Sb in ANFIS.
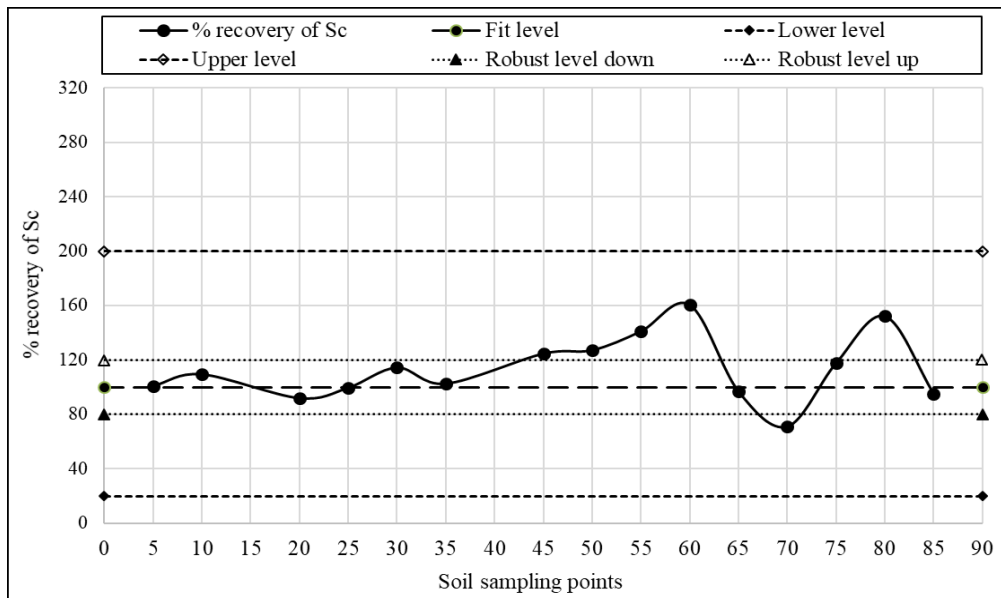
Figure 4.17: Variation of recovery level for Sc in ANFIS.

The percentage recovery of Sb were found in the acceptable range (Figure 4.16) for all soil sampling points except 75 as well as the percentage recovery for Sc were in the acceptable range (Figure 4.17) for all soil sampling points. Among of them, most of the soil sampling points were near to the fit level 100. In addition, GRI values were found 1.65, 1.37 and 1.26 (greater than 1) for Hg, Sb and Sc, respectively. Therefore, the prediction model for Hg, Sb and Sc were reliable. Moreover, relationships of predicted and measured concentrations of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils were also reported in Figure A-73 to Figure A-84, respectively, in the Appendix-A. In addition, the graphical representations of percentage recovery of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils were also reported in Figure A-85 to Figure A-96, respectively, in the Appendix-A. In contrast, the results of percentage recovery for Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils were also provided in Table B-1 to Table B-12, respectively, in Appendix-B.

4.4.1.3.3 Summary of Results in ANFIS Model

In this study, the genfis, sub-clustering partitioning (SCP) with different input membership function (MF) like gaussmf, trimf, trapmf, psigmf, gbellmf; output MF like linear and constant; optimization method such as hybrid or back-propagation (BP) as well as number of epochs was considered to select the best model of ANFIS. The model (SCP, gaussmf, linear and hybrid) showed the highest |R| (0.80) and lowest RMSE (1.52) values indicating the strong correlations and so it was selected as a general model of ANFIS for the prediction of heavy metal concentrations in soils at waste disposal site.

The concentration of various heavy metals with its sample location (latitude and longitude) were trained and tested to validate the selected model of ANFIS based on the acceptable limits of prediction parameters like R, RMSE, MAPE, GRI and percent recovery. In ANFIS analysis, the values of R for most of the heavy metals were found in the ranges of 0.81 to 0.999 indicating the best correlation in prediction. Among all the heavy metals, the maximum R-value was found 0.999 with the minimum RMSE 0.12 for Sc. The values of MAPE were found 56.57, 28.89 and 19.74 for Hg, Sb and Sc in soils, respectively. The value of error was minimum for Sc indicating closeness of predicted values with measured one of Sc in soils. The performance of Sb was

better than Hg due to less error (28.89 < 56.57). The values of percentage recovery were found 143.50, 115.53 and 113.73 for Hg, Sb and Sc, respectively. Among which for Sc, it was very close to fit level 100 than others. The percentage recovery of Sb were found in the acceptable range for all soil sampling points except 75 as well as the percentage recovery for Sc were in the acceptable range for all soil sampling points. Among of them, most of the soil sampling points were near to the fit level 100. In addition, GRI values were found 1.65, 1.37 and 1.26 (greater than 1) for Hg, Sb and Sc, respectively, which representing the reliability of prediction model. In ANFIS analysis, if the latitude and longitude of 22.798701 and 89.498029, respectively, were considered as input in the developed model for Sc, the output/ predicted value of Sc was obtained as 12.11, whereas measured value was 12.12. Here, it can be noted that one can easily be computed the concentration of particular heavy metals in soils of the selected waste disposal site by using latitude and longitude only in the developed model of ANFIS.

4.4.2 Support Vector Machine

In this study, the model validation for various kernel functions of support vector machine (SVM), performance of the model prediction with prediction parameters and the results of SVM were deliberated successively. The results of SVM for the heavy metals of Pb, Hg, Sb and Sc in soils were discussed in details here. Moreover, the results of other heavy metals in SVM of Cu, Ni, Zn, Co, Cd, Hg, Mn, Cr, Sb, Sr, V and Ba were reported in Appendix-C (Figures) and Appendix-D (Tables), respectively.

4.4.2.1 Validation of Models in SVM

In this study, total sixteen models (A to D with 5, 10, 15 and 20 fold numbers) for SVM analysis were formed with different kernel functions like linear-SVM (SVM-L), quadratic-SVM (SVM-Q), cubic-SVM (SVM-C) and gaussian or radial basis function-SVM (SVM-RBF) for fold numbers 5, 10, 15 and 20. The selected model was then compared in terms of the best values of R and RMSE to assess the performance of each model. Table 4.8 illustrates the sixteen models of different kernel functions of SVM for As, Co and Cd. The R-values were observed within the range of $0.2 < |R| < 0.8$ for most of the model in case of As and Co indicating correlation between the variables. The R-values of Cd were found negative for most of the models (B-5, B-10, A-

15, B-15, C-15, B-20 and C-20) indicating downhill correlation between the variables. In this analysis, Arsenic (As) was considered in compare to other results of heavy metals for selecting the best model of SVM.

Table 4.8: Validation of SVM model with different kernel functions

Model No.	Fold No.	Kernel Function	As		Co		Cd	
			R	RMSE	R	RMSE	R	RMSE
A-5	5 Fold	SVM-L	0.20	1.92	0.41	2.29	0.10	1.52
B-5		SVM-C	0.47	1.73	0.58	2.04	-0.63	1.81
C-5		SVM-Q	0.45	1.76	0.61	2.00	0.17	1.50
D-5		SVM- RBF	0.69	1.41	0.66	1.89	0.61	1.21
A-10	10 Fold	SVM-L	0.24	1.91	0.40	2.30	0.10	1.52
B-10		SVM-C	0.45	1.76	0.58	2.03	-0.54	1.74
C-10		SVM-Q	0.44	1.77	0.62	1.97	0.14	1.51
D-10		SVM- RBF	0.75	1.31	0.69	1.81	0.61	1.21
A-15	15 Fold	SVM-L	0.24	1.91	0.41	2.28	-0.10	1.52
B-15		SVM-C	0.51	1.69	0.62	1.97	-0.40	1.65
C-15		SVM-Q	0.42	1.78	0.64	1.93	-0.14	1.54
D-15		SVM- RBF	0.76	1.27	0.70	1.78	0.62	1.20
A-20	20 Fold	SVM-L	0.24	1.9	0.40	2.29	0.10	1.52
B-20		SVM-C	0.49	1.71	0.58	2.04	-0.47	1.69
C-20		SVM-Q	0.45	1.76	0.62	1.96	-0.24	1.57
D-20		SVM- RBF	0.75	1.3	0.69	1.80	0.60	1.22

The variation of R and RMSE for As were clearly expressed in Figure 4.18 and 4.19, respectively, for different kernel function and fold number. The SVM-RBF shows the higher value of R than that of other kernel functions of SVM-L, SVM-Q and SVM-C in SVM for fold numbers 5, 10, 15 and 20 (Figure 4.18). Among the entire folds considered in this analysis, the fold number 15 shows the maximum R-value (0.76) with minimum RMSE (1.27).
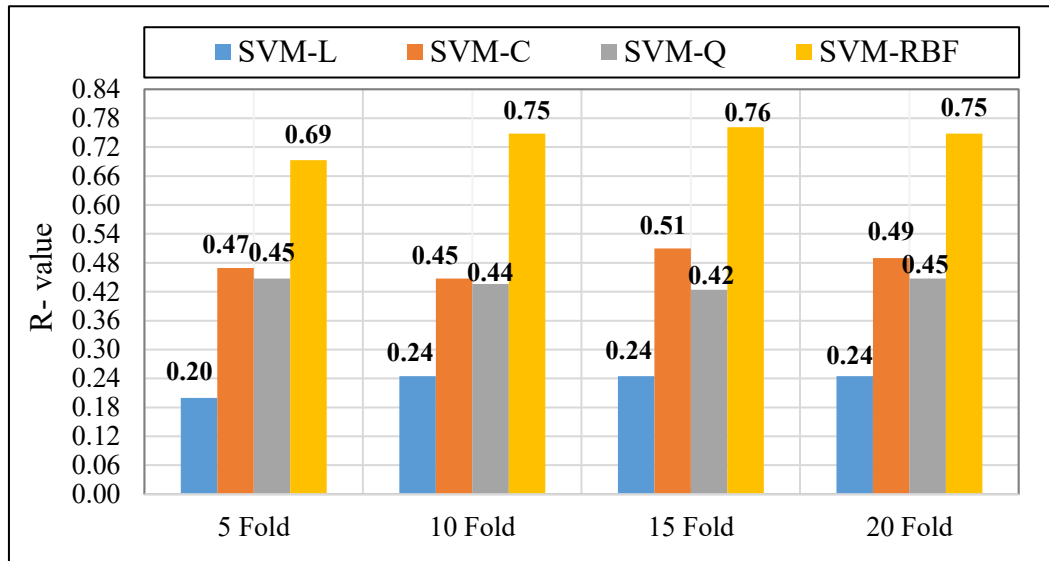
Figure 4. 18: Variation of R with different fold numbers and kernel functions of SVM for As.

Similarly, Figure 4.19 reveals the lower value of RMSE for SVM-RBF than that of other kernel functions of SVM-L, SVM-Q and SVM-C in case of fold numbers 5, 10, 15 and 20. It reveals the fold number 15 showing the minimum RMSE of 1.27 than that of other fold numbers.
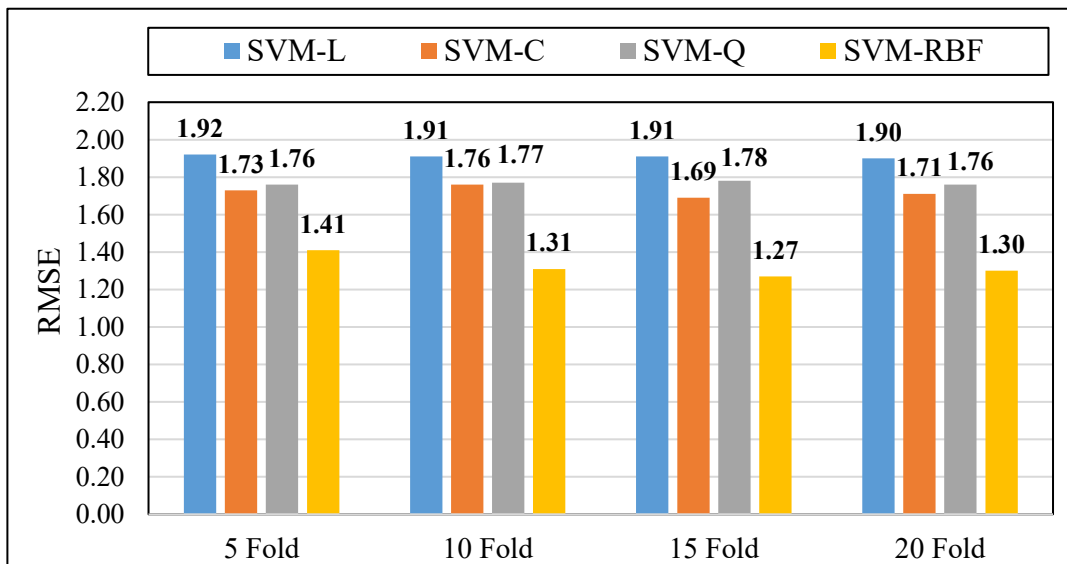
Figure 4.19: Variation of RMSE with different kernel function and fold number for As.
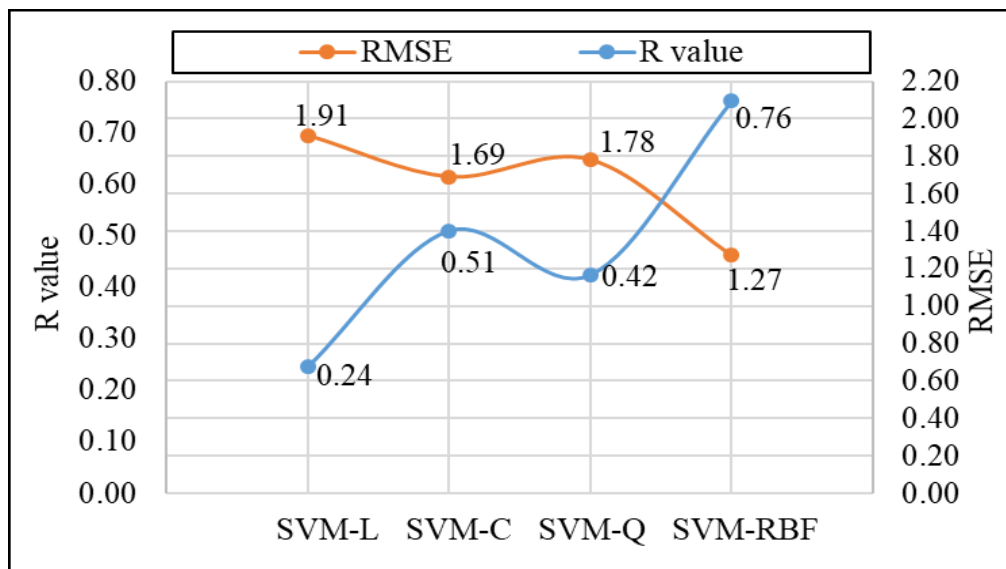
Figure 4.20: Variation of R and RMSE with different kernel functions of SVM
for selected fold number of As in soil.

Figure 4.20 entirely shows that fold number 15 provides the maximum R-value (0.76) with minimum RMSE (1.27) for model D-15 (SVM-RBF with 15 folds). Based on aforementioned, R and RMSE, model D-15 (SVM-RBF with 15 folds) was selected for the modelling of heavy metal analysis in soils of selected waste disposal site in SVM.

4.4.2.2 Graphical Representation of Results in SVM

The outputs of SVM-RBF model were represented by contour map, surface viewer and residuals as the results of predicted concentrations for heavy metals in soils. The representation of outputs for Pb, Hg, Sb and Sc in soils were explained in Figure 4.21 to Figure 4.24, respectively. In Figure 4.21, the surface viewer shows the predicted concentration of Pb in soils. The intensity of predicted Pb was represented by the colorbar whereas yellow color shows the higher intensity of the concentration of Pb. In SVM analysis, the predicted concentration of Pb varies from 13.34 to 49.75 mg/kg in soils (whereas measured value 12.11 to 90.55 mg/kg).
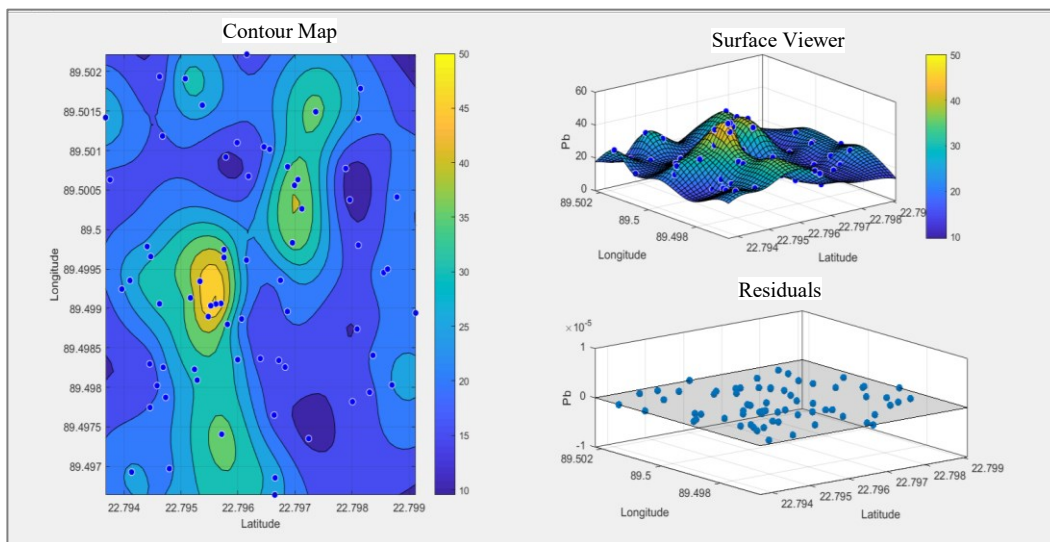
Figure 4.21: Graphical representation of the outputs for Pb.

Besides, contour map shows the individual color area of same concentration of Pb in soils. In addition, residuals represent the difference of predicted and measured heavy metal concentrations in soils. In Figure 4.22, surface viewer shows the better predicted concentration of Hg in the ranges of 0.96 to 8.68 mg/kg instead of measured concentration of 0.72 to 9.20 mg/kg than that of other heavy metals in soils.
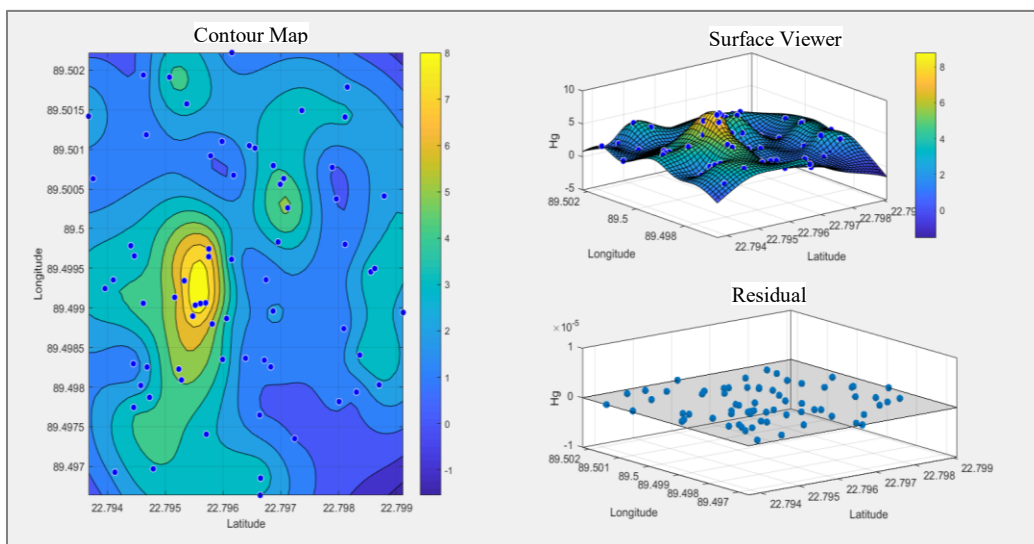
Figure 4.22: Representation of the outputs for Hg.

Similarly, Figure 4.23 and Figure 4.24 represents the concentration of Sb and Sc in the range of 1.41 to 9.70 mg/kg and 7.73 to 12.16 mg/kg, respectively.
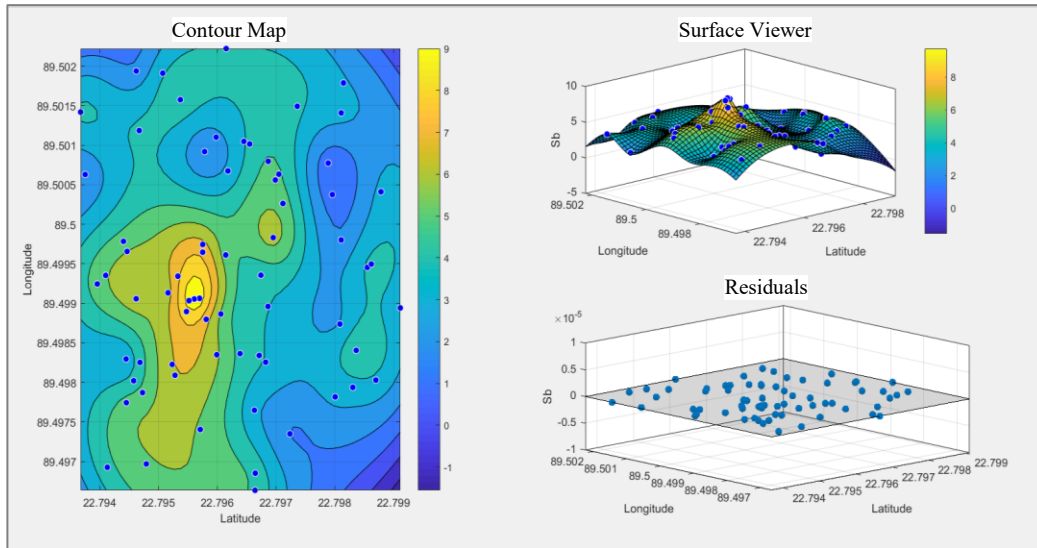
Figure 4.23: Representation of the outputs for Sb.
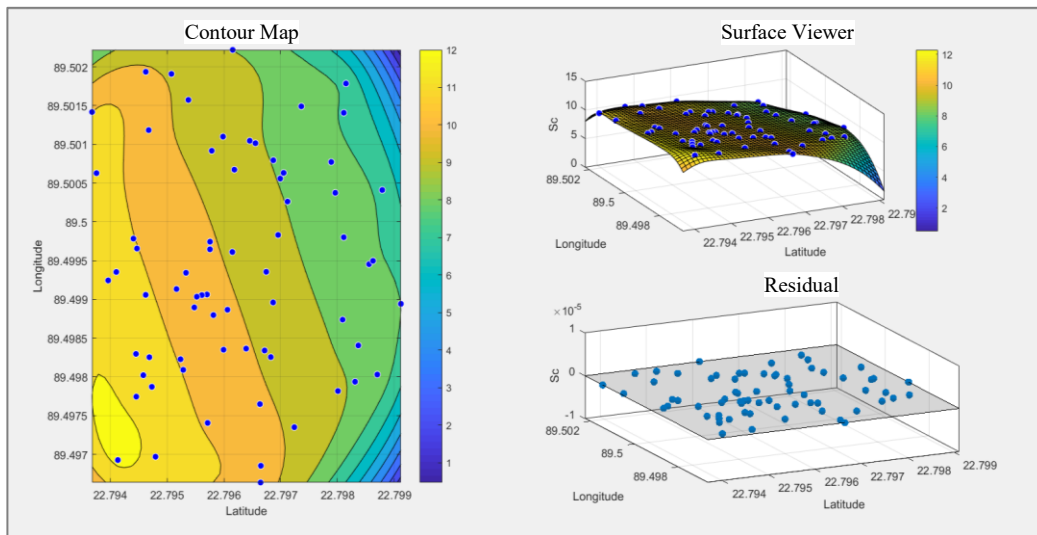
Figure 4.24: Representation of the outputs for Sc

Moreover, the representation of output for testing data with contour map, surface viewer and residuals of other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure C-1 to Figure C-12 in Appendix-C.

4.4.2.3 Assesment of Model Performance in SVM

The model SVM-RBF with 15-folds was selected based on the acceptable limits of R and RMSE for the prediction of heavy metal concentrations in soils. Table 4.9 shows the values of R and RMSE for training and testing data for all studied heavy metals in soils. In training, the value of $|R|$ were found in the ranges of $0.2 < |R| < 0.8$ for all heavy metals which specified the correlation between input and output variables. Among all the heavy metals, the maximum R-value was found 0.78 for Hg with the minimum RMSE 1.29 indicating the best correlation for predicting of Hg in soil. On the other hand, the minimum R-value was found 0.37 for Sc with the RMSE value 3.22 indicating the worst correlation in prediction for Sc than that of other studied heavy metals in soils. For Ti, it was found the maximum RMSE value 279.98 due to the higher intensity of concentrations than all other heavy metals.

Table 4.9: Variation of R and RMSE of heavy metals for SVM

Heavy metals	SVM(training)		SVM(testing)	
	R value	RMSE	R value	RMSE
Pb	0.66	10.64	0.43	10.73
Cu	0.71	2.30	0.73	2.09
Ni	0.65	1.31	0.60	1.53
Zn	0.63	7.59	0.37	11.01
Co	0.68	1.85	0.70	1.91
Cd	0.60	1.22	0.51	1.40
As	0.75	1.30	0.59	1.64
Sc	0.37	3.22	0.14	3.85
Hg	0.78	1.29	0.47	2.03
Mn	0.53	5.52	0.06	7.24
Cr	0.49	2.01	0.26	2.66
Ti	0.74	279.98	0.50	422.11
Sb	0.72	1.57	0.49	2.06
Sr	0.73	5.79	0.47	8.25
V	0.72	10.91	0.49	15.35
Ba	0.75	15.36	0.65	18.90

The relationships of predicted and measured concentrations of Pb, Hg, Sb and Sc in soils in training is shown in Figures 4.25a to 4.25d, respectively. Figures reveal all the predicted concentrations were closed to the fitted line and Hg (Figure 4.25b) shows the best performance

with better R-value than others do. Moreover, the results of other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils through SVM regression analysis (training) were reported in Figure C-13 to Figure C-24 in Appendix-C.
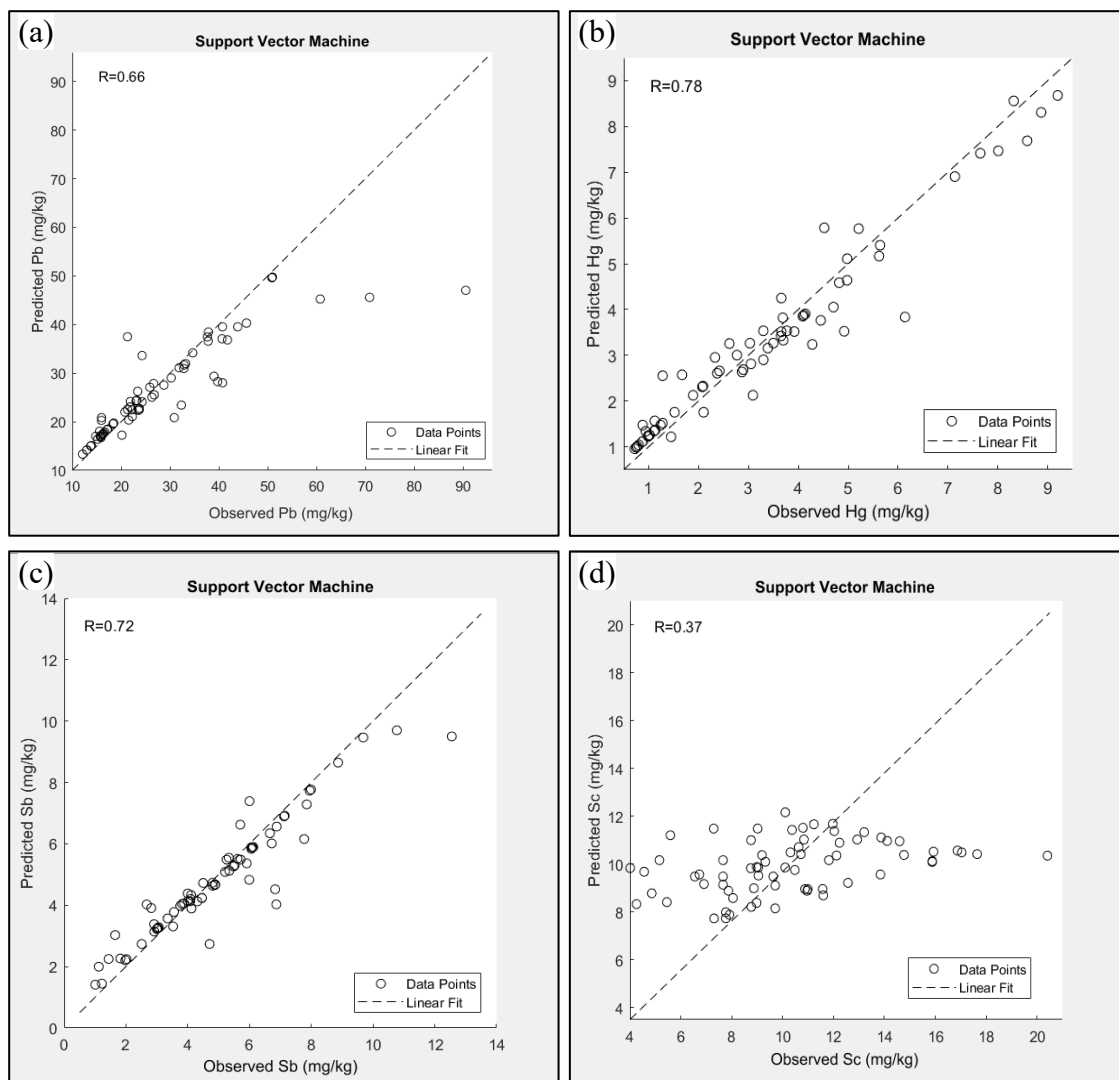
Figure 4.25: Regression analysis of SVM (training) for (a) Pb, (b) As, (c) Sb and (d) Sc.

In testing, the R-value was found between the ranges of $0.2 < |R| < 0.8$ for Pb, Cu, Ni, Zn, Co, Cd, As, Hg, Cr, Ti, Sb, Sr, V and Ba in soils which specified the correlation between the inputs and outputs of SVM-RBF model. Among all the heavy metals in testing, maximum R-value was found 0.73 for Cu with RMSE 2.03; in addition, minimum R-value was found 0.06 with maximum RMSE 7.24 for Mn. The relationship of the predicted and measured concentration of

Pb, Hg, Sb and Sc in soils were presented in Figures 4.26a to 4.26d, respectively. The plotted concentrations were closed to the fitted line for Sb (Figure 4.26c) whereas for Sc, concentrations were scattered from the fitted line (Figure 4.26d). Therefore, the predicted concentrations of Sb from SVM-RBF (testing) was more accurate than that of Pb, Hg and Sc in soils (Figure 4.26). Moreover, the results from SVM regression analysis (testing) for other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils were reported in Figure C-25 to Figure C-36 in Appendix-C.
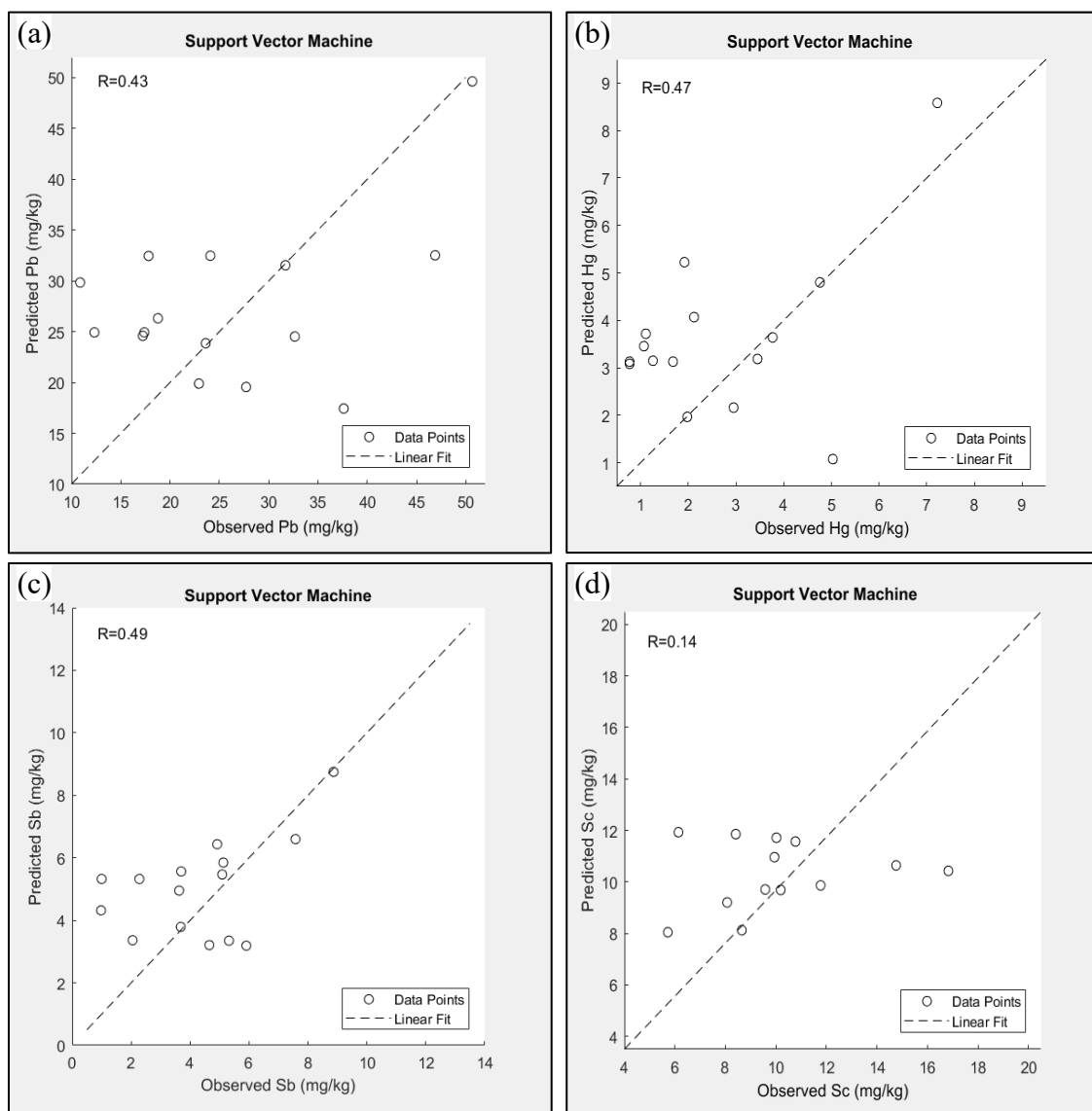
Figure 4.26: Regression analysis of SVM (testing) for (a) Pb, (b) As, (c) Sb and (d) Sc.

4.4.2.3.1 Assessment of Pb of SVM

In this study, SVM model was directed to predict heavy metal concentrations in soils of 15 unknown sampling points of the selected waste disposal site. The aim of this section was to evaluate the closeness of the predicted and measured concentration of Pb in soils. Table 4.10 represents the predicted and measured concentration of Pb in soils. The variation of predicted and measured concentration of Pb was clearly exposed in Figure 4.27. In this figure, only few soil sampling points (5, 35, 65 and 85) shows the minor variation as well as all other points' shows the highest variation of concentrations of Pb in soils. Result reveals the measured values of 31.71 and 23.62 mg/kg, while predicted values of 31.54 and 23.86 mg/kg for soil sampling points 35 and 65, respectively in case of Pb.

Table 4.10: Recovery level of Pb in SVM (testing)

Soils sampling points	Recovery level of Pb for testing in SVM		
	Measured concentration	Predicted concentration	% recovery
5	50.66	49.63	97.97
10	46.89	32.52	69.35
20	37.62	17.44	46.36
25	32.66	24.52	75.07
30	27.72	19.56	70.56
35	31.705	31.54	99.47
45	24.09	32.47	134.79
50	17.82	32.45	182.11
55	18.77	26.32	140.21
60	10.88	29.84	274.31
65	23.62	23.86	101.01
70	17.24	24.59	142.64
75	17.39	24.93	143.35
80	12.33	24.93	202.18
85	22.93	19.89	86.74
% mean recovery			124.41
RMSE			10.73
MAPE			45.01
GRI			1.58

108
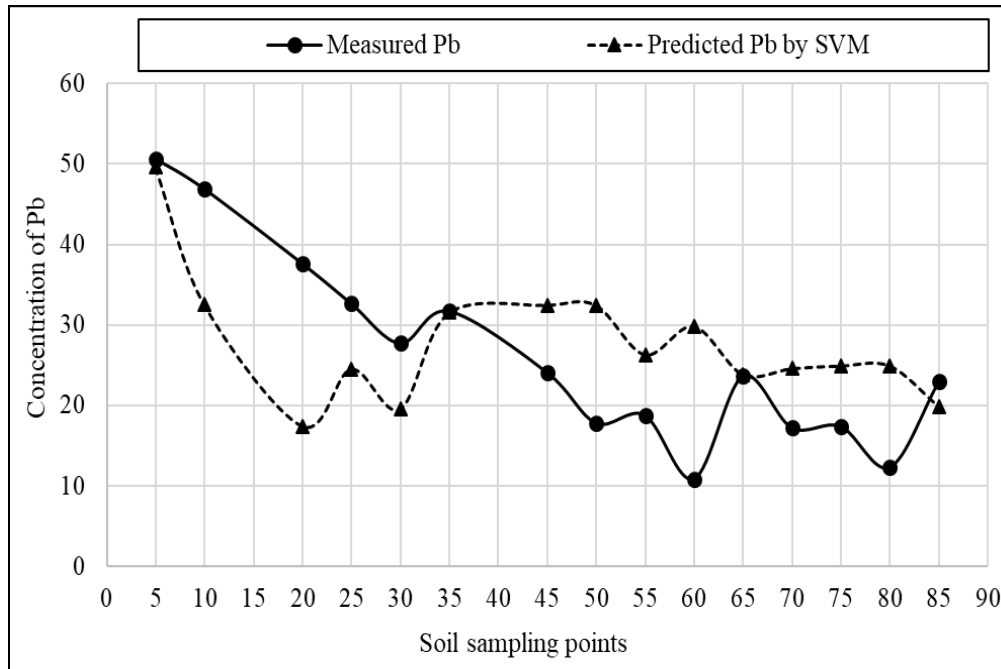
Figure 4.27: Comparison of predicted and measured concentration of Pb from SVM (testing).

Table 4.10 also demonstrates the results of RMSE, MAPE, GRI and percentage recovery. The percentage recovery means what percentage of measured value is recovered by the predicted value. The variation of recovery level for Pb in ANFIS model (testing) is shown in Figure 4.28. In this figure, all soil sampling points were in the acceptable ranges 20-200%, except soil sampling point 60. Besides, some soil sampling points like 5, 35, 65 and 85 were found very near and between the ranges of 80-120% which indicating the robustness of model for Pb in SVM analysis.
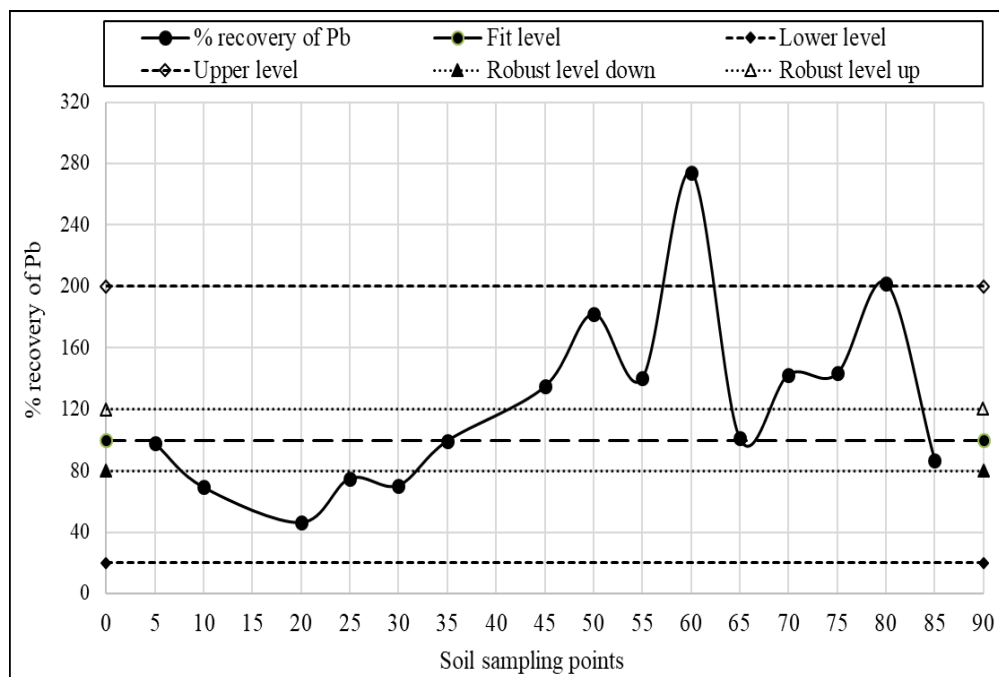
Figure 4.28: Variation of recovery level for Pb in SVM.

In addition, the MAPE value was found 45.01% which was very close to the acceptable ranges of 30 to 40% for Pb. Furthermore, the value of GRI was found 1.58 (greater than 1) for Pb which representing the reliability of model. Therefore, SVM model was considered as robust and reliable for predicting the Pb concentration in soils of the selected waste disposal site.

4.4.2.3.2 Assessment of Hg, Sb and Sc in SVM

The results of measured, predicted concentration as well as percentage recovery of Hg, Sb and Sc in soils depicts in Table 4.11. The values of MAPE were found 113.45, 57.14 and 50.19 for Hg, Sb and Sc, respectively. Minimum error was found for Sc indicating the closeness of the predicted values with measured one of Sc in soils. The performance of Sb was better than Hg due to less error (57.14 < 113.45).

Table 4.11: Recovery level for Hg, Sb and Sc in SVM (testing)

Soils sampling points	Hg			Sb			Sc		
	Measured	Predicted	(%) recovery	Measured	Predicted	(%) recovery	Measured	Predicted	(%) recovery
5	7.22	8.58	118.86	8.87	8.75	98.64	16.83	10.43	61.96
10	4.76	4.80	100.83	7.58	6.60	87.07	14.76	10.64	72.09
20	5.03	1.08	21.39	5.91	3.19	53.97	11.77	9.86	83.81
25	3.77	3.64	96.54	5.09	5.47	107.43	10.77	11.57	107.40
30	1.98	1.97	99.28	4.65	3.21	68.95	9.58	9.71	101.33
35	3.45	3.19	92.35	3.68	3.79	102.89	8.65	8.12	93.88
45	1.92	5.22	272.12	4.92	6.43	130.79	9.94	10.96	110.27
50	1.11	3.72	334.86	3.63	4.95	136.40	8.07	9.20	114.02
55	1.26	3.15	249.89	2.05	3.36	163.99	5.72	8.04	140.60
60	0.77	3.08	400.47	0.98	2.52	256.92	3.02	8.52	282.04
65	2.12	4.07	191.89	5.13	5.85	113.99	10.02	11.72	116.92
70	1.07	3.46	323.02	3.70	5.56	150.37	8.41	11.85	140.93
75	1.68	3.13	186.25	2.28	5.32	233.49	6.14	11.93	194.26
80	0.77	3.13	406.36	1.00	3.32	332.36	3.39	11.93	351.85
85	2.95	2.16	73.22	5.32	3.35	62.95	10.19	9.69	95.04
% mean recovery			197.82	% mean recovery		140.01	% mean recovery		137.76
RMSE			2.03	RMSE		2.06	RMSE		3.85
MAPE			113.45	MAPE		57.14	MAPE		50.19
GRI			2.30	GRI		1.68	GRI		1.60

The variation of predicted and measured concentration of Hg, Sb and Sc were illustrated in Figures 4.28, 4.29 and 4.30, respectively. The perfect prediction was observed in case of Hg in soils for the soil sampling points of 5, 10, 30, 35 and 85 (Figure 4.29). In contrast, 20, 45, 50, 55, 60, 65, 70, 75 and 80 were exposed comparatively worse prediction than that of others soil sampling points.
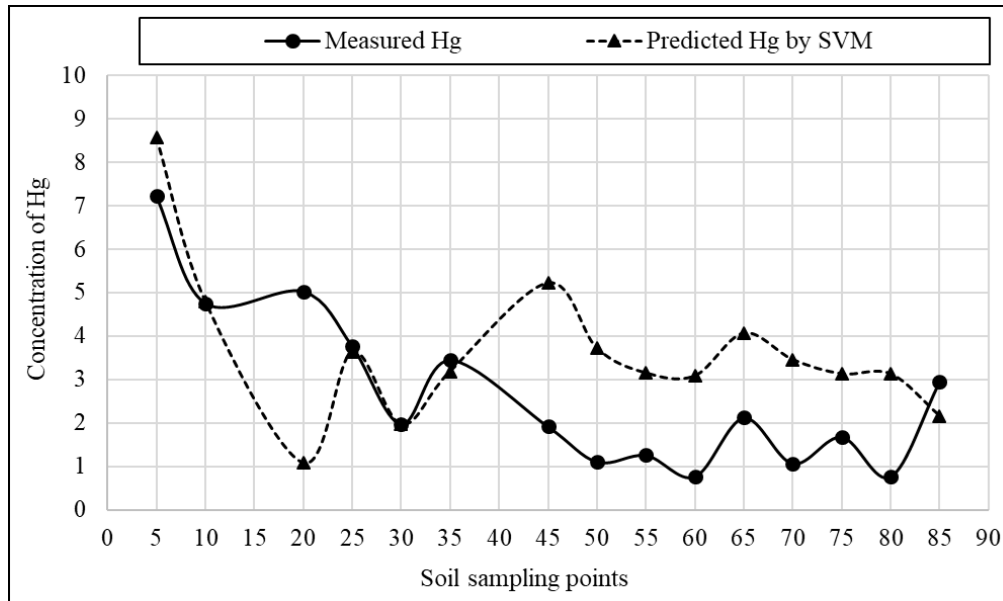
Figure 4. 29: Comparison of predicted and measured concentration of Hg
from SVM (testing).

Similarly, the perfect prediction was observed for the soil sampling points 5, 10, 25, 35 and 65 for Sb (Figure 4.30), whereas the soil sampling points 25, 30, 35, 45, 50 and 85 were denoted the perfect prediction for Sc in soils (Figure 4.31).
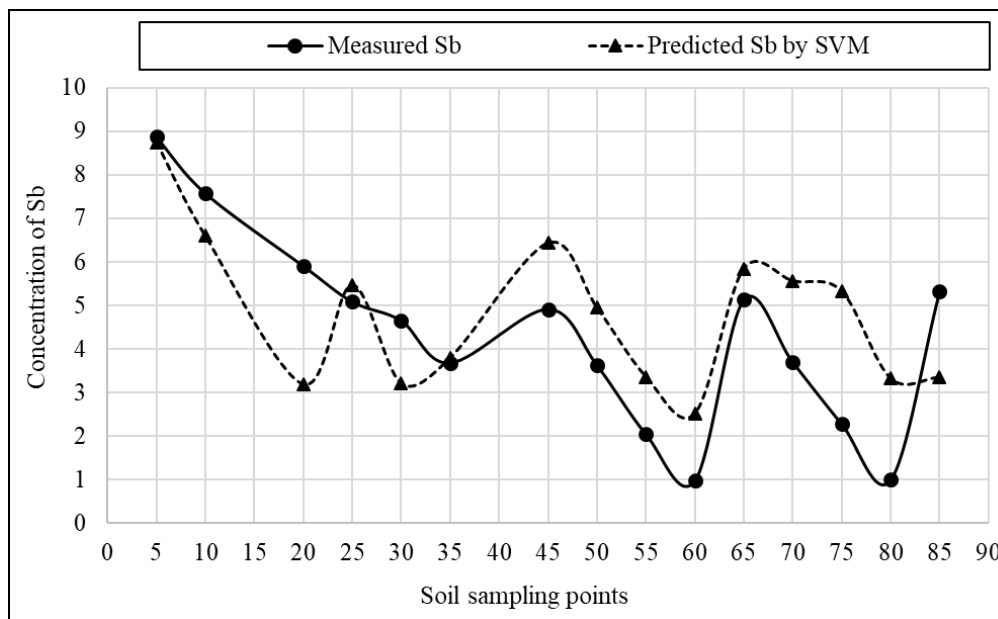
Figure 4.30: Comparison of predicted and measured concentration of Sb from
SVM (testing).
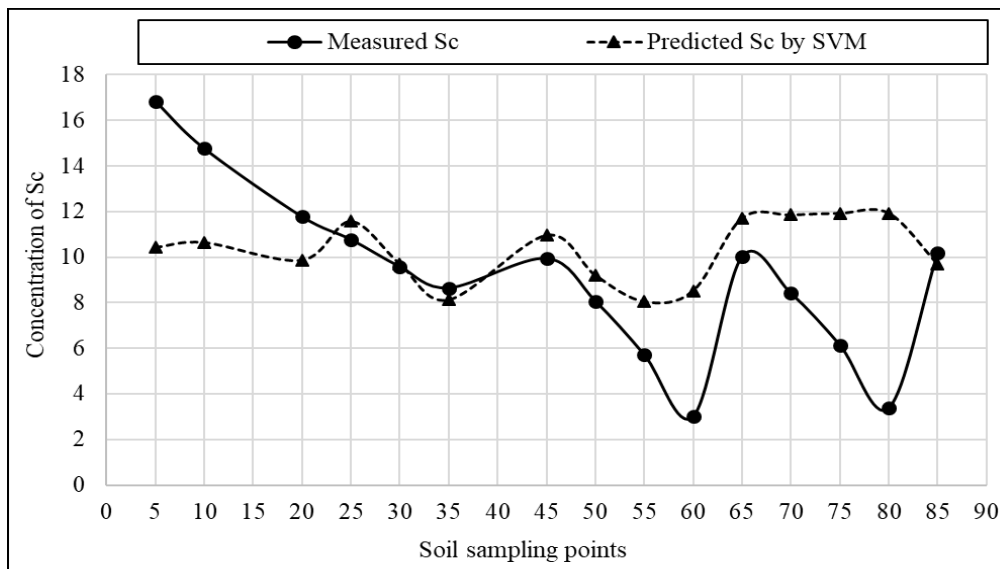
Figure 4.31: Comparison of predicted and measured concentration of Sc from SVM (testing).

Besides, the values of percentage recovery were found 197.82, 140.01 and 137.76 for Hg, Sb and Sc, respectively (Table 4.11). Among which for Sc, it was very close to fit level 100 than that of other studied heavy metals in soils. The level of percent recovery for Hg, Sb and Sc were expressed in Figure 4.32, 4.33 and 4.34, respectively.
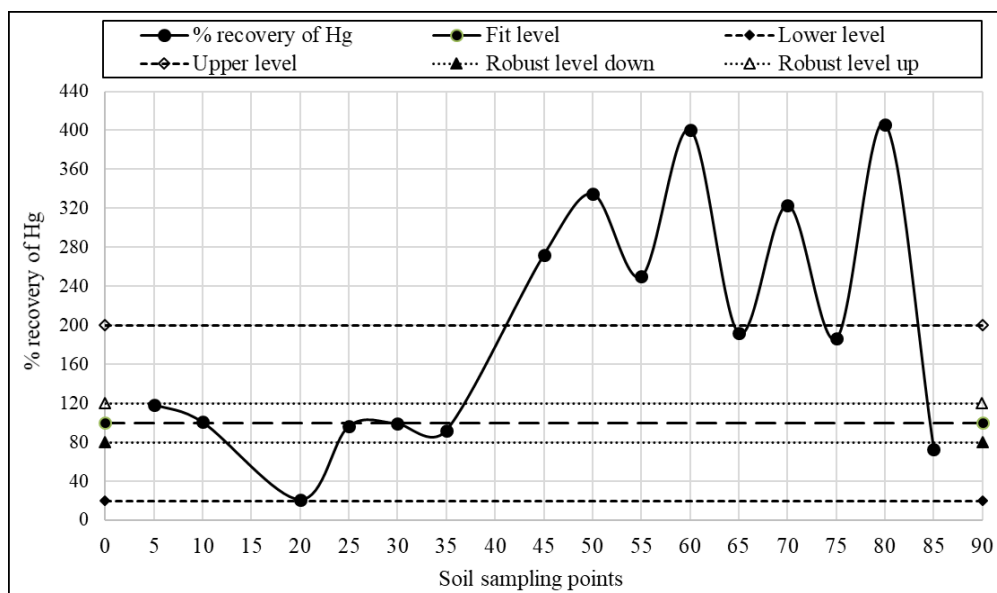
Figure 4.32: Variation of recovery level for Hg in SVM.

The percentage recovery of all soil sampling points were in the ranges of 20-200% except 45, 50, 55, 60, 70 and 80 in case of Hg (Figure 4.32). Based on this result, it was recommended that the prediction of Hg in soils was in acceptable range. Among the accepted soil sampling points; 5, 10, 25, 30 and 35 were in robust level ranges 80-120% (Figure 4.32).
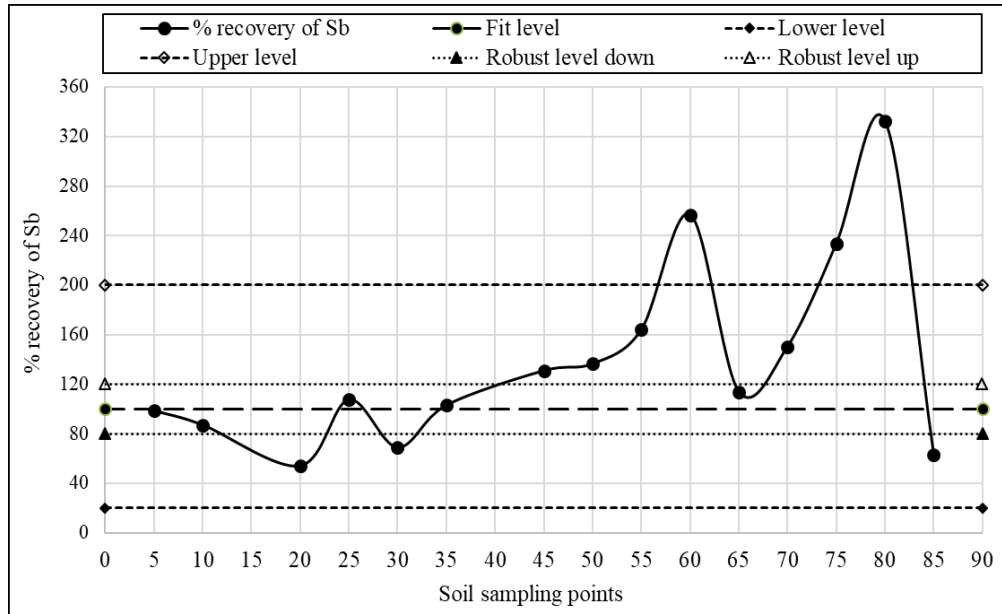
Figure 4.33: Variation of recovery level for Sb in ANFIS.
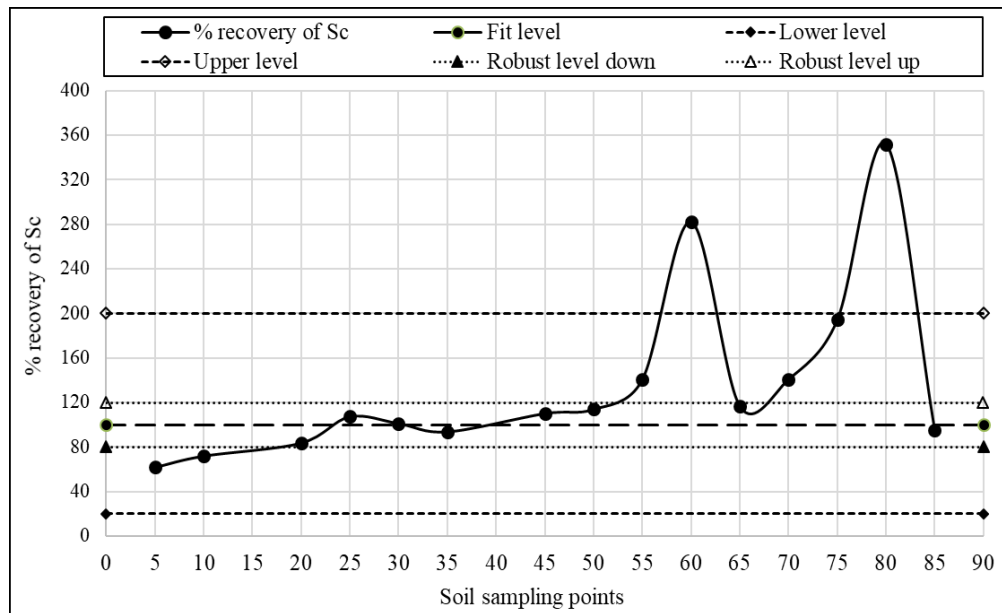
Figure 4.34: Variation of recovery level for Sc in ANFIS.

114

Similarly, the percentage recovery for Sb of soil sampling points 60, 75 and 80 were out of acceptable range (Figure 4.33) as well as the percentage recovery for Sc of soil sampling points 60 and 80 was out of acceptable range (Figure 4.34). Without the aforementioned sampling points of Sb and Sc, all values of percentage recovery were acceptable, as well as most of them were recommended as the robustness of prediction for Sb and Sc. In addition, GRI values were 2.30, 1.68 and 1.60 for Hg, Sb and Sc, respectively. All of this values were greater than 1. So, the prediction model for Hg, Sb and Sc were reliable. Moreover, relationships of predicted and measured concentrations of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils were also reported in Figure C-37 to Figure C-48, respectively, in the Appendix-C. In addition, the graphical representations of percentage recovery of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils were also reported in Figure C-49 to Figure C-60, respectively, in the Appendix-C. In contrast, the results of percentage recovery for Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba in soils were also provided in Table D-1 to Table D-12, respectively, in Appendix-D.

4.4.2.3.3 Summary of Results in SVM Model

In SVM analysis, various models with different kernel functions like SVM-L, SVM-Q, SVM-C and SVM-RBF was formed to select best-fitted model of SVM. The cross-validation with different folds (5, 10, 15 and 20) was also performed to control overfitting of the data. The model with kernel function SVM-RBF shows the higher value of R than that of other kernel functions of SVM-L, SVM-Q and SVM-C in SVM analysis for fold numbers 15. The maximum R-value (0.76) with minimum RMSE (1.27) represents the correlation (ranges of $0.2 < |R| < 0.8$) between the input and output variables for Cu in prediction. The values of MAPE were found 113.45, 57.14 and 50.19 for Hg, Sb and Sc, respectively. Minimum error was found for Sc indicating the closeness of the predicted values with measured one of Sc in soils. The performance of Sb was better than Hg due to less error ($57.14 < 113.45$). Besides, the values of percentage recovery were found 197.82, 140.01 and 137.76 for Hg, Sb and Sc, respectively. Among which for Sc, it was very close to fit level 100 than that of other studied heavy metals in soils. The percentage recovery of all soil sampling points were in the ranges of 20-200% except 45, 50, 55, 60, 70 and 80 in case of Hg. Based on this result, it was recommended that the prediction of Hg in soils was in acceptable range. Among the accepted soil sampling points;

5, 10, 25, 30 and 35 were in robust level ranges 80-120%. In addition, GRI values were 2.30, 1.68 and 1.60 (>1) for Hg, Sb and Sc, respectively representing the reliability of prediction model for SVM. The outputs of SVM-RBF model were represented by contour map, surface viewer and residuals as the results of predicted concentrations for heavy metals in soils. The intensity of predicted Pb was represented by the colorbar whereas yellow color shows the higher intensity of the concentration of Pb. In SVM analysis, the predicted concentration of Pb varies from 13.34 to 49.75 mg/kg in soils whereas measured value varies from 12.11 to 90.55 mg/kg.

4.4.3 Artificial Neuron Network

In this study, the model validation with changing of neuron number, various training and transfer functions, performance of the model prediction with prediction parameters and ANN results were deliberated successively. The results of ANN for the heavy metals of Pb, Hg, Sb and Sc were discussed in details in this section. Moreover, the outputs for the other heavy metals of Cu, Ni, Zn, Co, Cd, Hg, Mn, Cr, Sb, Sr, V and Ba in soils for ANN were reported in Appendix-E (Figures) and Appendix-F (Tables), respectively.

4.4.3.1 Validation of Models in ANN

In this study, different models of ANN were performed by changing the number of neurons as well as various training and transfer functions of ANN through MATLAB and hence discussed in following articles.

4.4.3.1.1 Selection of Neuron Number

In ANN analysis, four neuron structures were formed with different neuron numbers of 5, 10, 15 and 20 successively. In this analysis, the selection of neuron structure was performed based on the best values of R and RMSE. Table 4.12 illustrates the performance of four neuron structures for all heavy metals in soils with R and RMSE. R-values were found greater than 0.8 for most of the heavy metals for the neuron structure 2-10-1 indicating the strong correlation between the variables.

Table 4.12: Model validation performance for different neuron number

SL. No.	Heavy metals	2--5--1		2--10--1		2--15--1		2--20--1	
		R-Value	RMSE	R-Value	RMSE	R-Value	RMSE	R-Value	RMSE
1	Pb	0.75	9.52	0.80	8.47	0.61	11.35	0.76	9.30
2	Cu	0.79	1.98	0.80	2.36	0.79	2.03	0.71	3.65
3	Ni	0.67	1.29	0.72	1.19	0.73	1.20	0.78	1.15
4	Zn	0.74	6.62	0.73	7.37	0.77	6.56	0.59	8.25
5	Co	0.80	1.51	0.76	1.65	0.76	1.72	0.76	1.78
6	Cd	0.36	1.45	0.80	0.94	0.69	1.14	0.74	1.05
7	As	0.83	1.12	0.88	1.01	0.54	1.65	0.84	1.09
8	Sc	0.71	2.45	0.82	2.00	0.62	2.79	0.82	2.02
9	Hg	0.74	1.52	0.85	1.18	0.62	1.89	0.80	1.34
10	Mn	0.61	5.20	0.68	4.79	0.55	5.46	0.57	5.42
11	Cr	0.62	1.84	0.67	1.73	0.73	1.57	0.56	2.23
12	Ti	0.65	330.02	0.81	243.33	0.75	314.14	0.81	265.90
13	Sb	0.62	2.05	0.75	1.52	0.61	3.11	0.82	1.33
14	Sr	0.60	6.83	0.83	5.06	0.79	5.34	0.81	5.10
15	V	0.57	13.27	0.73	10.88	0.75	10.60	0.73	11.28
16	Ba	0.80	14.04	0.82	13.55	0.57	33.66	0.67	18.46

The variation of R with different neuron structures represents in Figure 4.35 for all studied heavy metals in soils. In this figure, the R value was found higher for the neuron structure 2-10-1 for most of the heavy metals than that for the other neuron structures (like 2-5-1, 2-15-1 and 2-20-1) as well as As shows the maximum value of R (0.88).
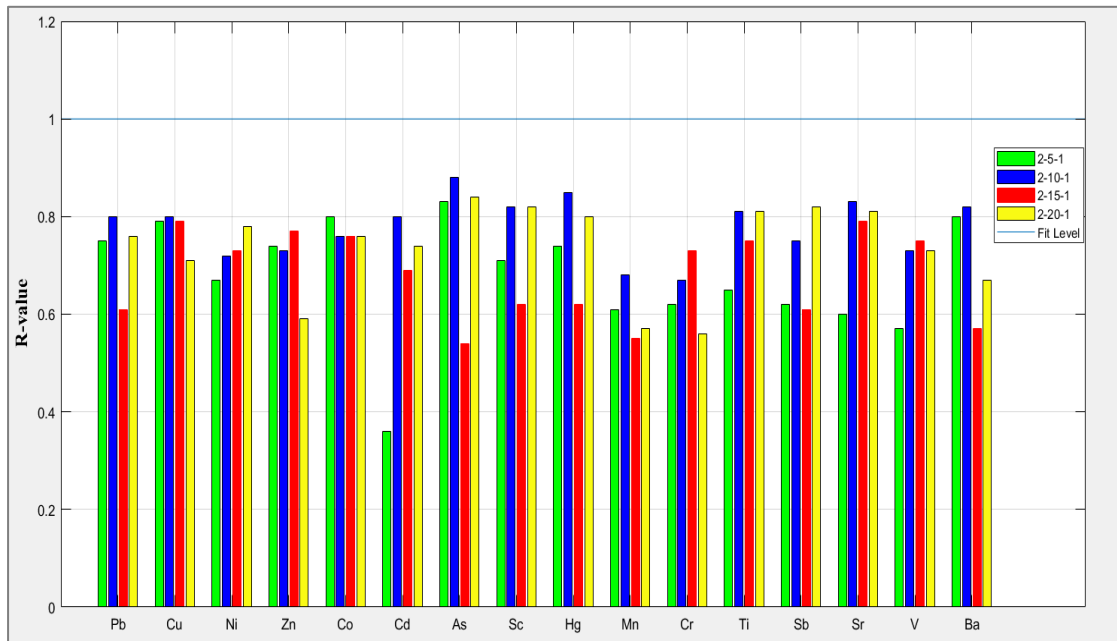
Figure 4.35: Variation of R with different neuron structures for all heavy metals.
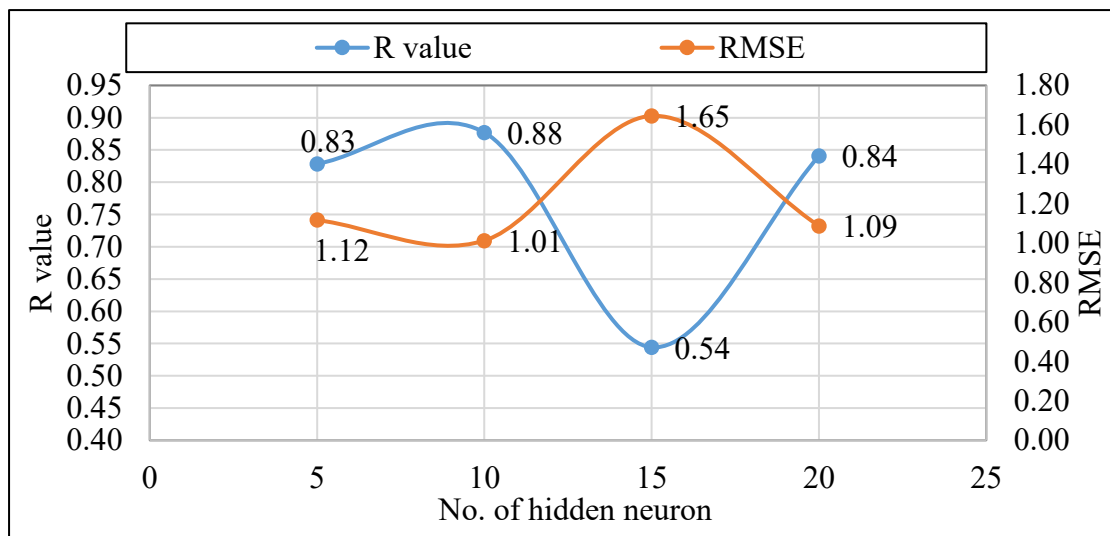
Figure 4.36: Comparison of R and RMSE of As for different number of hidden neuron.

Figure 4.36 also clearly expresses the variation of R and RMSE of As with different number of neurons. In this figure, for neuron number 10, As shows the maximum value of R (0.88) with minimum RMSE (1.01) whereas, neuron number 15 shows the minimum value of R (0.54) with maximum RMSE (1.65). Therefore, neuron structure 2-10-1 was selected for modelling of heavy metal analysis in soils of selected waste disposal site in ANN model.

4.4.3.1.2 Selection of Training and Transfer Functions

In this study, for ANN analysis, nine models were formed by interchanging different training functions like levenberg-marquardt (TRAINLM), one-step secant (TRAINOSS) and scaled conjugate gradient (TRAINSCG) as well as transfer functions like tangent sigmoid transfer function (TANSIG), linear transfer function (PURELIN) and log-sigmoid transfer function (LOGSIG). The models (combined with training and transfer functions) with neuron structure 2-10-1 were compared in terms of the best R and RMSE to assess the performance of each model. Table 4.13 illustrates the performance of different training and transfer functions for As with R and RMSE.

Table 4.13: Model validation performance for different training and transfer functions

Training function	Transfer function	Designation	Train		Neuron structure	No. of iterations
			R	RMSE		
TRAINLM	TANSIG	LT	0.88	1.01	2--10--1	36
TRAINLM	LOGSIG	LL	0.86	1.03	2--10--1	17
TRAINLM	PURELIN	LP	0.31	1.86	2--10--1	6
TRAINOSS	TANSIG	OT	0.67	1.49	2--10--1	12
TRAINOSS	LOGSIG	OL	0.72	1.38	2--10--1	17
TRAINOSS	PURELIN	OP	0.29	1.88	2--10--1	9
TRAINSCG	TANSIG	ST	0.76	1.33	2--10--1	14
TRAINSCG	LOGSIG	SL	0.57	1.85	2--10--1	12
TRAINSCG	PURELIN	SP	0.31	1.87	2--10--1	21

R-values of models LT and LL were greater than 0.8, which indicated the strong correlation between the variables. Excepting the models LT and LL, R-values of all other models (LP, OT, OL, OP, ST, SL and SP) were within the ranges of $0.2 < |R| < 0.8$, indicating the correlation between variables. Figure 4.37 illustrates the performance of training and transfer function for

As. In this figure, model LT shows the higher value of R (0.88) than that of other models with minimum RMSE (1.01). On the contrary, model SP shows the lower value of R (0.31) than the other models with maximum RMSE (1.87).

	LT	LL	LP	OT	OL	OP	ST	SL	SP
R	0.88	0.86	0.31	0.67	0.72	0.29	0.76	0.57	0.31
RMSE	1.01	1.03	1.86	1.49	1.38	1.88	1.33	1.85	1.87
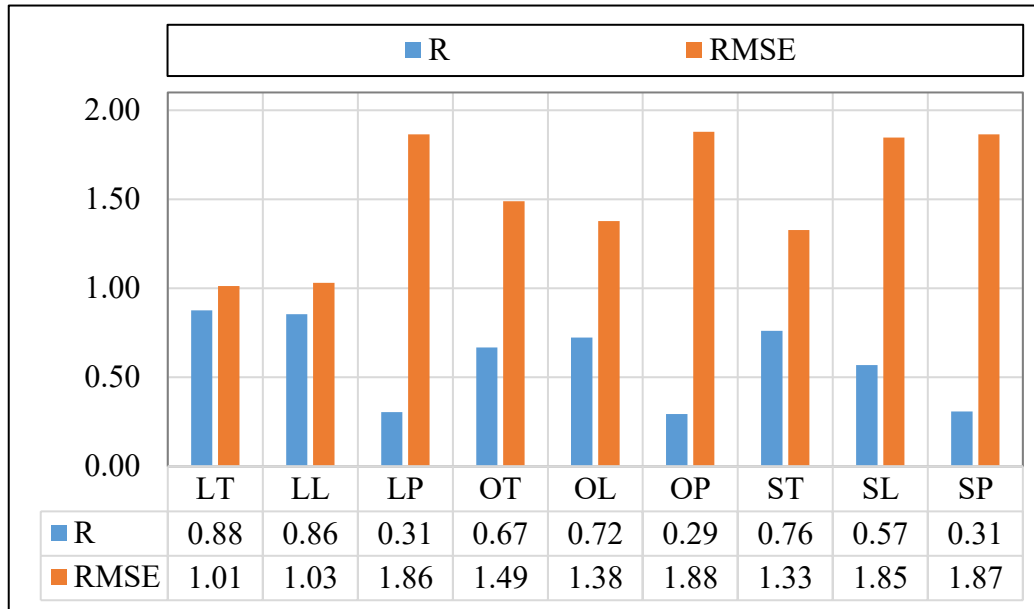
Figure 4.37: Performance of training and transfer function for As.

From the aforementioned discussion, the model LT with neuron structure 2-10-1 was selected for the prediction of heavy metal concentrations in soils of waste disposal site.

4.4.3.2 Assessment of Model Performance in ANN

The model LT with training function levenberg-marquardt (TRAINLM), transfer function tangent sigmoid (TANSIG) and neuron structure 2-10-1 was selected based on the acceptable limits of R and RMSE for the prediction of heavy metal concentrations. Table 4.14 provides the values of R and RMSE for training and testing for all studied heavy metals in soils.

Table 4.14: Variation of R and RMSE of all heavy metals for ANN

Heavy metals	ANN (training)		ANN (testing)	
	R value	RMSE	R value	RMSE
Pb	0.80	8.47	0.62	10.47
Cu	0.80	2.36	0.78	2.21
Ni	0.72	1.19	0.48	1.59
Zn	0.73	7.37	0.02	13.66
Co	0.76	1.65	0.24	3.35
Cd	0.80	0.94	0.26	2.22
As	0.88	1.01	0.80	1.31
Sc	0.82	2.00	0.52	3.45
Hg	0.85	1.18	0.37	2.29
Mn	0.68	4.79	0.42	6.44
Cr	0.67	1.73	0.10	2.97
Ti	0.81	243.33	0.34	531.52
Sb	0.75	1.52	0.33	2.59
Sr	0.83	5.06	0.53	8.36
V	0.73	10.88	0.56	13.78
Ba	0.82	13.55	0.71	16.03

In training, the values were found in the ranges of $|R| \geq 0.8$ for Pb, Cu, Cd, As, Sc, Hg, Ti, Sr and Ba indicating strong correlation between the inputs and outputs. In addition, R-value for Ni, Zn, Co, Mn, Cr, Sb and V was found in the ranges of $0.2 < |R| < 0.8$ which specified the existing of correlation between the inputs and outputs. Among all the heavy metals, the maximum R-value was found 0.88 for As with the minimum RMSE 1.01 indicating the best correlation in prediction for As. On the other hand, the minimum R-value was found 0.67 for Cr with the RMSE value 1.73 indicating the worst correlation in prediction for Cr than that of other studied heavy metals. For Ti, it was presented that the value of RMSE was maximum (243.33) due to the higher intensity of concentrations than all other heavy metals. The relationships of predicted and measured concentrations of Pb, Hg, Sb and Sc in soils in training is shown in Figure 4.38a, 4.38b, 4.38c and 4.38d, respectively. Figures reveal all the predicted concentrations were closed to the fitted line and Hg (Figure 4.38b) shows the best performance with better R value than others. Moreover, the ANN regression for training of other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure E-1 to Figure E-12 in Appendix-E.
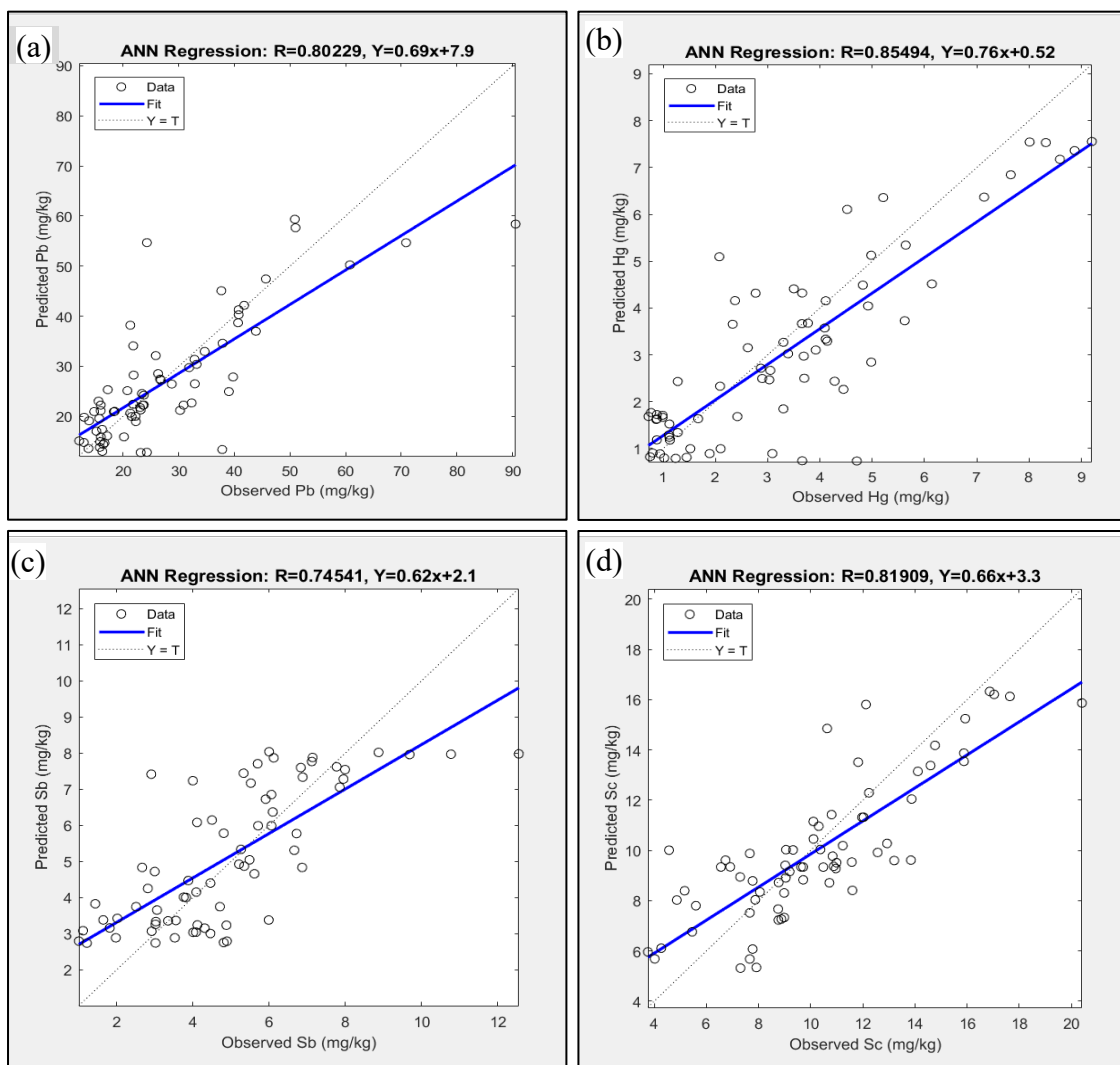
Figure 4.38: Regression analysis for ANN (training) of (a) Pb, (b) As, (c) Sb and (d) Sc.

In testing, the value of |R| for only As was found in the range |R| ≥ 0.8, which were indicated the existing of strong correlation. The R-value between the ranges of 0.2 < |R| < 0.8 was found for Pb, Cu, Ni, Zn, Co, Cd, As, Sc, Hg, Mn, Cr, Ti, Sb, Sr, V and Ba which specified the existing of correlation between the inputs and outputs. Among all the heavy metals in testing, the maximum R-value (0.80) with minimum RMSE (1.31) for As as well as R=0.02 (minimum) and RMSE=13.66 (maximum) for Zn. The relationships of predicted and measured concentrations of Pb, Hg, Sb and Sc in soils were represented in Figure 4.39a, 4.39b, 4.39c and 4.39d, respectively. In figure predicted concentration of Pb, Hg, Sb and Sc were plotted against the measured concentration respectively. The plotted data for Pb was close to the fitted line (Figure

4.39a) whereas for Sb data was scattered from the fitted line (Figure 4.39c). Moreover, ANN regression for testing data of other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure E-13 to Figure E-24 in Appendix-E.
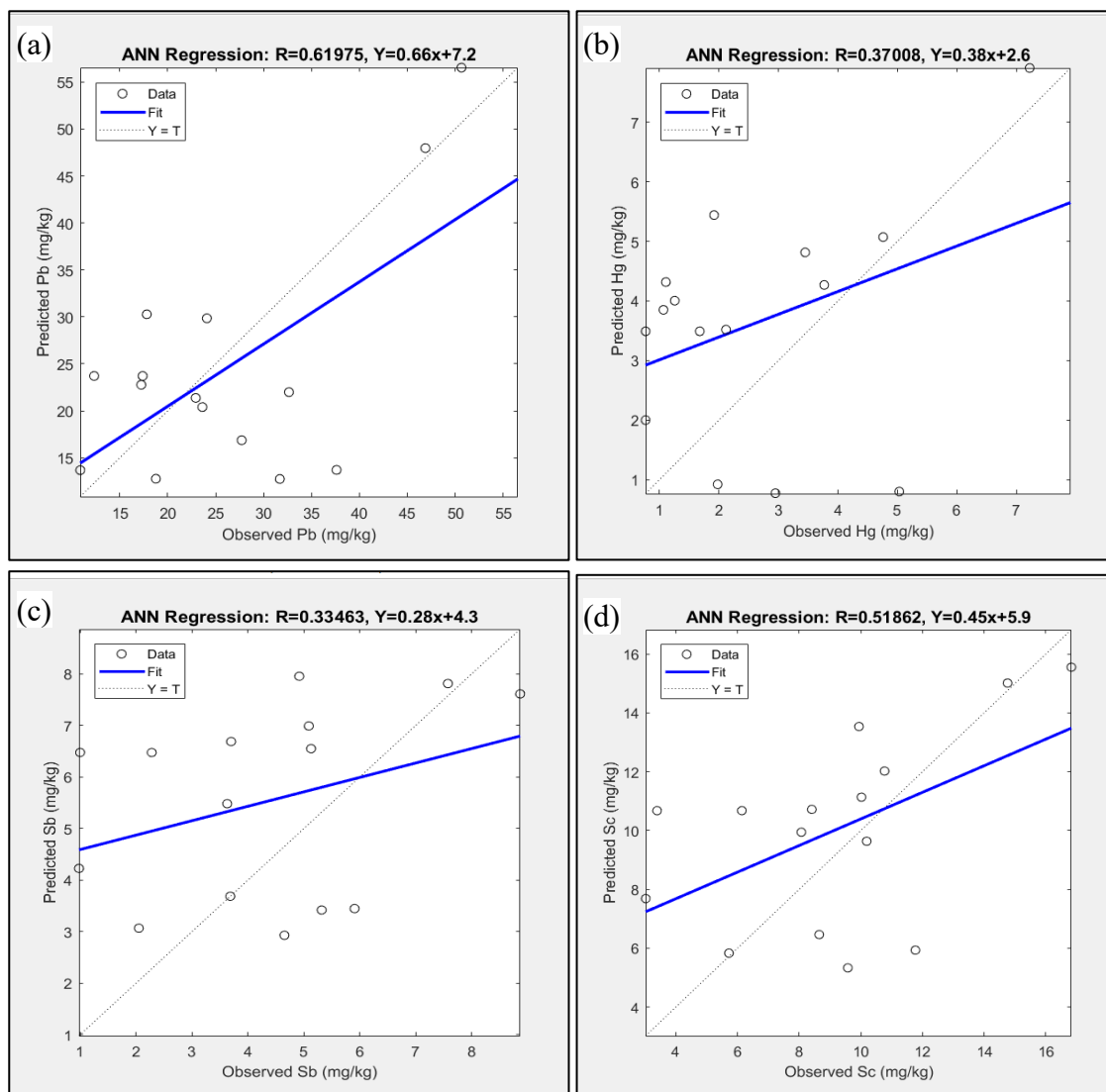
Figure 4.39: Regression analysis for ANN (testing) of (a) Pb, (b) As, (c) Sb and (d) Sc.

4.4.3.2.1 Assessment of Pb of ANN

In this study, ANN model was directed to predict heavy metal concentrations in soils of 15 unknown sampling points of the selected waste disposal site. The aim of this section was to evaluate the closeness of the predicted and measured concentration of Pb in soils for ANN

model. Table 4.15 represents the predicted and measured concentration of Pb. The variation of measured and predicted concentration of Pb was clearly exposed in Figure 4.40.

Table 4.15: Recovery level of Pb in ANN (testing)

| Soils sampling points | Recovery level of Pb for testing in ANN | | |
	Measured concentration	Predicted concentration	% recovery
5	50.66	56.53	111.59
10	46.89	47.96	102.28
20	37.62	13.71	36.44
25	32.66	22.00	67.35
30	27.72	16.88	60.88
35	31.705	12.75	40.21
45	24.09	29.85	123.93
50	17.82	30.28	169.90
55	18.77	12.77	68.04
60	10.88	13.69	125.84
65	23.62	20.41	86.40
70	17.24	22.77	132.10
75	17.39	23.71	136.36
80	12.33	23.71	192.32
85	22.93	21.38	93.24
% mean recovery			103.13
RMSE			10.47
MAPE			36.12
GRI			1.59

In this figure, only few soil sampling points (10, 60, 65 and 85) shows the minor variation as well as all others points shows the high variation. Considering soil sampling points 10 and 85, measured values were found as 46.89 and 22.93 mg/kg for Pb whereas predicted values were 47.96 and 23.86 mg/kg, respectively.
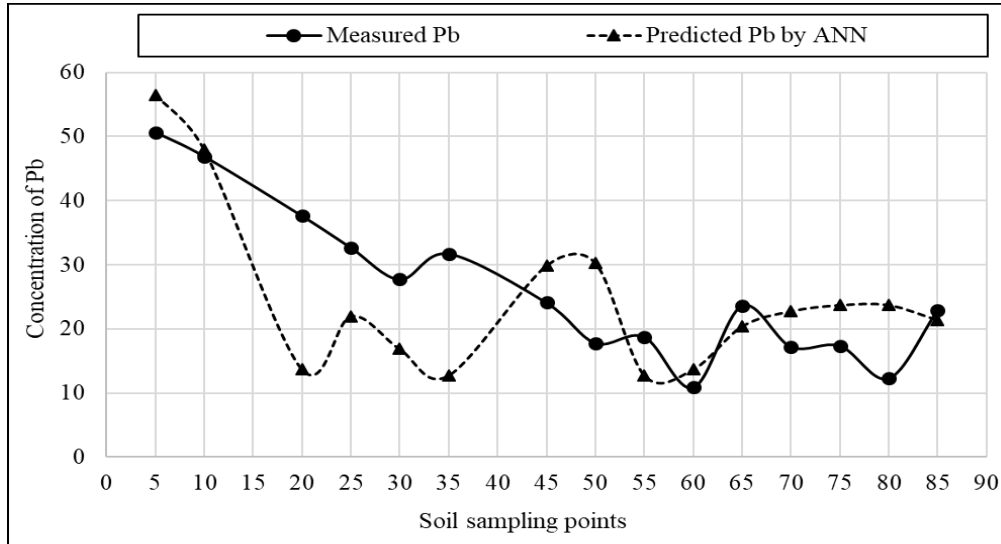
Figure 4.40: Comparison of predicted and measured concentration of Pb from ANN (testing).

Table 4.15 also demonstrates the results of RMSE, MAPE, GRI and percentage recovery. The percentage recovery means what percentage of measured value is recovered by the predicted value. The variation of recovery level for Pb in ANN model (testing) is shown in Figure 4.41. In this figure, all sampling points are in the acceptable ranges 20-200%. Besides, some sampling points like 5, 10, 45, 60, 65 and 85 were found very near and between the ranges of 80-120% indicating the robustness of model for Pb in ANN analysis.
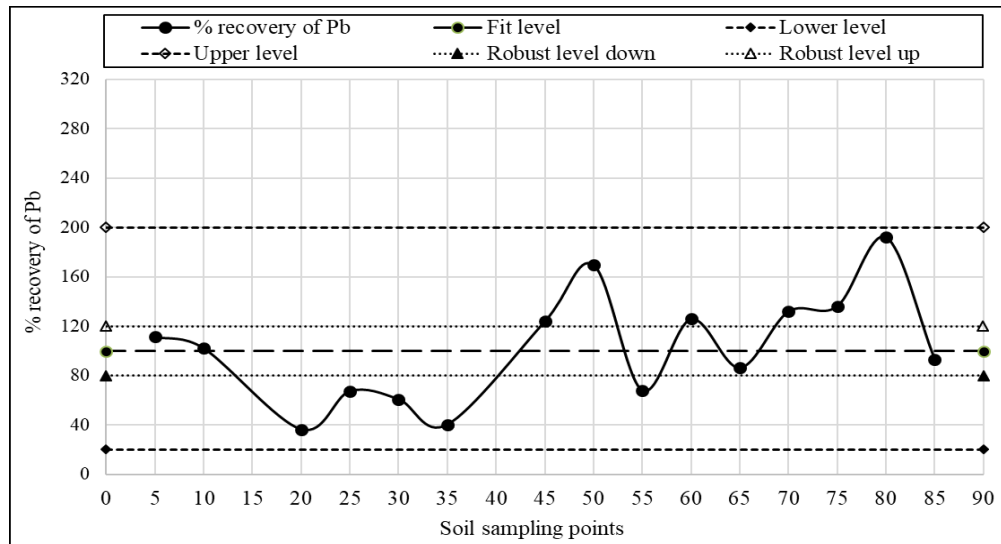
Figure 4.41: Variation of recovery level for Pb in ANN.

125

Again, the MAPE value for Pb was found 36.12% signifying quite good results of MAPE. Furthermore, the value of GRI was found 1.59 (greater than 1) for Pb representing the reliability of model. Therefore, the model of ANN was considered as robust and reliable for predicting the Pb concentration in soils.

4.4.3.2.2 Assessment of Hg, Sb and Sc in ANN

The results of measured, predicted concentration as well as percentage recovery of Hg, Sb and Sc in soils depicts in Table 4.16. The values of MAPE were found 127.80, 100.22 and 45.92 for Hg, Sb and Sc, respectively. Minimum error was found for Sc indicating the closeness of the predicted values with measured one of Sc in soils. The performance of Sb was better than Hg due to less error (100.22 < 127.80).

Table 4.16: Recovery level for Hg, Sb and Sc in ANN (testing)

Soils sampling points	Hg			Sb			Sc		
	Measured	Predicted	(%) recovery	Measured	Predicted	(%) recovery	Measured	Predicted	(%) recovery
5	7.22	7.91	109.51	8.87	7.62	85.85	16.83	15.55	92.41
10	4.76	5.07	106.53	7.58	7.82	103.13	14.76	15.01	101.73
20	5.03	0.80	15.97	5.91	3.45	58.35	11.77	5.93	50.41
25	3.77	4.27	113.21	5.09	6.99	137.37	10.77	12.02	111.63
30	1.98	0.92	46.71	4.65	2.93	62.94	9.58	5.33	55.65
35	3.45	4.81	139.53	3.68	3.69	100.07	8.65	6.46	74.70
45	1.92	5.44	283.33	4.92	7.96	161.73	9.94	13.54	136.18
50	1.11	4.32	388.91	3.63	5.48	151.08	8.07	9.94	123.21
55	1.26	4.00	317.80	2.05	3.07	149.65	5.72	5.83	101.93
60	0.77	2.00	260.08	0.98	4.23	431.35	3.02	7.69	254.52
65	2.12	3.52	166.01	5.13	6.55	127.72	10.02	11.13	111.12
70	1.07	3.85	359.90	3.70	6.69	180.80	8.41	10.72	127.45
75	1.68	3.49	207.81	2.28	6.48	284.11	6.14	10.68	173.86
80	0.77	3.49	453.41	1.00	6.48	647.77	3.39	10.68	314.90
85	2.95	0.78	26.30	5.32	3.42	64.29	10.19	9.64	94.57
% mean recovery			199.67	% mean recovery		183.08	% mean recovery		128.29
RMSE			2.29	RMSE		2.59	RMSE		3.45
MAPE			127.80	MAPE		100.22	MAPE		45.92
GRI			2.61	GRI		1.96	GRI		1.60
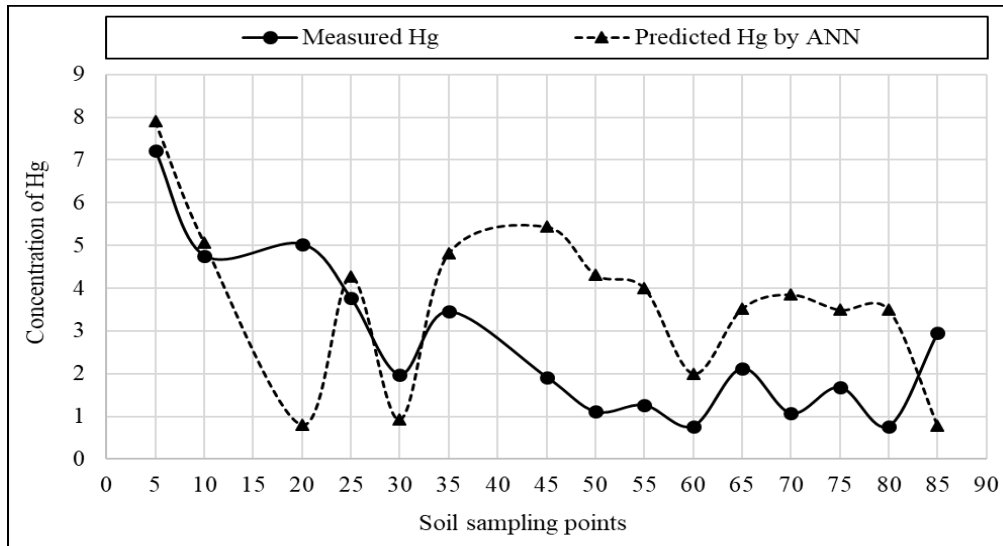
Figure 4.42: Comparison of predicted and measured concentration of Hg from ANN (testing).

The variation of predicted and measured concentration of Hg, Sb and Sc were illustrated in Figures 4.42, 4.43 and 4.44, respectively. The perfect prediction was observed in case of Hg in soils for the soil sampling points of 5, 10, 25 and 30 (Figure 4.42). In contrast, 20, 45, 50, 55, 60, 65, 70, 75 and 80 were exposed comparatively worse prediction than that of others soil sampling points in ANN analysis.
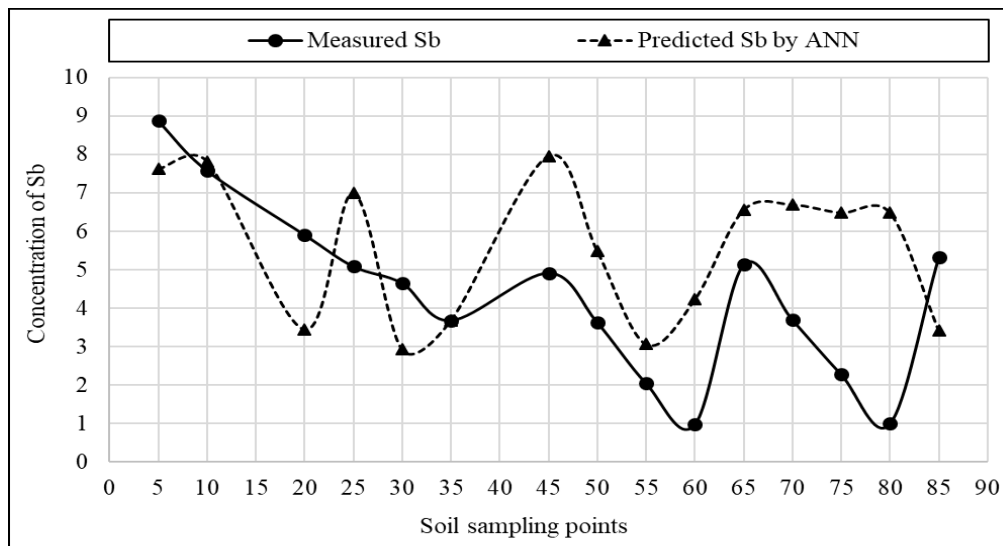
Figure 4.43: Comparison of predicted and measured concentration of Sb from ANN (testing).

127
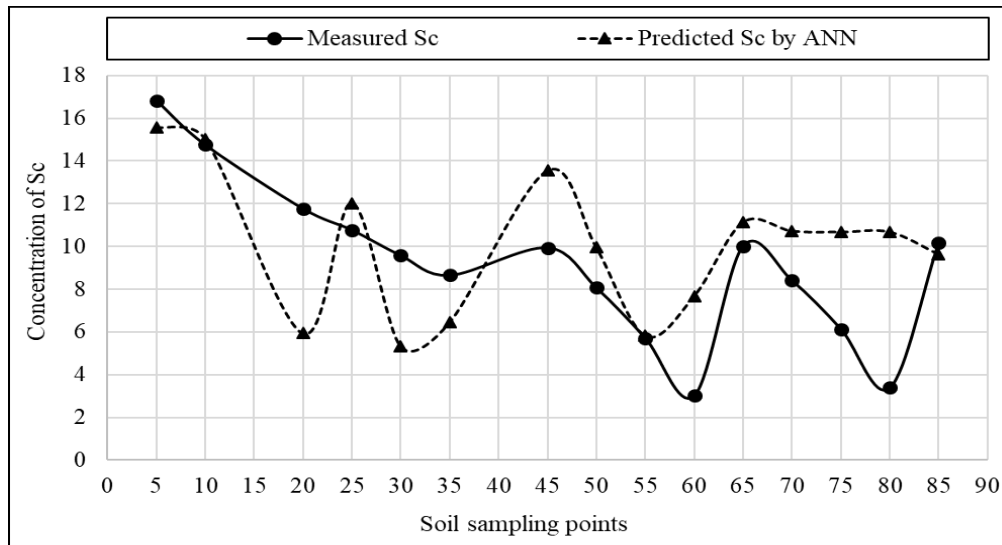
Figure 4.44: Comparison of predicted and measured concentration of Sc from ANN (testing).

Similarly, the perfect prediction was observed for the soil sampling points 5, 10, 35 and 55 for Sb (Figure 4.43), whereas the soil sampling points 5, 10, 25, 50, 55, 65 and 85 were denoted the perfect prediction for Sc in soils (Figure 4.44).
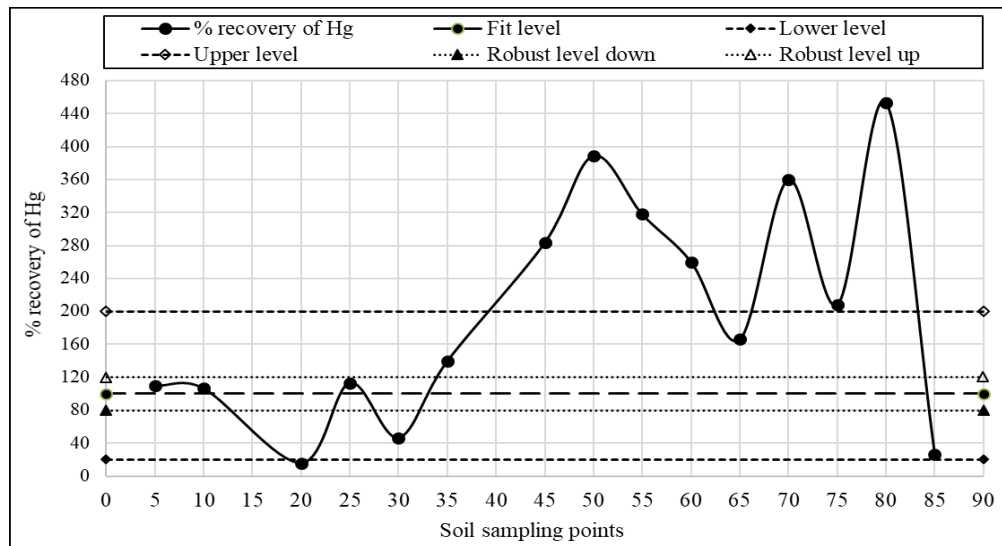
Figure 4.45: Variation of recovery level for Hg in ANN.

Besides, the values of percentage recovery were found 199.67, 183.08 and 128.29 for Hg, Sb and Sc, respectively (Table 4.16). Among which for Sc, it was very closed to fit level 100 than that of other studied heavy metals in soils. The level of recovery for Hg, Sb and Sc were

expressed in Figure 4.45, 4.46 and 4.47 respectively. The percentage recovery were found in the ranges of 20-200% for all soil sampling points except 45, 50, 55, 60, 70, 75 and 80 in case of Hg (Figure 4.45). Based on this result, it was recommended that the prediction of Hg in soils was in acceptable range. Among the accepted soil sampling points; most of them were in robust level ranges 80-120% (Figure 4.45).
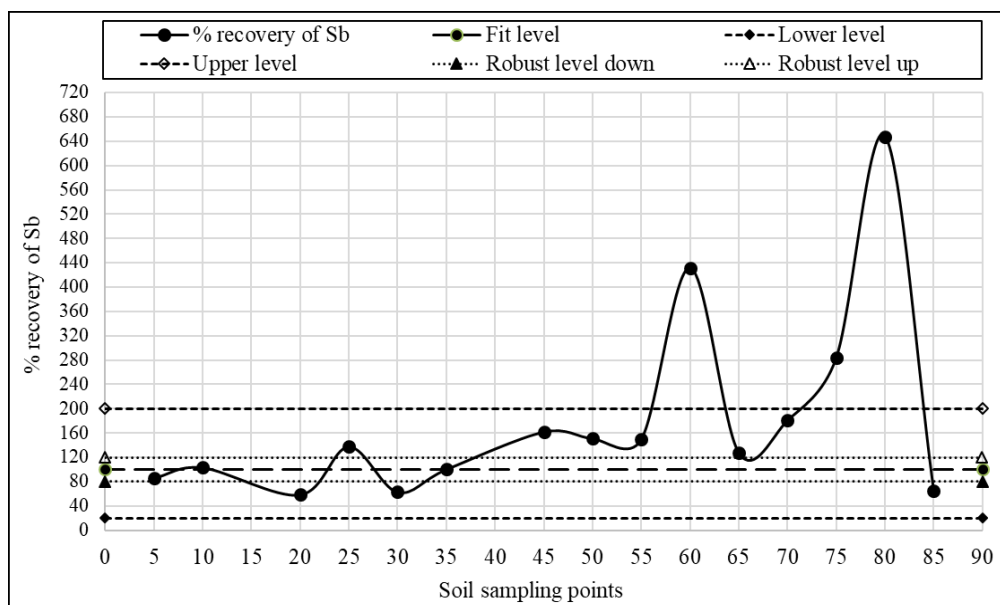
Figure 4.46: Variation of recovery level for Sb in ANN.
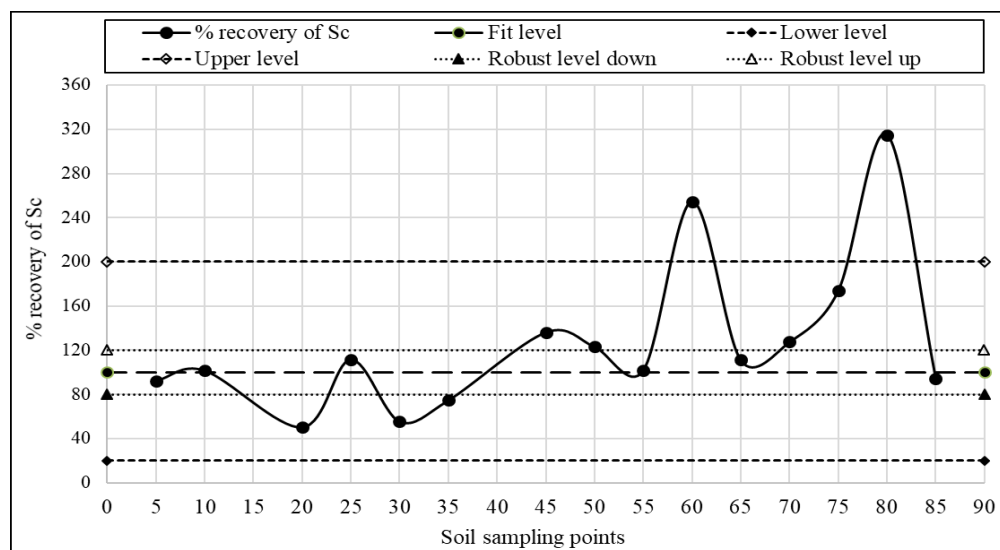
Figure 4.47: Variation of recovery level for Sc in ANN.

Similarly, the percentage recovery for Sb of soil sampling points 60, 75 and 80 were out of acceptable range (Figure 4.46) as well as the percentage recovery for Sc of soil sampling points 60 and 80 was out of acceptable range (Figure 4.47). Without the aforementioned sampling points of Sb and Sc, all values of percentage recovery were acceptable, as well as most of them were recommended as the robustness of prediction for Sb and Sc. In addition, GRI values were found 2.61, 1.96 and 1.60 (>1) for Hg, Sb and Sc, respectively. Therefore, the prediction model for Hg, Sb and Sc were reliable. Moreover, the comparison graph, recovery graph and recovery table in ANN model for other heavy metals in ANN of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, , Sr, V and Ba were reported in Figure E-25 to E-36 Figure E-37 to Figure E-48 as well as Table F-1 to Table F-12 respectively, in the Appendix-E and Appendix-F.

4.4.3.2.3 Summary of Results in ANN Model

In ANN analysis, different neuron structures (like 2-5-1, 2-10-1, 2-15-1 and 2-20-1), different training functions like TRAINLM, TRAINOSS and TRAINSCG as well as transfer functions like TANSIG, PURELIN and LOGSIG were implemented for the selection of best-fitted model. The model LT with training function TRAINLM, transfer function TANSIG and neuron structure 2-10-1 was selected based on the acceptable limits of R and RMSE for the prediction of heavy metal concentrations. The maximum R-value (0.88) for As with the minimum RMSE (1.01) represents the strong correlation (ranges of $|R| \geq 0.8$) in prediction. The values of MAPE were found 56.57, 28.89 and 19.74 for Hg, Sb and Sc in soils, respectively. The values of percentage recovery were found 143.50, 115.53 and 113.73 for Hg, Sb and Sc, respectively. Among which for Sc, it was very close to fit level 100 than others. the percentage recovery ranges from 20-200% for all soil sampling points except 50, 70 and 80 in case Hg. Based on this result, it was recommended that the prediction of Hg in soils was in acceptable range. Among the accepted soil sampling points; 5, 10, 20, 25, 30 and 85 were in robust level (80-120%). The percentage recovery were found in the acceptable range for all soil sampling points except 75 in case of Sb. In addition, GRI values were found 1.65, 1.37 and 1.26 (greater than 1) for Hg, Sb and Sc, respectively.

4.5 Comparisons of Artificial Intelligence Techniques

In this study, the best-fitted models of AI techniques (ANFIS, SVM and ANN) were selected for the prediction of heavy metals concentration in soils of unknown sampling points of the selected waste disposal site. The accuracy of the predicted results depend on the acceptable limits of prediction parameters like R-value, RMSE, MAPE, GRI and percentage recovery. In the literature, various researchers published the acceptable limits of these prediction parameters. In this study, to check the validity of predicted concentrations from AI techniques (ANFIS, SVM and ANN), the acceptable limits of these prediction parameters were considered and hence discussed briefly in the following articles.

4.5.1 Prediction Parameters

The values of the prediction parameters represents the higher accuracy of the selected model or not. The comparison of the results of R-value, RMSE, MAPE, GRI and mean percentage recovery of AI techniques were analysed and hence discussed in the following articles.

4.5.1.1 R Value

A research conducted by Smith (1986) suggested that the value of $|R| \geq 0.8$: a strong correlation exists, $0.2 < |R| < 0.8$: correlation exists and $|R| \leq 0.2$: a weak correlation exists. When the value of $|R|$ is greater than 0.9, then a very strong correlation exists /represents robustness between the variables. Table 4.17 represents the results of R for all studied heavy metals with various AI techniques in both training and testing. In addition, the results of R were orderly expressed in Figure 4.48 and Figure 4.49 for training and testing, respectively.

Table 4.17: Results of R of various AI techniques

Heavy metals	R value (training)			R value (testing)		
	ANFIS	SVM	ANN	ANFIS	SVM	ANN
Pb	0.99	0.66	0.80	0.55	0.43	0.62
Cu	0.98	0.71	0.80	0.86	0.73	0.78
Ni	0.81	0.65	0.72	0.71	0.60	0.48
Zn	0.998	0.63	0.73	0.63	0.37	0.02
Co	0.97	0.68	0.76	0.58	0.70	0.24
Cd	0.91	0.60	0.80	0.79	0.51	0.26
As	0.98	0.75	0.88	0.88	0.59	0.80
Sc	0.999	0.37	0.82	0.70	0.14	0.52
Hg	0.86	0.78	0.83	0.53	0.47	0.37
Mn	0.86	0.53	0.68	0.45	0.06	0.42
Cr	0.998	0.49	0.67	0.72	0.26	0.10
Ti	0.997	0.74	0.81	0.76	0.50	0.34
Sb	0.92	0.72	0.75	0.40	0.49	0.33
Sr	0.99	0.73	0.83	0.75	0.47	0.53
V	0.999	0.72	0.73	0.72	0.49	0.56
Ba	0.999	0.75	0.82	0.82	0.65	0.71

In Figure 4.48, ANFIS model shows the higher value of R (training) for all studied heavy metals than other models of AI techniques. The values of R for most of the heavy metals were found in the ranges of 0.81 to 0.999. According to Smith (1986), it indicated the robustness of ANFIS model. Moreover, ANN shows comparatively the better performance than that of SVM. Therefore, the performance of R (training) can be expressed as ANFIS > ANN > SVM (Figure 4.48).
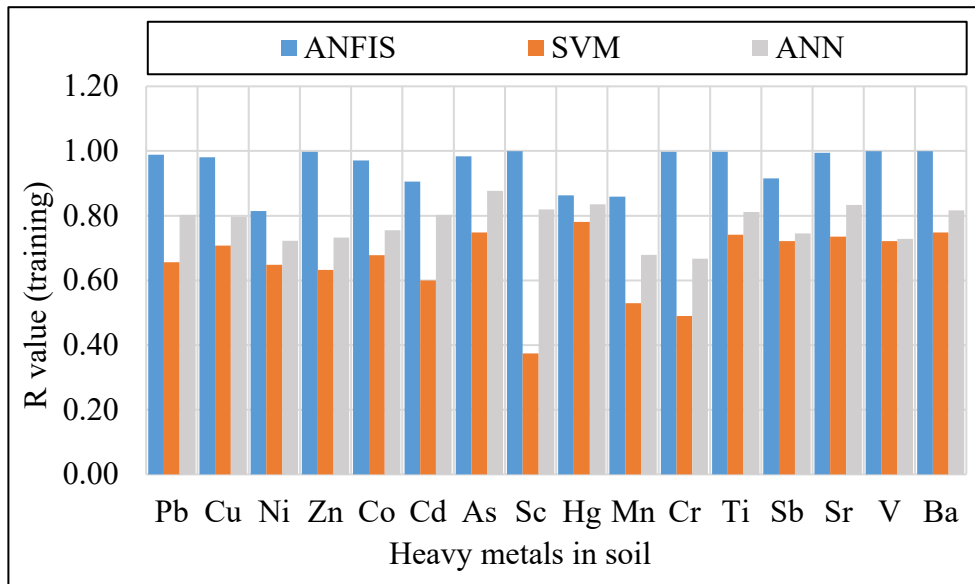
Figure 4.48: Variation of R values (training) with various AI techniques.

In addition, Figure 4.49 also shows the higher value of R (testing) for ANFIS model than that of other models of AI techniques for all studied heavy metals in soils. In ANN model, the R-value of eight heavy metals (Pb, Cu, As, Sc, Mn, Sr, V and Ba) were found comparatively better than that of SVM. In addition, the R-value of other eight heavy metals were found better in SVM than that of ANN. Therefore, the performance of R in testing can be expressed as ANFIS > SVM ≥ ANN.
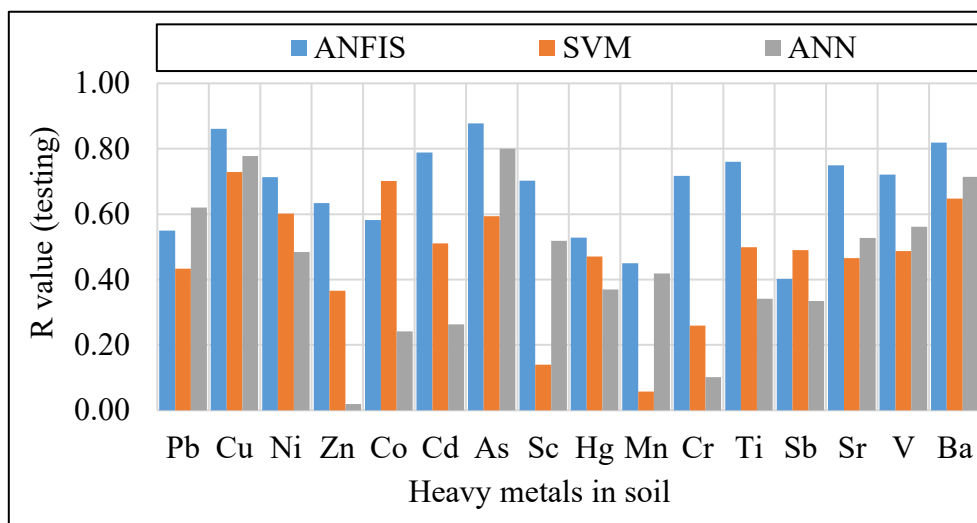
Figure 4.49: Variation of R-values (testing) with various AI techniques.

4.5.1.2 RMSE Value

A research conducted by Schweizer (2010) and stated that the lower value of RMSE indicates better fit of model and zero means no error. Table 4.18 illustrates the results of RMSE for all studied heavy metals with various AI techniques in both training and testing. The results of RMSE were clearly expressed for training and testing in Figure 4.50 and Figure 4.51, respectively, for the judgement of various AI techniques

Table 4.18: Results of RMSE of various AI techniques

Heavy metals	RMSE (training)			RMSE (testing)		
	ANFIS	SVM	ANN	ANFIS	SVM	ANN
Pb	1.38	10.64	8.47	15.25	10.73	10.47
Cu	0.63	2.30	2.36	2.45	2.09	2.21
Ni	1.00	1.31	1.19	1.17	1.53	1.59
Zn	0.67	7.59	7.37	9.18	11.01	13.66
Co	0.60	1.85	1.65	2.35	1.91	3.35
Cd	0.65	1.22	0.94	1.17	1.40	2.22
As	0.36	1.30	1.01	1.44	1.64	1.31
Sc	0.12	3.22	2.00	2.97	3.85	3.45
Hg	1.42	1.37	1.18	1.95	2.03	2.29
Mn	3.33	5.52	4.79	6.24	7.24	6.44
Cr	0.15	2.01	1.73	2.02	2.66	2.97
Ti	29.60	279.98	243.33	341.76	422.11	531.52
Sb	0.91	1.57	1.52	2.30	2.06	2.59
Sr	0.88	5.79	5.06	7.17	8.25	8.36
V	0.69	10.91	10.88	12.40	15.35	13.78
Ba	0.88	15.36	13.55	15.52	18.90	16.03

In Figure 4.50, ANFIS model shows the minimum value of RMSE in training for all studied heavy metals than that of other models of AI techniques. On the other hand, SVM shows the maximum RMSE indicating worse performance than ANN does. Therefore, the performance of RMSE in training can be expressed as ANFIS > ANN > SVM.
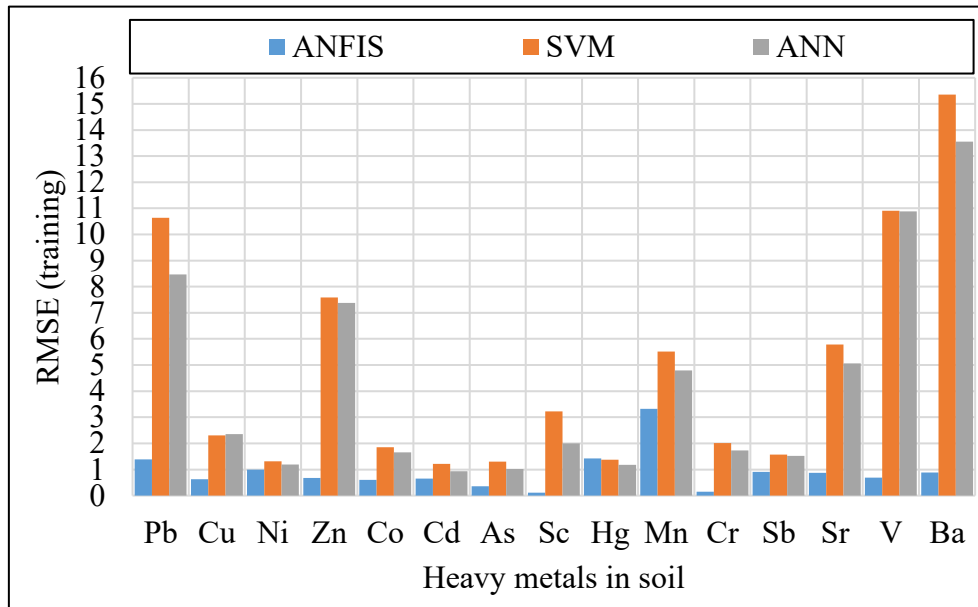
Figure 4.50: Variation of RMSE (training) with various AI techniques.

In Figure 4.51, RMSE value (testing) for most of the heavy metals in ANFIS model shows the less error than that of other models of AI techniques. The RMSE value of ten heavy metals (Pb, Ni, Zn, Co, Cd, Hg, Cr, Ti, Sb and Sr) in soils were found better in SVM than that of ANN.
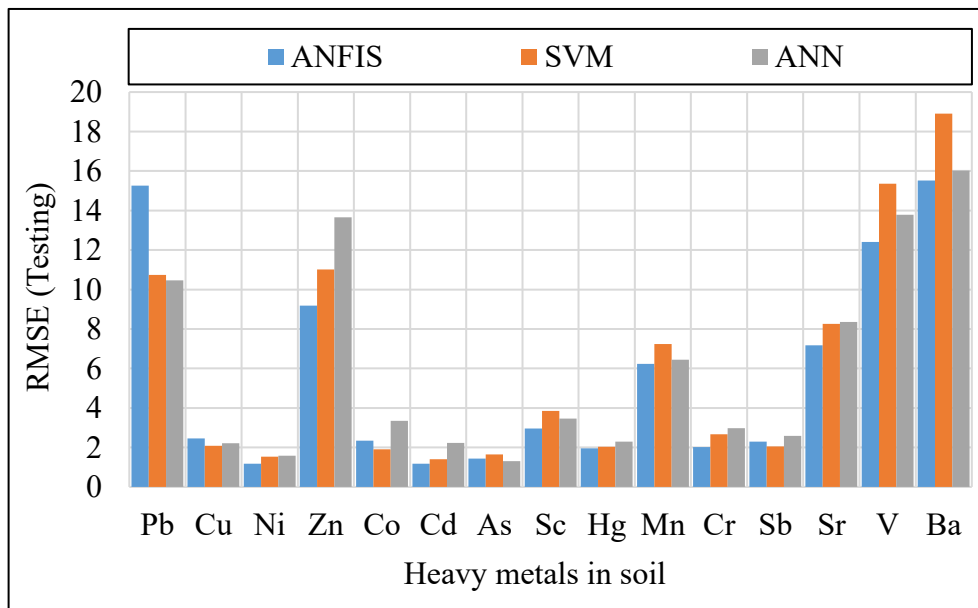
Figure 4.51: Variation of RMSE (testing) with various AI techniques.

In contrast, the RMSE value of other heavy metals (Cu, As, Sc, Mn, V and Ba) in soils were found better in ANN than that of SVM model. Therefore, the performance of RMSE in testing can be expressed as ANFIS > SVM > ANN.

4.5.1.3 MAPE Value

According to Rayer (2007), the MAPE value about 30–40% was acceptable and zero means no error. Table 4.19 illustrates the results of MAPE for all studied heavy metals with various AI techniques in both training and testing. The results of MAPE were clearly expressed in Figure 4.52 and Figure 4.53 for training and testing, respectively for the judgement of various AI techniques.

Table 4.19: Results of MAPE of various AI techniques

Heavy metals	MAPE (training)			MAPE (testing)		
	ANFIS	SVM	ANN	ANFIS	SVM	ANN
Pb	2.41	11.44	21.71	15.58	45.01	36.12
Cu	6.90	17.41	40.27	64.12	80.48	69.09
Ni	23.72	15.87	27.55	38.42	58.08	56.14
Zn	0.32	9.93	25.17	38.22	52.45	58.93
Co	6.46	26.65	21.26	50.51	47.16	76.42
Cd	11.49	12.17	18.83	45.21	60.11	89.96
As	9.93	17.03	21.06	59.90	99.32	55.15
Sc	0.39	27.13	17.47	19.74	50.19	45.92
Hg	20.03	18.51	47.02	56.57	113.45	127.80
Mn	34.43	44.61	42.02	130.45	174.75	123.46
Cr	0.60	19.12	38.17	70.13	150.75	146.56
Ti	1.50	10.30	20.41	45.40	68.75	76.83
Sb	10.76	12.92	25.73	28.89	57.14	100.22
Sr	1.13	8.40	23.38	33.55	47.96	42.68
V	0.31	12.81	30.44	53.38	86.98	63.64
Ba	0.50	11.51	21.67	33.94	49.88	38.42

In Figure 4.52, most of all heavy metals show MAPE value in the ranges of 30-40% in ANFIS model for training than other models of AI techniques. On the other hand, ANN shows the maximum MAPE, which indicating worse performance than SVM. Therefore, the performance of MAPE in training can be expressed as ANFIS > SVM > ANN.
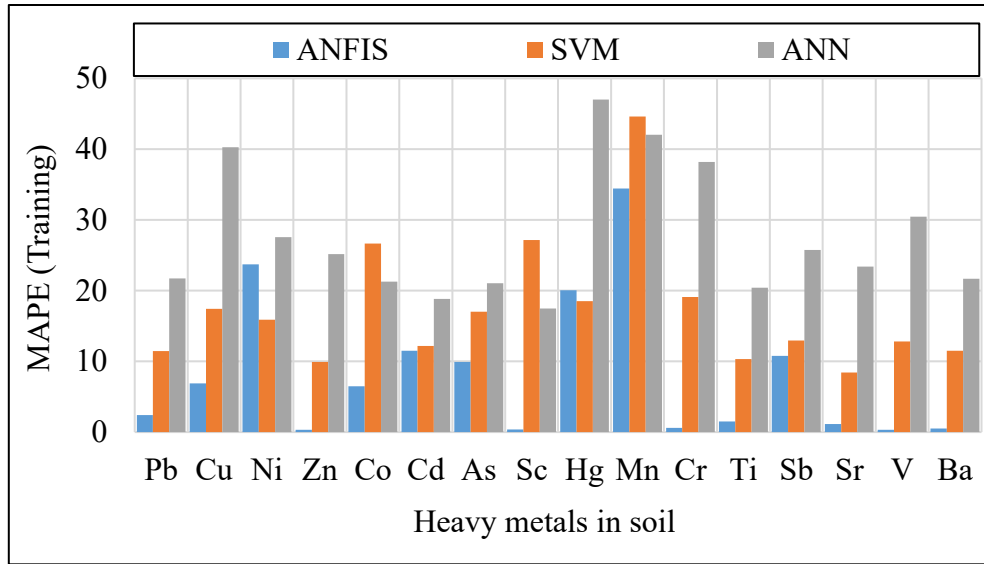
Figure 4.52: Variation of MAPE (in training) with various AI techniques.

In Figure 4.53, MAPE value for most of all the heavy metals in ANFIS model shows the less error than other models of AI techniques in testing. The MAPE value of ten heavy metals (Pb, Cu, Ni, As, Sc, Mn, Cr, Sr, V and Ba) were found better in ANN than SVM as well as the MAPE value of other heavy metals were found better in SVM than ANN model. Therefore, the performance of MAPE in testing can be expressed as ANFIS > ANN > SVM.
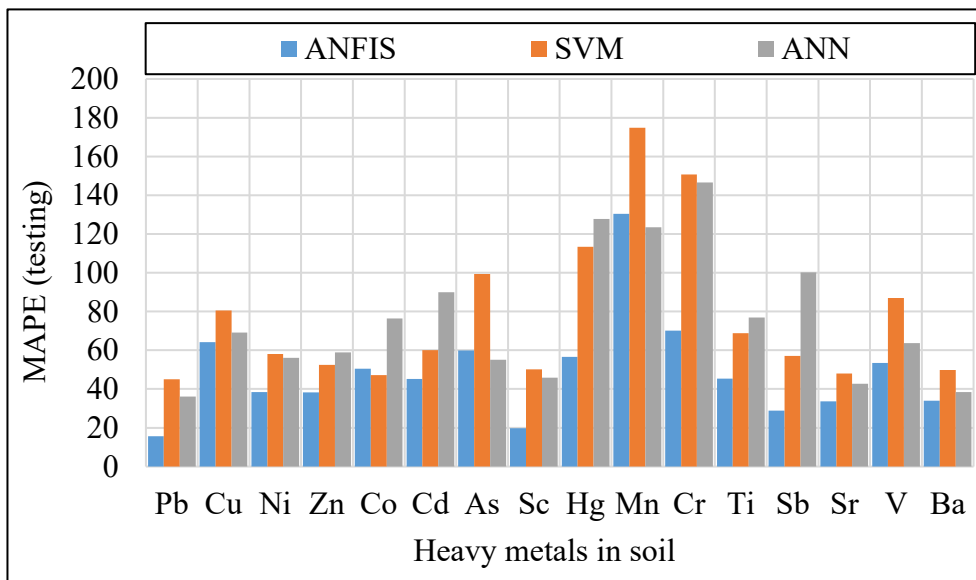
Figure 4.53: Variation of MAPE (in testing) with various AI techniques.

4.5.1.4 GRI Value

GRI is a statistical method to determine the reliability of a model. According to Leggett and Williams (1981), GRI value must be greater or equal 1 and GRI value 1 represents the perfectness and reliability of a model. Table 4.20 illustrates the results of GRI for all studied heavy metals with various AI techniques in both training and testing. Figure 4.54 and Figure 4.55 represent the variation of GRI value of different AI techniques for training and testing, respectively.

Table 4.20: Results of GRI of various AI techniques

Heavy metals	GRI (training)			GRI (testing)		
	ANFIS	SVM	ANN	ANFIS	SVM	ANN
Pb	1.08	1.19	1.32	1.20	1.58	1.59
Cu	1.14	1.28	1.48	1.91	1.86	2.27
Ni	1.35	1.25	1.49	1.46	1.64	1.69
Zn	1.01	1.16	1.35	1.48	1.63	1.71
Co	1.10	1.38	1.37	1.80	1.56	1.81
Cd	1.20	1.19	1.32	1.55	1.70	2.03
As	1.20	1.27	1.38	1.69	2.11	1.92
Sc	1.01	1.38	1.26	1.26	1.60	1.60
Hg	1.30	1.24	1.60	1.65	2.30	2.61
Mn	4.20	1.56	1.85	2.40	2.65	2.47
Cr	1.02	1.30	1.73	1.90	2.52	2.40
Ti	1.03	1.17	1.34	1.57	1.75	1.84
Sb	1.17	1.21	1.45	1.37	1.68	1.96
Sr	1.03	1.14	1.32	1.42	1.57	1.56
V	1.01	1.20	1.49	1.71	1.91	1.84
Ba	1.01	1.19	1.34	1.47	1.58	1.48

In Figure 4.54, the GRI value for most of all heavy metals were found very close to 1 in ANFIS model than that of other models of AI techniques for training. On the contrary, ANN shows more scattered values of GRI than that of SVM model. Therefore, the performance of GRI in training can be expressed as ANFIS > SVM > ANN.
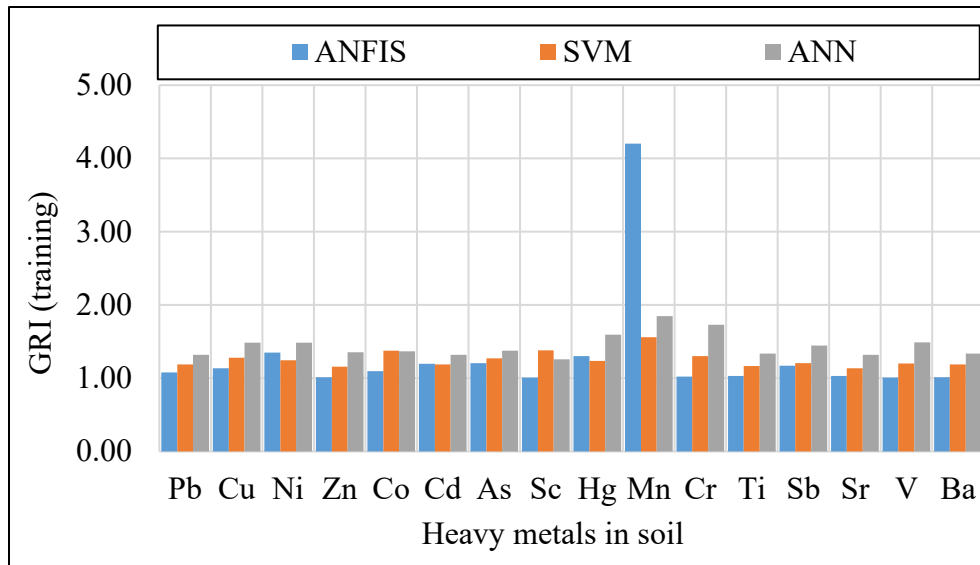
Figure 4.54: Variation of GRI (training) with various AI techniques.

In Figure 4.55, the GRI value for most of all heavy metals were found near about 1 in ANFIS model and it shows lower value of GRI than other models of AI techniques for testing. On the contrary, ANN shows more scattered values of GRI from 1 than SVM model. Therefore, the performance of GRI in testing can be expressed as ANFIS > SVM > ANN.
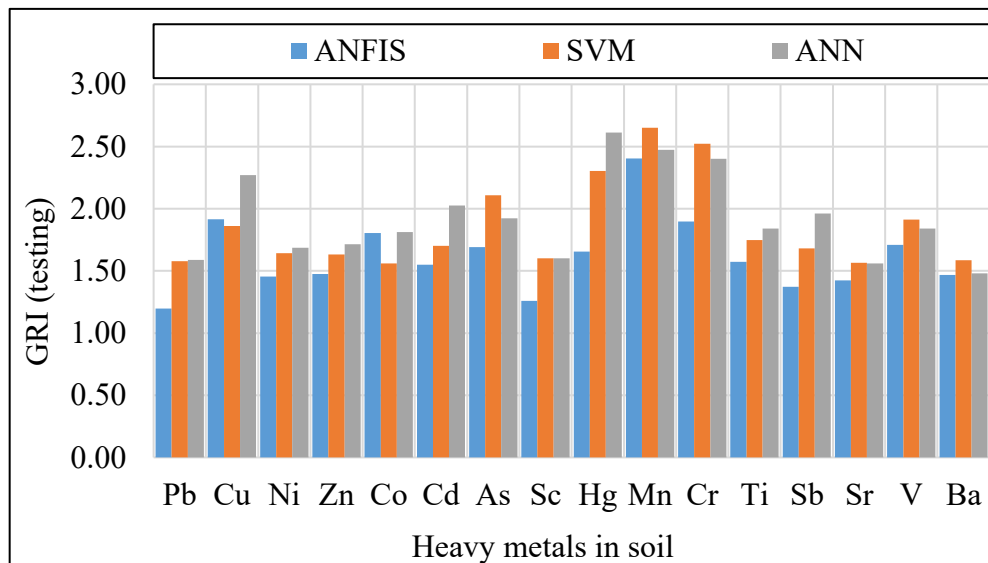
Figure 4.55: Variation of GRI (testing) with various AI techniques.

4.5.1.5 Percentage Recovery

The results of mean percentage recovery for all studied heavy metals in soils with various AI techniques in both training and testing are provided in Table 4.21. The percentage recovery means what percentage of measured value is recovered by the predicted value.

Table 4. 21: Mean percentage recovery of various AI techniques for training and testing

Heavy metals	Mean percentage recovery (training)			Mean percentage recovery (testing)		
	ANFIS	SVM	ANN	ANFIS	SVM	ANN
Pb	100.35	100.47	102.84	110.28	124.41	103.13
Cu	100.85	104.57	80.23	143.23	162.35	114.60
Ni	109.47	107.65	117.17	125.15	147.66	135.68
Zn	100.02	102.78	96.93	124.59	137.16	135.69
Co	100.80	115.05	113.17	114.86	134.12	158.96
Cd	103.18	102.53	105.80	136.98	143.98	171.83
As	101.76	106.24	106.29	151.14	178.60	116.94
Sc	100.01	108.37	103.18	113.73	137.76	128.29
Hg	109.73	107.67	105.14	143.50	197.82	199.67
Mn	117.20	107.92	148.16	208.35	247.67	196.97
Cr	100.11	105.71	140.81	151.14	221.40	221.65
Ti	100.08	102.46	108.92	133.63	155.36	159.17
Sb	102.77	103.81	117.99	115.53	140.01	183.08
Sr	100.05	101.41	97.58	125.14	134.71	122.73
V	100.01	105.16	114.04	137.61	170.52	139.00
Ba	100.02	103.41	105.30	122.36	137.29	124.63

According to Walfish (2006), 20-200% recovery represents the acceptance of model, 100% recovery represents the fit level and 80-120% recovery represents the robustness of the model. The results of mean percentage recovery were clearly expressed in Figure 4.56 and Figure 4.57 for training and testing, respectively for the judgement of various AI techniques.

In Figure 4.56, the mean percentage recovery values for most of all heavy metals were found very closer to the fit level (100%) in ANFIS model than that of other models of AI techniques for training. In SVM, the mean percentage recovery were found in the ranges of 80-120% and most of them were near to fit level (100%). On the contrary, ANN shows more scattered values

of mean percentage recovery from fit level than SVM model. Therefore, the performance of mean percentage recovery in training can be expressed as ANFIS > SVM > ANN.
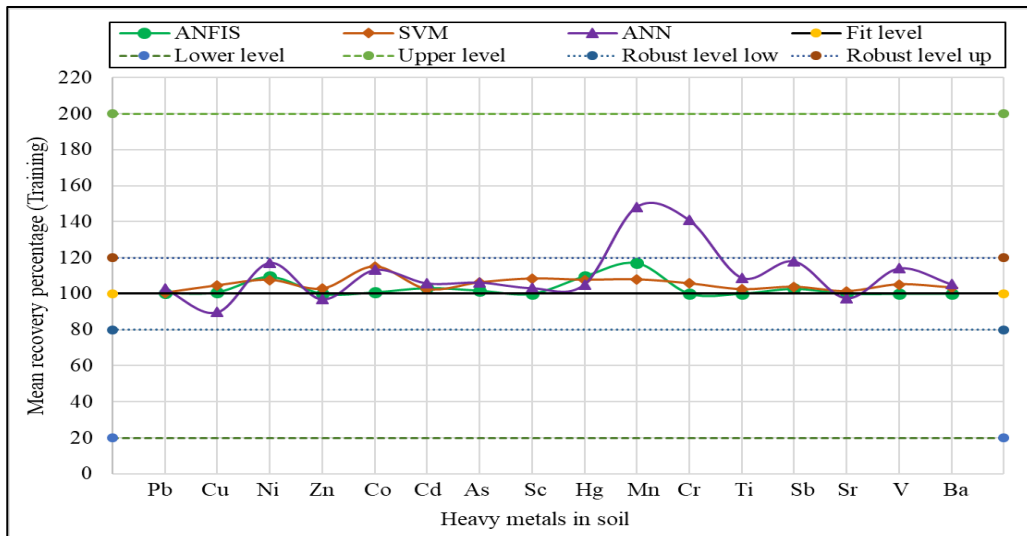
Figure 4.56: Classify models of AI techniques for percentage recovery (training).

In Figure 4.57, the mean percentage recovery values were found in acceptable ranges 20-200% for most of all heavy metals in case of all AI techniques in testing. Among three AI techniques, ANFIS shows the mean percentage recovery values near in robust level (120%) and represents better performance than other models.
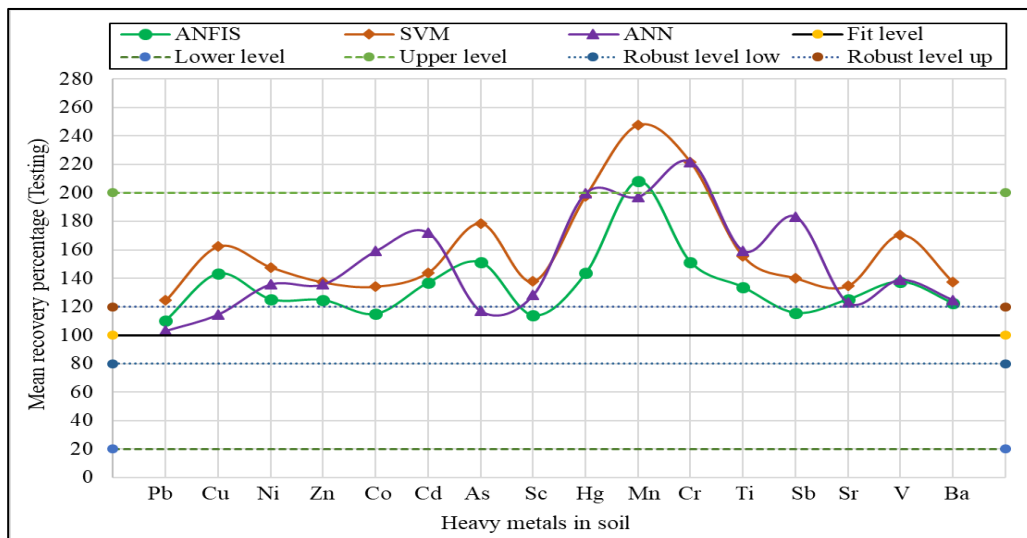
Figure 4.57: Classify models of AI techniques for percentage recovery (testing).

On the contrary, SVM shows more scattered of mean percentage recovery from robust level (120%) than that of ANN model. Therefore, the performance of mean percentage recovery in testing can be expressed as ANFIS > ANN > SVM.

4.5.2 Predicted Results from AI Techniques

A comparative analysis between predicted and measured concentrations from various AI techniques such as ANFIS, SVM and ANN through MATLAB for both of training and testing were performed and hence discussed in the following sections.

The predicted concentrations of Pb, Hg, Sb and Sc in soils was expressed in Figure 4.58 to Figure 4.61, respectively from various AI techniques for training. The predicted concentration of Pb (Figure 4.58) in ANFIS model was very closed to the measured concentration than other AI techniques, whereas, ANN model was shown less closer value of measured concentration than that of other AI techniques.
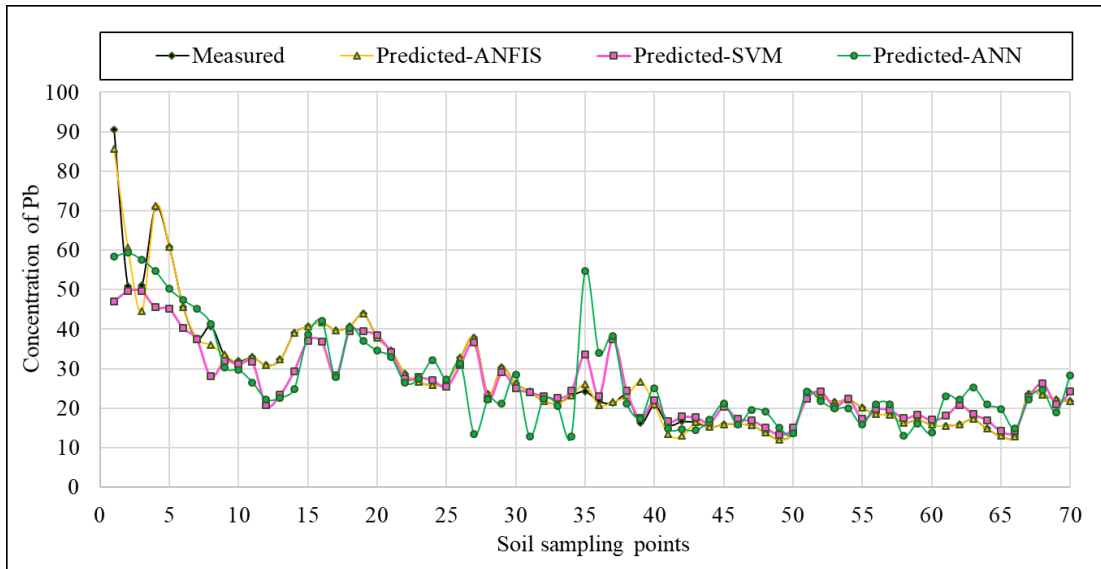
Figure 4.58: Variation of predicted results of Pb for different AI techniques in training.

Similarly, Figures 4.59, 4.60 and 4.61 represents very closer relationship between predicted and measured concentrations for Hg, Sb and Sc, respectively in ANFIS model than others model. ANN model shown less closer predicted concentrations with measured concentration than SVM

142

model. Therefore, the performance of the predicted results of various AI techniques for training can be expressed as ANFIS > SVM > ANN.
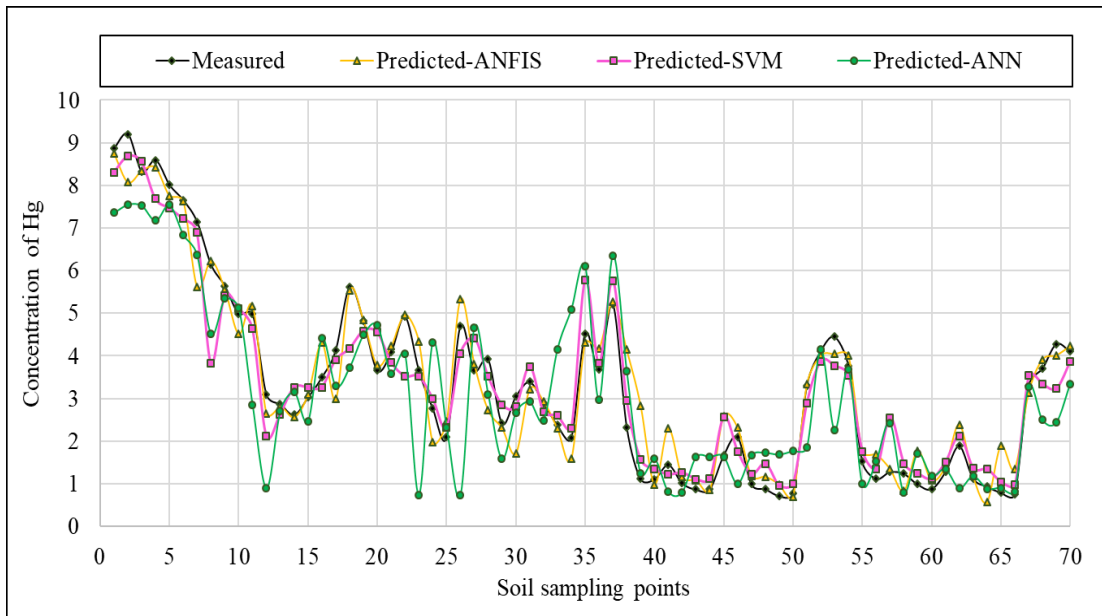
Figure 4.59: Variation of predicted results of Hg for different AI techniques in training.
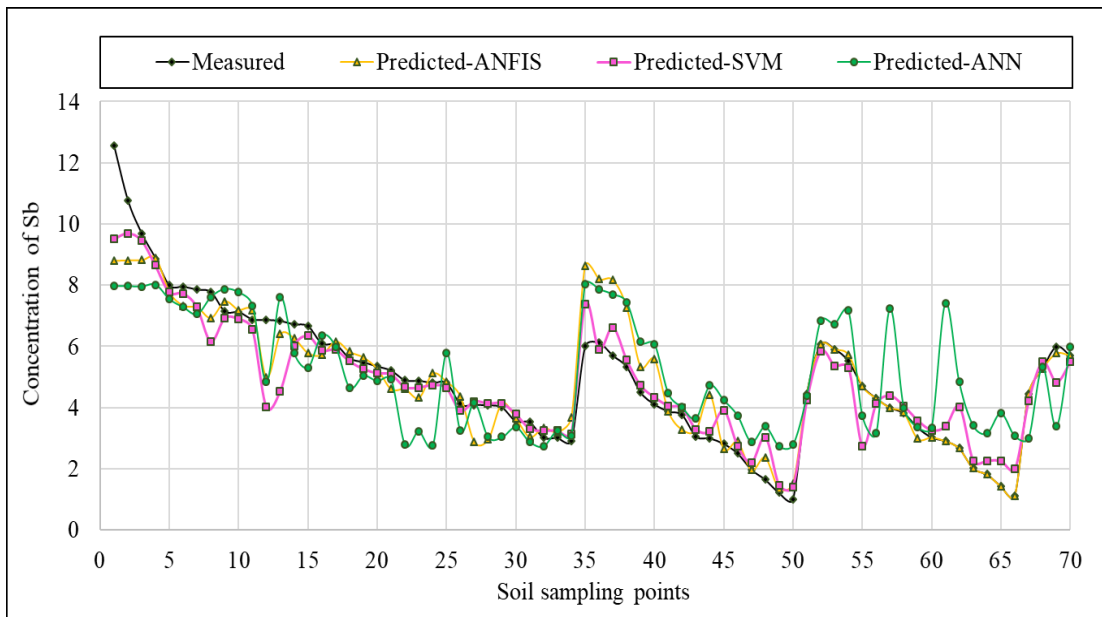
Figure 4.60: Variation of predicted results of Sb for different AI techniques in training.
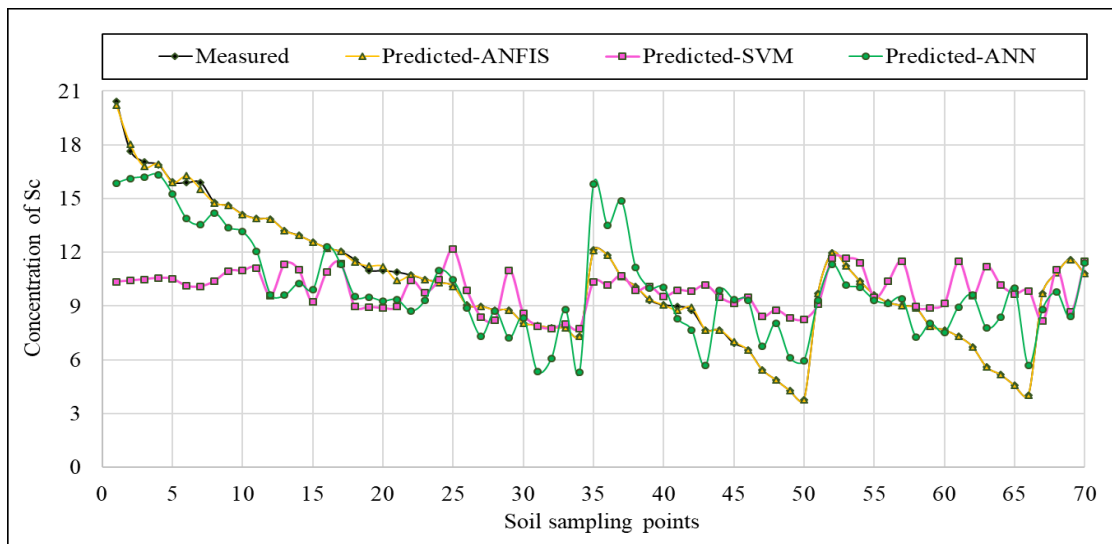
Figure 4.61: Variation of predicted results of Sc for different AI techniques in training.

Moreover, the predicted results of various AI techniques in training for other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure G-49 to G-60 in the Appendix-G.

The predicted results of various AI techniques for testing was expressed in Figure 4.62 to Figure 4.65 for Pb, Hg, Sb and Sc, respectively. Predicted concentration of Pb (Figure 4.62) in ANFIS model was closer to the measured concentration than other AI techniques whereas, ANN model was shown less closer predicted value with measured concentration than others model.
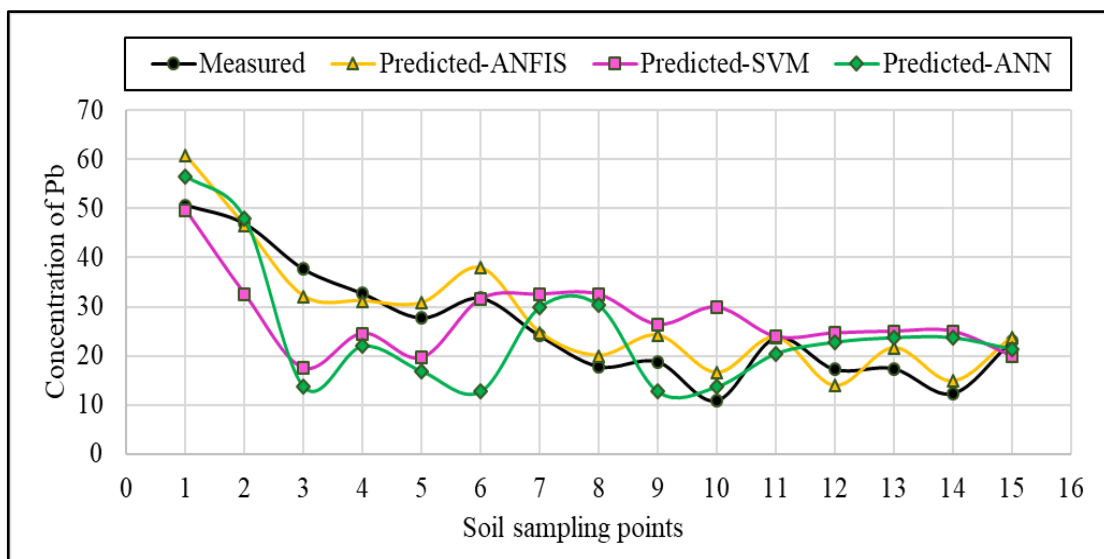
Figure 4.62: Variation of predicted results of Pb for different AI techniques in testing.

144

Similarly, Figures 4.63 for Hg, 4.64 for Sb and 4.65 for Sc were shown very closer predicted concentration to measured concentration in ANFIS model than others model.
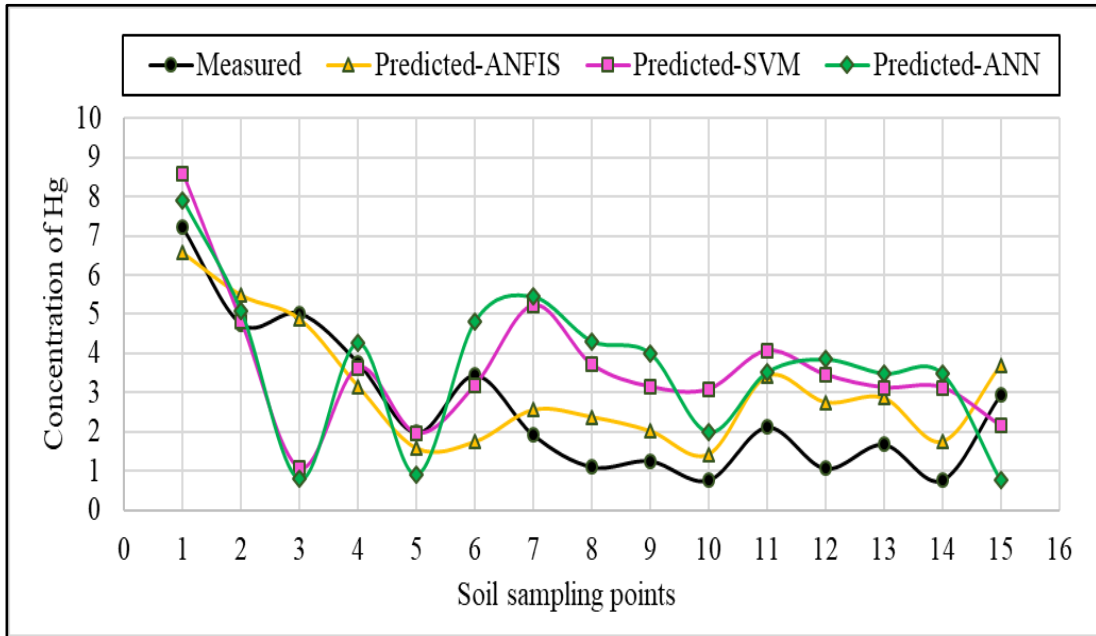
Figure 4.63: Variation of predicted results of Hg for different AI techniques in testing.
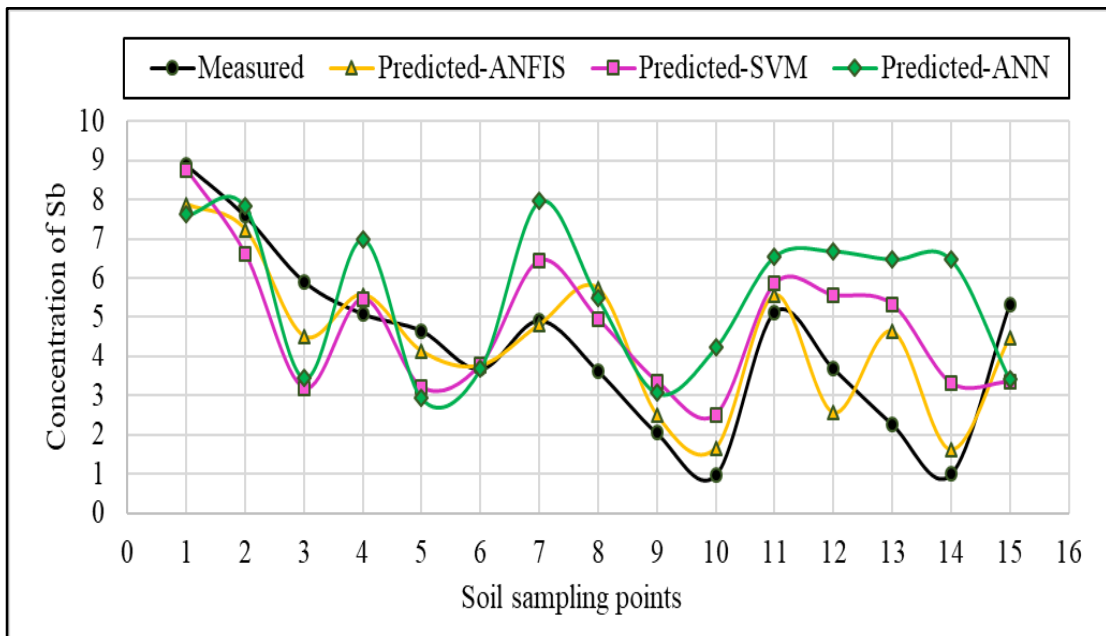
Figure 4.64: Variation of predicted results of Sb for different AI techniques in testing.
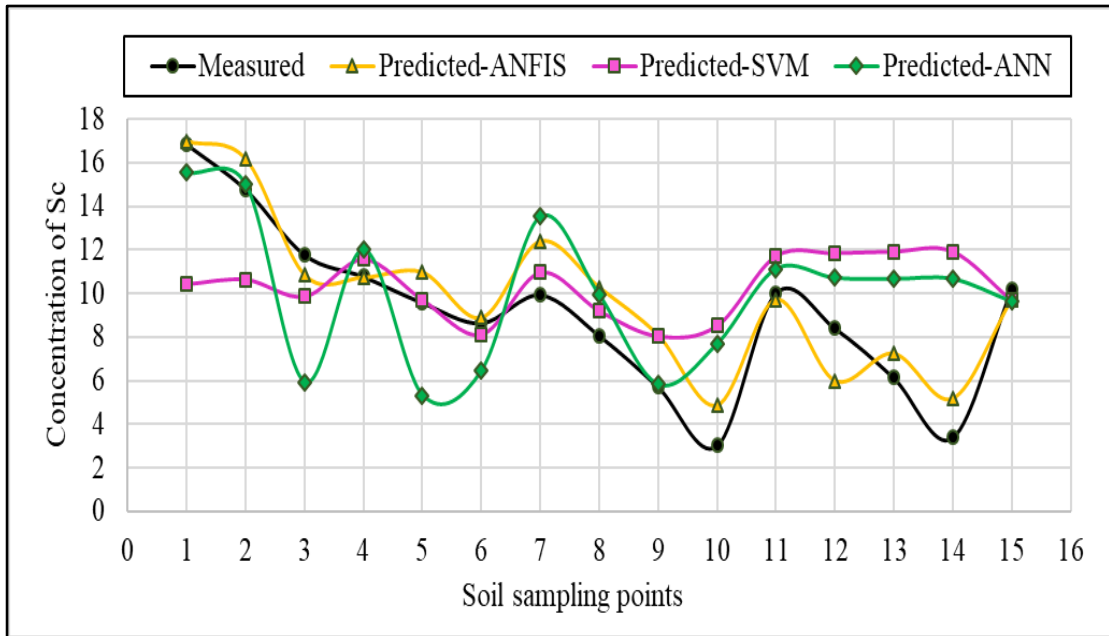
Figure 4.65: Variation of predicted results of Sc for different AI techniques in testing.

ANN model was shown less closer value of measured concentration than SVM model. Therefore, the performance of the predicted results of various AI techniques for testing can be expressed as ANFIS > SVM > ANN. Moreover, the predicted results of various AI techniques in testing for other heavy metals of Cu, Ni, Zn, Co, Cd, As, Mn, Cr, Ti, Sr, V and Ba were reported in Figure H-61 to H-72 in the Appendix-H.

4.6 Summary of Selected Functions and Algorithms of Models

In this study, for the analysis of heavy metal concentrations in soils, best-fitted AI technique was selected. To these attempts, ANFIS, SVM and ANN with various functions and algorithms were running through MATLAB to fix the best functions and algorithms. An analysis was also ready for comparing the functions, algorithms, inputs and outputs of various AI techniques used by various researchers in different fields in the literatures provided in Table 4.22. In the present study for ANFIS, different input membership function (MF) like gaussmf, trimf, trapmf, psigmf, gbellmf; output MF like linear and constant as well as optimization method such as hybrid or back-propagation (BP) through sub-clustering partitioning (SCP) were considered. Finally, SCP with gaussmf, linear and hybrid was proved as the best functions and algorithms to analyse the

146

heavy metal concentrations in soils (Table 4.23) by using inputs as latitude and longitude of the soil sampling points. The researchers of Sadrossadat et al. (2016) conducted a study for the prediction of the resilient modulus of flexible pavement of subgrade soils and proved that FCM with gaussmf, linear was the best functions and algorithms to analyse the resilient modulus by using input as P#200, LL, PI, $w_{opt.}$ w_c, S_r, q_u, σ_3, σ_d (Table 4.22).

In addition, in the present study for SVM analysis, different kernel functions like SVM-L, SVM-Q and SVM-RBF for fold numbers 5, 10, 15 and 20 were used. The SVM-RBF with fold number 15 was selected as the best-fitted kernel function for the prediction of heavy metal concentrations in soils (Table 4.23). A study conducted by Gholami et al. (2011) for the prediction of the concentration of Ni and Fe and proved that SVM-RBF as the best-fitted kernel function in the field of surface water with input variables pH, SO_4, HCO_3, TDS, EC, Mg, and Ca (Table 4.22).

Furthermore, in the present study, for ANN analysis, four neuron structures were formed with different neuron numbers of 5, 10, 15 and 20 successively. In addition, for ANN analysis, nine models were formed by interchanging different training functions like TRAINLM, TRAINOSS and TRAINSCG as well as transfer functions like TANSIG, PURELIN and LOGSIG. In present study, the model LT (including the functions levenberg-marquardt and TANSIG) with neuron structure 2-10-1 proved the best-fitted algorithm for the prediction of heavy metal concentrations in soils (Table 4.23). However, Gandhimathi et al. (2016) conducted a research for predicting the concentration of Pb, Cr, As, Fe, Hg, and Cd in soils and stated that FFNN with neuron structure 2-20-1 as the best combination of ANN model using input variables as latitude and longitude of soil sampling point (Table 4.22).

Here, it can be concluded that one can easily be evaluated the heavy metal concentrations in soils using these selected functions and algorithms (described in Table 4.23) of AI techniques without further analysis.

Table 4.22: Various Functions and Algorithms of Models in Literatures

Serial no.	Objectives of literatures	Model	Used functions and algorithms		Input and output		Reference
			Present study	Literature	Present study	Literature	
1	Prediction of the resilient modulus of flexible pavement subgrade soils	ANFIS	GENFIS: Sub-clustering partitioning and grid partitioning; Input MF: Gaussmf, trimf, trapmf, psigmf, gbellmf; Output MF: Linear and constant; Optimization method: Hybrid and back-propagation; No. of epoach: 100	GENFIS: Fuzzy c-means (FCM) clustering, Input (MF): Gaussianmf, Output MF: Linear	Input: Latitude and longitude; Output: Concentration of heavy metals	Input: P#200, LL, PI, wopt. wc, Sr, qu, σ3, σd; Output: MR	Sadrossadat et al., 2016
2	Prediction of water quality parameters (BOD, COD) of Karoon River			Input MF: Gaussmf, Output MF: Constant, Optimization method: Hybrid		Input: EC, pH, Ca, Mg, Na, Turbidity, PO4, NO3, NO2; Output: DO, BOD, COD	Emamgholizadeh et al., 2014
3	Predict of Cadmium (Cd) Concentrations in the Filyos River, Turkey			Input MF: Gaussmf, Output MF: Linear, Optimization method: Hybrid		Input: Concentration of Fe, Cu, Mn, Zn, Ni, Cr; Output: Concentration of Cd	Sonmez et al., 2018
4	Prediction of the concentration of Ni and Fe using support vector machine (SVM).	SVM	Kernel function: SVM-L, SVM-C, SVM-Q and SVM-RBF Fold number: 5, 10, 15 and 20	Kernel function: SVM-RBF	Input: Latitude and longitude; Output: Concentration of heavy metals	Input: pH, SO4, HCO3, TDS, EC, Mg, and Ca; Output: Fe and Ni	Gholami et al., 2011
5	Prediction of the heavy metals (Mn, Cu, Pb and Fe) in Sarcheshmeh copper mine, Iran using support vector machine (SVM).			Kernel function: SVM-RBF		Input: pH, SO4 and Mg; Output: Cu, Fe, Mn and Zn	Aryafar et al., 2012
6	Prediction of soil cation exchange capacity of an agricultural research station by SVM			Kernel function: SVM-L, SVM-P, SVM-S and SVM-RBF		Input: Clay (%), Silt (%), Sand (%), Gypsum (%) and OM (%) Output: CEC	Jafarzadeh et al., 2016
7	Prediction of As, Cu, Pb, and Zn concentration in the groundwater resources of Ghahavand Plain	ANN	Feed-forward back propagation neural network (FFBPNN); Training functions: TRAINLM, TRAINOSS, TRAINSCG; Transfer functions: Tansig, purelin, logsig; Neural structure: 2-5-1, 2-10-1, 2-15-1 and 2-20-1	FFNN with Levenberg-marquardt (LM), Sigmoid and linear functions	Input: Latitude and longitude; Output: Concentration of heavy metals	Output: Concentration of As, Cu, Pb, and Zn	Alizamir, M., & Sobhanardakani, S. 2016
8	Forecasting the arsenic (As) lead (Pb), and zinc (Zn) concentration of groundwater resources of Asadabad plain			FFNN with Levenberg-marquardt (LM) and Bayesian regularization (BR) algorithms		Output: Concentration of As, Pb, and Zn	Alizamir, M., & Sobhanardakani, S. 2017
9	Predicting the concentration of Pb, Cr, As, Fe, Hg, and Cd in soils			FFNN & 6 neuron structure (2-10-1, 2-12-1, 2-14-1, 2-16-1, 2-18-1 and 2-20-1)		Input: latitude and longitude Output: Concentration of Pb, Cr, As, Fe, Hg, and Cd	Gandhimathi et al., 2016

Table 4.23: Results of various functions and algorithms of different models

| AI techniques | Selected functions and algorithms | |
	Present study	Literature
ANFIS	Genfis: SCP Input MF: gaussmf Output MF: linear and Optimization method: hybrid	Input MF: gaussmf Output MF: linear and Optimization method: hybrid
SVM	Kernel Function: SVM-RBF Fold Number: 15	Kernel Function: SVM-RBF
ANN	Feed-forward back propagation neural network (FFNN) Training Function: Levenberg-marquardt Transfer Function: Tangent Sigmoid (tansig) Neuron structure: 2-10-1	Feed-forward back propagation neural network (FFNN) Training Function: Levenberg-marquardt
		Feed-forward back propagation neural network (FFNN) Neuron structure: 2-20-1

4.7 Final Evaluation of Results of AI Techniques

The main focus of this section to select the best fitted model of AI techniques. For this reason, the detailed results where various researches used AI techniques in different field like soil, surface water (SW), ground water (GW) etc. were collected from literature. A comparative study was performed to select the best fitted model of AI techniques. A study conducted by Emamgholizadeh et al. (2014) and showed that the R values of DO, BOD and COD were found greater for ANFIS than that of ANN both for training and testing whereas RMSE was found lower for ANFIS than that of ANN. Among them, the results of COD revealed the best performance with R=0.959 and RMSE=4.21 for ANFIS as well as R=0.917 and RMSE=5.57 for ANN. In addition, the performance of SVM and GRNN methods in the training and testing steps in Shur River, Sarcheshmeh copper mine, Iran for Cu, Mn, Zn and Fe (Rooki et al., 2011) was also highlighted in this section. In this literature, SVM shows the better performance of prediction of Cu, Mn, Zn and Fe with higher value of R and lower value of RMSE than that of ANN in both training and testing steps. The results of Cu were found R=0.990 and RMSE= 3.21

(training) in SVM whereas R=0.990 and RMSE= 5.47 (training) in ANN. In the present study, to evaluate the performance of various AI techniques, different prediction parameters like R, RMSE, MAPE, GRI and percentage recovery were considered. The results of various AI techniques from present study and literature were summarised in Table 4.24. In the present study, ANFIS shows the best performance for all criteria of goodness like R, RMSE, MAPE, GRI and percentage recovery. In addition, SVM shows the comparatively better results for most of the prediction parameters than ANN.

Table 4.24: Summary of results of various AI techniques

Prediction parameters		Present study	Literature	Final remarks
R value	Training	ANFIS > ANN > SVM	ANFIS > ANN	
			SVM > ANN	
	Testing	ANFIS > SVM ≥ ANN	ANFIS > ANN	
			SVM > ANN	
RMSE	Training	ANFIS > ANN > SVM	ANFIS > ANN	
			SVM > ANN	
	Testing	ANFIS > SVM > ANN	ANFIS > ANN	ANFIS > SVM > ANN
			SVM > ANN	
MAPE	Training	ANFIS > SVM > ANN	-----	
	Testing	ANFIS > ANN > SVM	-----	
Mean percentage recovery	Training	ANFIS > SVM > ANN	-----	
	Testing	ANFIS > ANN > SVM	-----	
GRI	Training	ANFIS > SVM > ANN	-----	
	Testing	ANFIS > SVM > ANN	-----	
Performance of predicted results in training		ANFIS > SVM > ANN	-----	
Performance of predicted results in testing		ANFIS > SVM > ANN	-----	

Based on results published by Emamgholizadeh et al. (2014) in the literature, it was observed the following order of AI techniques as ANFIS > ANN for both training and testing. In addition, the results also stated by Rooki et al. (2011) and proved SVM > ANN for both training and testing. Therefore, it can be finally expressed that the performance of AI techniques was considered by the sequence of ANFIS > SVM > ANN for best prediction of heavy metal concentrations in soils of waste disposal site (Table 4.24).

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This study aims to fix the functions and algorithms of various AI techniques for the analysis of heavy metal concentrations in soils of a selected waste disposal site in old Rajbandh, Khulna. The AI techniques can be used to solve engineering problems without mathematical modelling. In this study, AI techniques such as ANFIS, SVM and ANN were implemented through MATLAB. The prediction of heavy metal concentrations in soils through AI techniques play an important role to minimize laboratory error, sampling time and testing budget. The performance for all criterion of goodness like R, RMSE, MAPE, GRI and percentage recovery available in the literatures were considered to validate the predicted results from AI techniques. The present study executed the following conclusions:

1. ANFIS model was a reliable technique than that of other counterparts of SVM and ANN to analyse the heavy metal concentrations in soil with the acceptable degree of robustness and accuracy.

2. A combination of best functions and algorithms in ANFIS model was selected for the prediction of heavy metal concentrations in soils. This combination of model was GENFIS: SCP, Input MF: Gaussmf, Output MF: Linear, Optimization Method: Hybrid and no. of epoch: 100.

3. A rule viewer in ANFIS was developed to compute the concentration of a particular heavy metal in soils of the selected waste disposal site by inserting GPS values (latitude and longitude) only.

4. Among all the heavy metals in ANFIS analysis, the maximum R-value was found 0.999 with the minimum RMSE 0.12 indicating the best correlation in prediction of Sc in soils. The others value of prediction parameters (MAPE= 36.00, GRI=1.50, percentage recovery=123.43%) for Sc were found within the acceptable limits.

5. The predicted concentrations of Sc varies from 4.85 to 16.99, whereas, measured from 3.02 to 16.83 mg/kg in ANFIS (testing) for unknown soil sampling points indicating the predicted and measured concentrations are close to each other.

6. A combination of functions in SVM model was selected based on its best performance with kernel function: SVM-RBF and fold number: 15.

7. Among all the heavy metals in testing for SVM analysis, maximum R-value was found 0.73 for Cu with RMSE 2.03; in addition, minimum R-value was found 0.06 with maximum RMSE 7.24 for Mn. (MAPE= 80.48, GRI=1.86, percentage recovery=162.35% for Cu).

8. The model of ANN with training function (Levenberg-Marquardt), transfer function (Tansig) and no. of neurons 10 was selected for the prediction of heavy metals.

9. The performance of artificial intelligence techniques can be expressed in a sequence of ANFIS > SVM > ANN among the AI techniques used in this study.

Finally, it can be concluded that these selected AI techniques with fixed functions and algorithms may be used of other researchers without further analysis of AI techniques to predict heavy metal concentrations in soils of a selected waste disposal site.

5.2 Recommendations for Further Studies

The following recommendations are required for further studies:

1. The intensity of heavy metal concentration should be higher in the centre of waste disposal site; however, in this study, higher intensity of Hg was found at the outer edge of waste disposal site. This reason need to be addressed in further studies.

2. The values of RMSE for most of the heavy metals were close to zero. But the values of RMSE for Mn and Ti were found 3.5 and 29.60, respectively. This sudden rise of error should be defined in further studies.

3. The MAPE value of Mn was in out of range, the reason should be identified in further researches.

4. There are more AI techniques in literature among which only three techniques were used in present study. Other AI techniques can be included for comparing and selecting the best in further studies.

5. Other membership functions (MF) can be used when this study will be performed in future.

6. Time variation may be considered in future studies.

References

Addae, E. 2013, "An assessment of heavy metal contamination in soils and vegetation: a case study of Korle Lagoon Reclamation site". Doctoral dissertation, University of Ghana.

Alizamir, M., and Sobhanardakani, S. 2016, "Forecasting of heavy metals concentration in groundwater resources of Asadabad plain using artificial neural network approach". Journal of Advances in Environmental Health Research, Vol. 4(2), pp. 68-77.

Alizamir, M., and Sobhanardakani, S. 2017, "Predicting arsenic and heavy metals contamination in groundwater resources of Ghahavand plain based on an artificial neural network optimized by imperialist competitive algorithm". Journal of Environmental Health Engineering and Management, Vol. 4(4), pp. 225-23.

Alkaiem, L., and Sternberg, H. 2016, "Analysis of inclination measurement by means of artificial neural networks–A comparison of static and dynamic networks". Paper presented at the 3rd Joint International Symposium on Deformation Monitoring.

Alloway, B. J., Jackson, A. P., and Morgan, H. 1990, "The accumulation of cadmium by vegetables grown on soils contaminated from a variety of sources". Science of the total Environment, Vol. 91, pp. 223-236.

Aryafar, A., Gholami, R., Rooki, R., and Ardejani, F.D. 2012, "Heavy metal pollution assessment using support vector machine in the Shur River, Sarcheshmeh copper mine, Iran". Environmental earth sciences, Vol. 67(4), pp. 1191-1199.

Battiti, R., and Tecchiolli, G. 1995, "Training neural nets with the reactive tabu search. IEEE transactions on neural networks". Vol. 6(5), pp. 1185-1200.

Behzad, M., Asghari, K., Eazi, M., and Palhang, M. 2009, "Generalization performance of support vector machines and neural networks in runoff modelling". Expert Systems with applications, Vol. 36(4), pp. 7624-7629.

Branch, S.K. 2005, "Guidelines from the international conference on harmonisation". Journal of pharmaceutical and biomedical analysis, Vol. 38(5), pp. 798-805.

Chang, K. Y., Chen, C. S., and Hung, Y. P. 2010, "A ranking approach for human age's estimation based on face images". 20th International Conference on Pattern Recognition, IEEE, pp. 3396-3399.

Cherkassky, V., and Ma, Y. 2002, "Selection of meta-parameters for support vector regression". Paper presented at the International Conference on Artificial Neural Networks.

Choobbasti, A.J., Farrokhzad, F., and Rahim Mashaie, S., and Azar, P. H. 2015, "Mapping of soils layers using artificial neural network (case study of Babol, northern Iran) ". Journal of the South African Institution of Civil Engineering, Vol. 57(1), pp. 59-66.

Cortes, C., and Vapnik, V. 1995, "Support-vector networks". Machine learning, Vol. 20(3), pp. 273-297.

Daniel, D.E., and Koerner, R.M. 1995, "Waste containment facilities, Guidance for Construction Quality Assurance and Quality Control of Linear and Cover System". ASCE, Vol. 13, pp. 25-27.

Das, S. K., Samui, P., and Sabat, A. K. 2011, "Application of artificial intelligence to maximum dry density and unconfined compressive strength of cement stabilized soil". Geotechnical and Geological Engineering, Vol. 29(3), pp. 329-342.

Das, S. K., Samui, P., Sabat, A. K., and Sitharam, T. G. 2010, "Prediction of swelling pressure of soil using artificial intelligence techniques". Environmental Earth Sciences, Vol. 61(2), pp. 393-403.

Emamgholizadeh, S., Kashi, H., Marofpoor, I., and Zalaghi, E. 2014, "Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models". International Journal of Environmental Science and Technology, Vol. 11(3), pp. 645-656.

epa.gov/epawaste/nonhaz/municipal/web/html/, h. a.

Fahmida, K., & Rafizul, I. M. 2017, "An Investigation on Soil Quality and Heavy Metal Levels in Soil of Rajbandh Waste Disposal Site at Khulna, Bangladesh".

Gandhimathi, A., and Anbarasi, A. 2016, "Environmental Impact Assessments of Heavy Metal on Soil and Water for Coimbatore, India". Fifth World Conference on Applied Sciences, Engineering and Technology 02-04 June 2016, HCMUT, Vietnam.

George, D., and Mallery, P. 2010, "SPSS for Windows step by step: A simple study guide and reference". 17.0 update (10a ed.) Boston: Pearson.

Ghadimi, F. 2014, "Assessment of the sources of chemical elements in sediment from Arak Mighan Lake". International Journal of Sediment Research, Vol. 29(2), pp. 159-170.

Gholami, R., Kamkar-Rouhani, A., Ardejani, F. D., and Maleki, S. 2011, "Prediction of toxic metals concentration using artificial intelligence techniques". Applied Water Science, Vol. 1(3-4), pp. 125-134.

Gribble, G. W. 1994, "The natural production of chlorinated compounds". Environmental science and technology, Vol. 28(7), pp. A (310-319).

Hagan, M. T., and Menhaj, M. B. 1994, "Training feedforward networks with the Marquardt algorithm". IEEE transactions on Neural Networks, Vol. 5(6), pp. 989-993.

Hashim, M. A., and Chu, K. H. 2004, "Biosorption of cadmium by brown, green, and red seaweeds". Chemical Engineering Journal, Vol. 97(2-3), pp. 249-255.

Hasofer, A. M., and Lind, N. C. 1974, "Exact and invariant second-moment code format". Journal of the Engineering Mechanics division, Vol. 100(1), pp. 111-121.

Haykin, S. S. 2009, "Neural networks and learning machines/Simon Haykin". New York: Prentice Hall.

Hornik, K., Stinchcombe, M., and White, H. 1990, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks". Neural networks, Vol. 3(5), pp. 551-560.

Hossein Pour, M., Lashkaripour, G. R., and Dehghan, P. 2014, "Environmental pollution evaluation of steel plants for achieving sustainable development; case study: khorasan steel complex of Iran". Journal of Biodiversity and Environmental Sciences, 4.

https://cdn-images-1.medium.com/max/1600/0*ecA4Ls8kBYSM5nza.jpg.

https://cdn-images-1.medium.com/max/2000/1*ZpkLQf2FNfzfH4HXeMw4MQ.png.

https://cdn-images-1.medium.com/max/750/0*0o8xIA4k3gXUDCFU.png.

https://en.wikipedia.org/wiki/Contaminated_land.

https://en.wikipedia.org/wiki/Khulna.

https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcR3lHQJMZL_CglzCp9-rWhcFAo5AdPUmd-WMIUIyrIGS1qh4a9K.

https://i.stack.imgur.com/iIcbq.gif.

https://i0.wp.com/kashmirreader.com/wp-content/uploads/2017/02/Dumping-site.jpg?fit=701%2C311andssl=1).

https://link.springer.com/article/10.100. (2018).

https://pixfeeds.com/images/33/610239/1280-sanitary-landfill-vs-open-dump.png.

https://whatis.techtarget.com/definition/correlation-coefficient.

https://www.aarki.com/hsfs/hubfs/Blog/machine_learning.png?F.png.

https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/.

https://www.epa.gov/report-environment/contaminated-land.

https://www.mathworks.com/help/fuzzy/train-adaptive-neuro-fuzzy-inference-systems-gui.html.

https://www.saedsayad.com/support_vector_machine_reg.htm.

Jafarzadeh, A. A., Pal, M., Servati, M., FazeliFard, M. H., and Ghorbani, M. A. 2016, "Comparative analysis of support vector machine and artificial neural network models for soil cation exchange capacity prediction". International journal of environmental science and technology, Vol. 13(1), pp. 87-96.

Jang, J. S. 1993, "ANFIS: adaptive-network-based fuzzy inference system". IEEE transactions on systems, man, and cybernetics, Vol. 23(3), pp. 665-685.

Jang, J. S. R., Sun, C. T., and Mizutani, E. 1997, "Neuro-fuzzy and soft computing computational approach to learning and machine intelligence". IEEE Transactions on automatic control, Vol. 42(10), pp. 1482-1484.

Johansson, E. M., Dowla, F. U., and Goodman, D. M. 1991, "Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method". International Journal of Neural Systems, Vol. 2(4), pp. 291-301.

Kişi, Ö, and Öztürk, Ö. 2007, "Adaptive neurofuzzy computing technique for evapotranspiration estimation". Journal of Irrigation and Drainage Engineering, Vol. 133(4), pp. 368-379.

Kişi, Ö. 2008, "River flow forecasting and estimation using different artificial neural network techniques". Hydrology Research, Vol. 39(1), pp. 27-40.

Korzun, E. A., and Heck, H. H. 1990, "Sources and fates of lead and cadmium in municipal solid waste". Journal of the Air and Waste Management Association, Vol. 40(9), 1220-1226.

Lee, C. Y., and Chern, S. G. 2013, "Application of a support vector machine for liquefaction assessment". Journal of Marine Science and Technology, Vol. 21(3), pp. 318-324.

Leggett, R. W., and Williams, L. R. 1981, "A reliability index for models". Ecological Modelling, Vol. 13(4), pp. 303-312.

Ludwig, H., and Keller, A. 2003, "The WSLA framework: Specifying and monitoring service level agreements for web services". Journal of Network and Systems Management, Vol. 11(1), pp. 57-81.

Markopoulos, A. P., Manolakos, D. E., and Vaxevanidis, N. M. 2008, "Artificial neural network models for the prediction of surface roughness in electrical discharge machining". Journal of Intelligent Manufacturing, Vol. 19(3), pp. 283-292.

Mayfield, D. B., and Fairbrother, A. 2013, "Efforts to standardize wildlife toxicity values remain unrealized". Integrated environmental assessment and management, Vol. 9(1), pp. 114-123.

Morillo, J., Usero, J., and Gracia, I. 2002, "Partitioning of metals in sediments from the Odiel River (Spain) ". Environment international, Vol. 28(4), pp. 263-271.

Nasrabadi. 2015, "An Index Approach to Metallic Pollution in River Waters". International Journal of Environmental Research, Vol. 9 (1), pp. 385-394.

Oyeku, O. T., and Eludoyin, A. O. 2010, "Heavy metal contamination of groundwater resources in a Nigerian urban settlement". African Journal of Environmental Science and Technology, Vol. 4(4).

Pugh, M. 1999, "Path to affordable landfills: landfill technology in the developing countries". Wastes Management, pp. 58-59.

Quina, M. J., Santos, R. C., Bordado, J. C., and Quinta-Ferreira, R. M. 2008, "Characterization of air pollution control residues produced in a municipal solid waste incinerator in Portugal". Journal of Hazardous Materials, Vol. 152(2), pp. 853-869.

Rafizul, I. M., Alamgir, M., and Islam, M. M. 2011, "Evaluation of contamination potential of sanitary landfill lysimeter using leachate pollution index". In Proceedings Sardinia.

Rayer, S. 2007, "Population forecast accuracy: does the choice of summary measure of error matter?" Population Research and Policy Review, Vol. 26(2), pp. 163.

Rooki, R., Ardejani, F. D., Aryafar, A., and Asadi, A. B. 2011, "Prediction of heavy metals in acid mine drainage using artificial neural network from the Shur River of the Sarcheshmeh porphyry copper mine, Southeast Iran". Environmental earth sciences, Vol. 64(5), pp. 1303-1316.

Roy, S., and Dass, G. 2014, "Statistical models for the prediction of shear strength parameters at Sirsa, India". International Journal of Civil and Structural Engineering, Vol. 4(4), pp. 483.

Rumsey, D. J. 2015, "U Can: statistics for dummies". John Wiley and Sons.

Sadrossadat, E., Heidaripanah, A., and Osouli, S. 2016, "Prediction of the resilient modulus of flexible pavement subgrade soils using adaptive neuro-fuzzy inference systems". Construction and Building Materials, Vol. 123, pp. 235-247.

Samui, P. 2008, "Support vector machine applied to settlement of shallow foundations on cohesionless soils". Computers and Geotechnics, Vol. 35(3), pp. 419-427.

Sanjida, K., and Rafizul, I. M. 2018, "Multivariate statistics and spatial distribution of heavy metals in soils of waste disposal site in south western region of Bangladesh". Paper presented at the 33rd Int. Conference on Solid Waste Technology and Management (ICSW 2018), Annapolis, Washington, DC, USA, (ISSN: 1091-8043)

Schweizer, K. 2010, "Some guidelines concerning the modelling of traits and abilities in test construction".

Singh, D. J., and Kalamdhad, A. 2011, "Effects of Heavy Metals on Soils, Plants, Human Health and Aquatic Life". Vol. 1.

Singh, R., Gautam, N., Mishra, A., and Gupta, R. 2011, "Heavy metals and living systems: An overview". Indian journal of pharmacology, Vol. 43(3), pp. 246.

Smith, S. L., and Mosier, J. N. 1986, "Guidelines for designing user interface software (No. MTR-10090) ". Bedford, MA: Mitre Corporation

Smola, A. J., and Schölkopf, B. 2004, "A tutorial on support vector regression". Statistics and computing, Vol. 14(3), pp. 199-222.

Sonmez, A. Y., Kale, S., Ozdemir, R. C., and Kadak, A. E. 2018, "An Adaptive Neuro-Fuzzy Inference System (ANFIS) to Predict of Cadmium (Cd) Concentrations in the Filyos River, Turkey". Turkish Journal of Fisheries and Aquatic Sciences, Vol. 18(12), pp. 1333-1343.

Soyupak, S., Karaer, F., Gürbüz, H., Kivrak, E., Sentürk, E., and Yazici, A. 2003, "A neural network-based approach for calculating dissolved oxygen profiles in reservoirs". Neural Computing and Applications, Vol. 12(3-4), pp. 166-172.

Suri, F. M. 2017, "Signal classification using Bayesian regularization and Levenberg-Marquardt algorithm". In 2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), pp. 1-6.

Tahir, N. M., Chee, P. S., and Jaafar, M. 2007, "Determination of heavy metals content in soils

and indoor dusts from nurseries in Dungun, Terengganu". The Malaysian Journal of Analytical Sciences, Vol. 11(1), pp. 280-286.

Talib, H. A., Ali, K. M., and Jamaludin, K. R. 2008, "Quality assurance in halal food manufacturing in Malaysia: A preliminary study". In Proceedings of International Conference on Mechanical and Manufacturing Engineering (ICME2008), pp. 21-23.

Thurgood, M. 1999, "Solid Waste Landfills: Decision-Makers Guide Summary".

Vaalgamaa, S., and Conley, D. J. 2008, "Detecting environmental change in estuaries: Nutrient and heavy metal distributions in sediment cores in estuaries from the Gulf of Finland, Baltic Sea". Estuarine, Coastal and Shelf Science, Vol. 76(1), pp. 45-56.

Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T., and Alkon, D. L. 1988, "Accelerating the convergence of the back-propagation method". Biological cybernetics, Vol. 59(4-5), pp. 257-263.

Walfish, S. 2006, "Analytical methods: a statistical perspective on the ICH Q2A and Q2B guidelines for validation of analytical methods". Bio Pharm International, Vol. 19(12), pp. 1-6.

Walker, D. J., Clemente, R., Roig, A., and Bernal, M. P. 2003, "The effects of soil amendments on heavy metal bioavailability in two contaminated Mediterranean soils". Environmental Pollution, Vol. 122(2), pp. 303-312.

Wang, L. 2005, "Support vector machines: theory and applications". Springer Science and Business Media, Vol. 177.

Wei, B., and Yang, L. 2010, "A review of heavy metal contaminations in urban soils, urban road dusts and agricultural soils from China". Micro chemical journal, Vol. 94(2), pp. 99-107.

Xiao, R., Bai, J., Huang, L., Zhang, H., Cui, B., and Liu, X. 2013, "Distribution and pollution, toxicity and risk assessment of heavy metals in sediments from urban and rural rivers of the Pearl River delta in southern China". Ecotoxicology, Vol. 22(10), pp. 1564-1575.

Zhang, J., Tang, M., and Viikari, L. 2012, "Xylans inhibit enzymatic hydrolysis of lignocellulosic materials by cellulases". Bio resource technology, Vol. 121, pp. 8-12.

Appendix-A

Results of Heavy Metal Analysis and Assessment of ANFIS
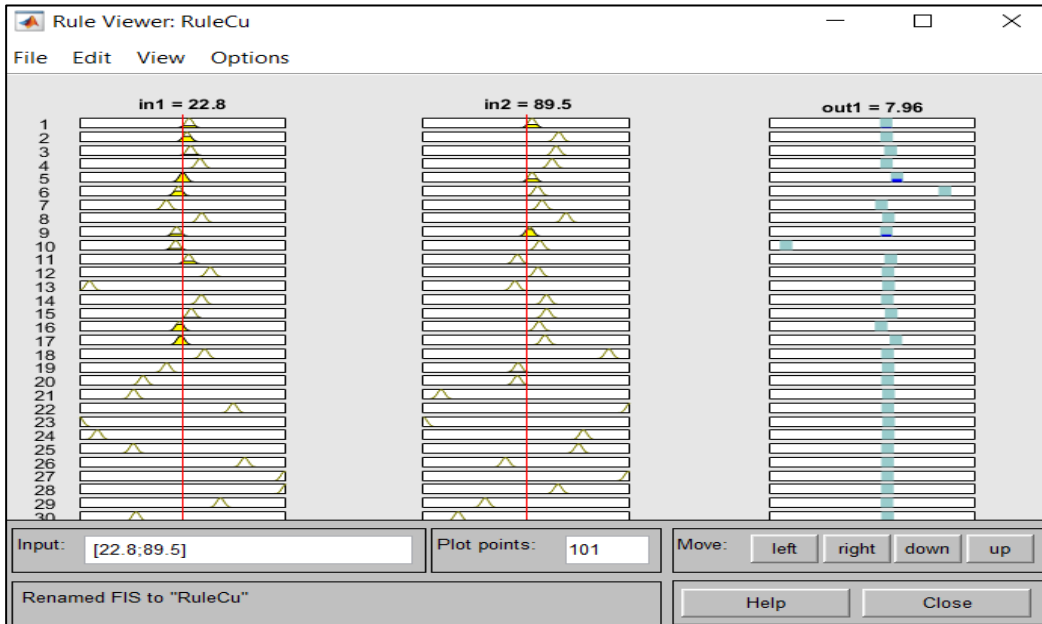
Figure A-1: ANFIS rule viewer for Cu (number of fuzzy rules: 59).
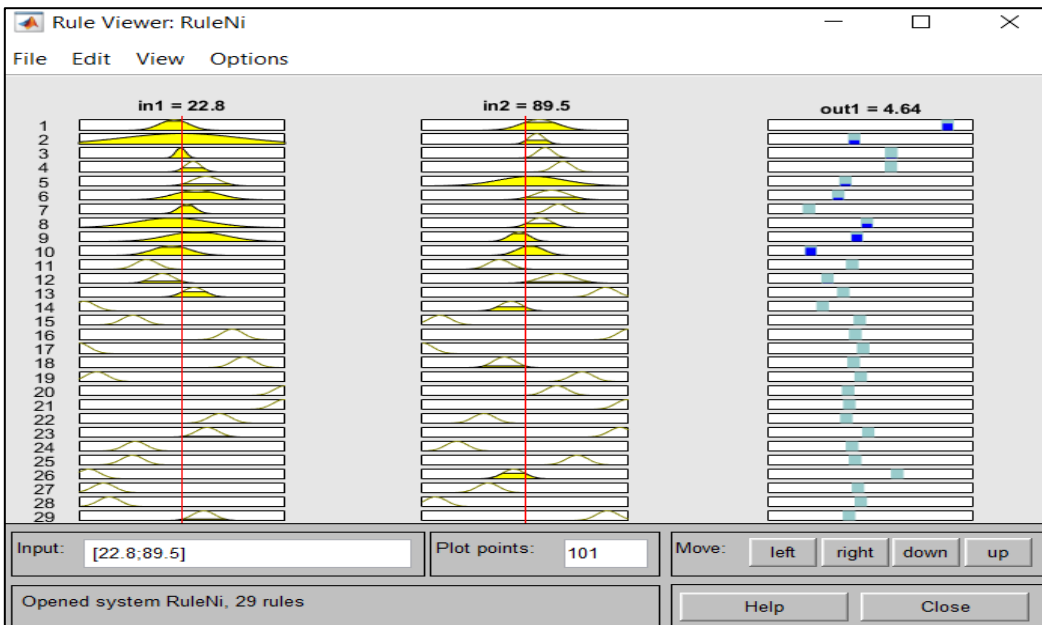
Figure A-2: ANFIS rule viewer for Ni (number of fuzzy rules: 29).
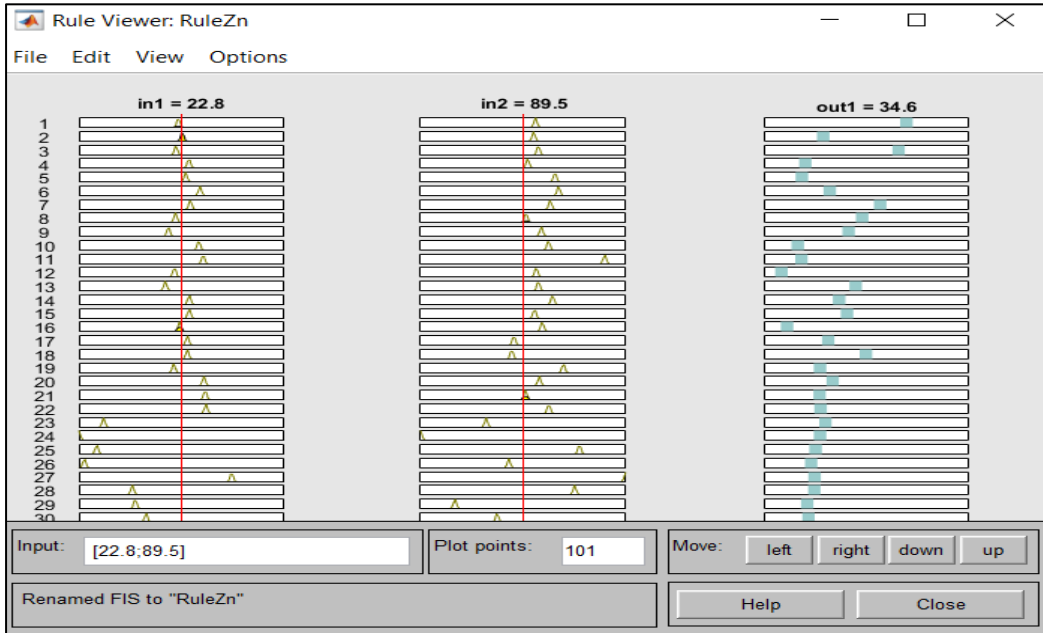
Figure A-3: ANFIS rule viewer for Zn (number of fuzzy rules: 68).
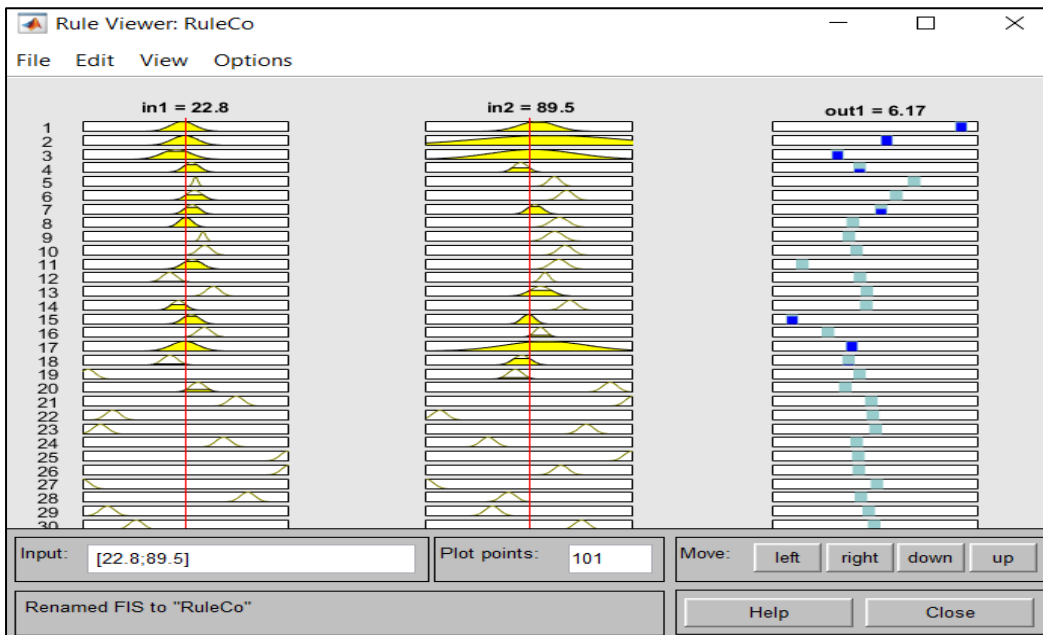
Figure A-4: ANFIS rule viewer for Co (number of fuzzy rules: 38).
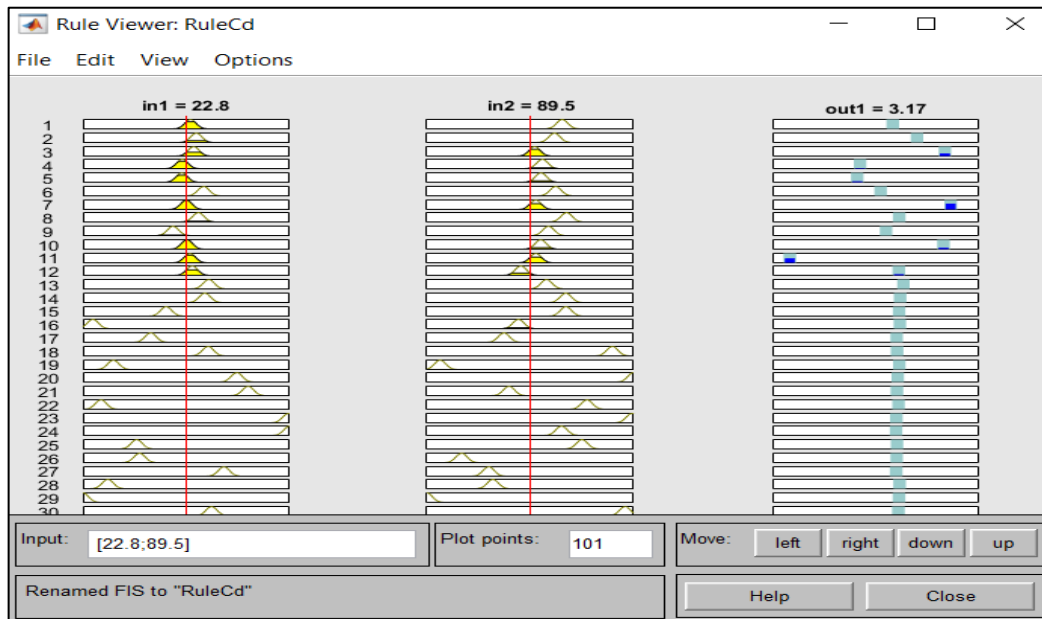
Figure A-5: ANFIS rule viewer for Cd (number of fuzzy rules: 48).
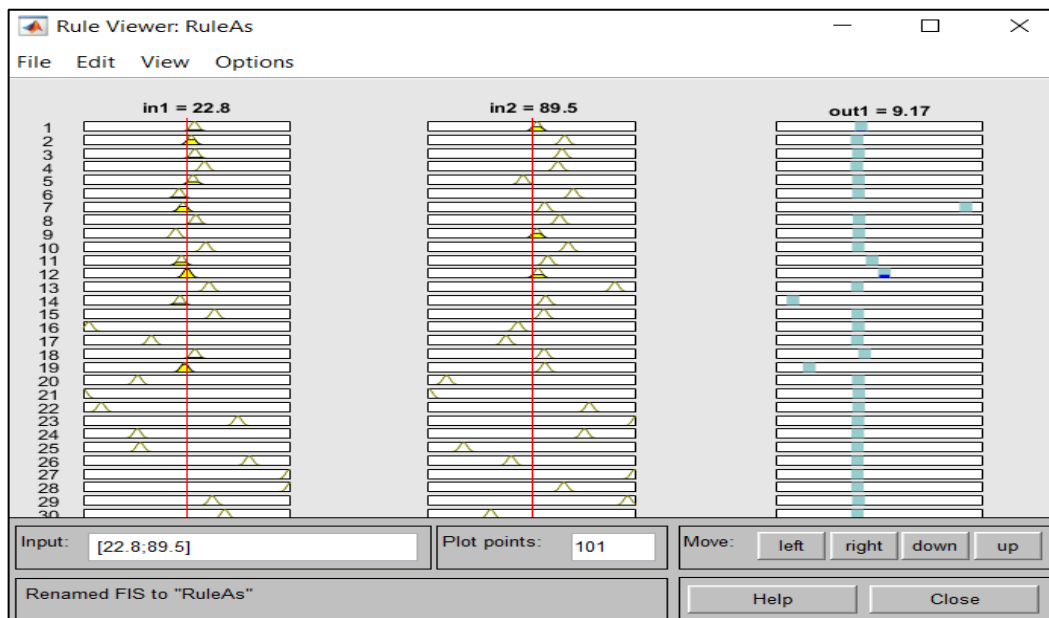
Figure A-6: ANFIS rule viewer for As (number of fuzzy rules: 56).
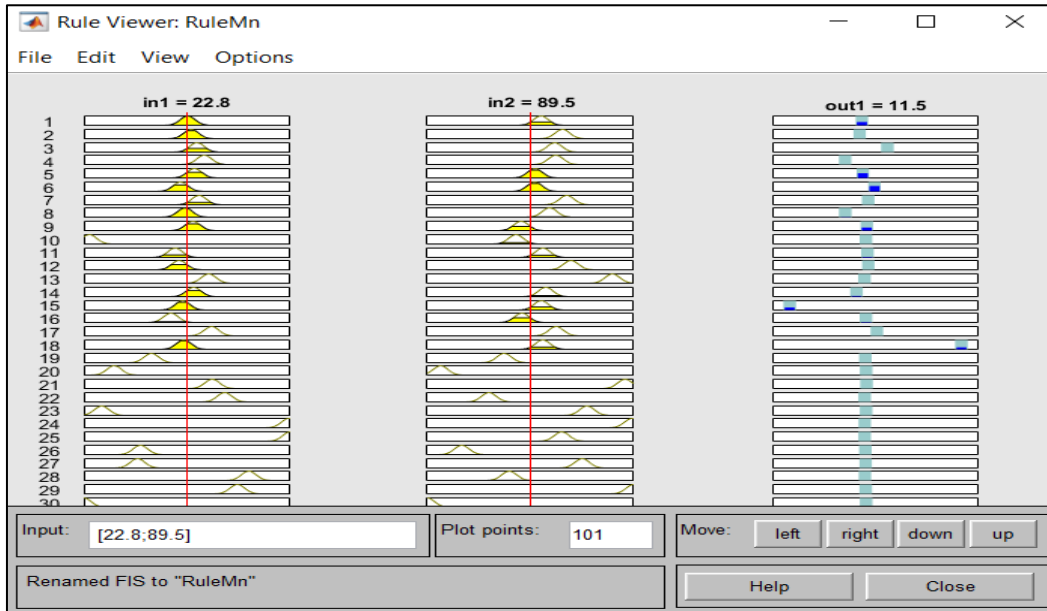
Figure A-7: ANFIS rule viewer for Mn (number of fuzzy rules: 37).
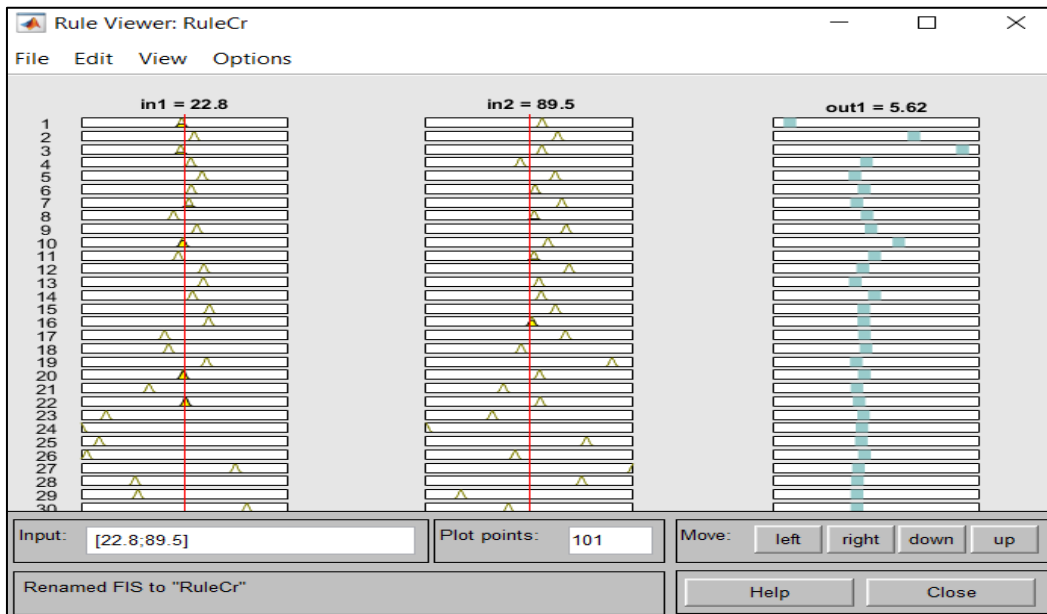
Figure A-8: ANFIS rule viewer for Cr (number of fuzzy rules: 66).
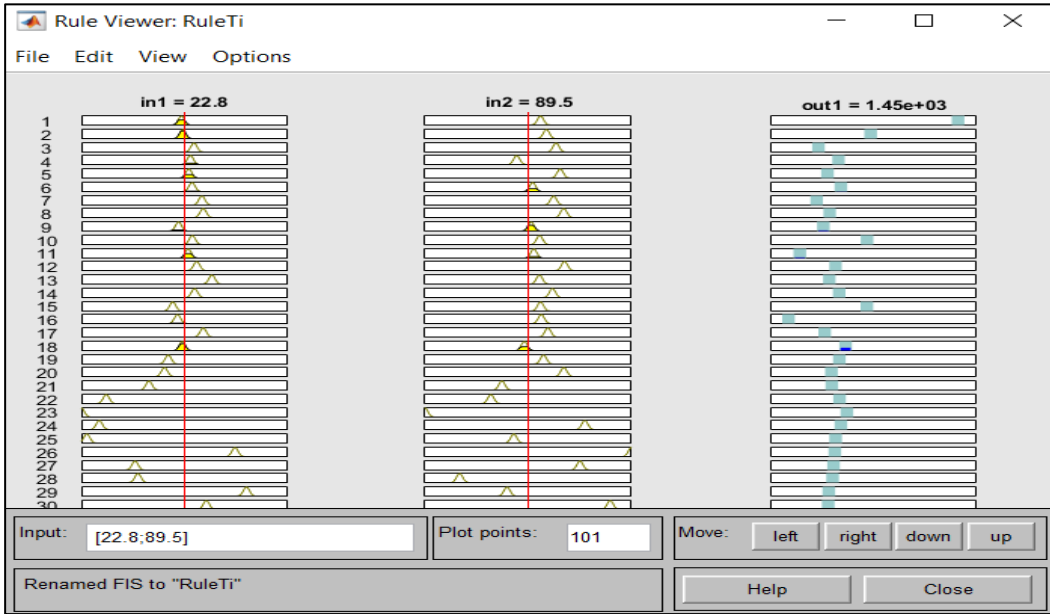
Figure A-9: ANFIS rule viewer for Ti (number of fuzzy rules: 61).
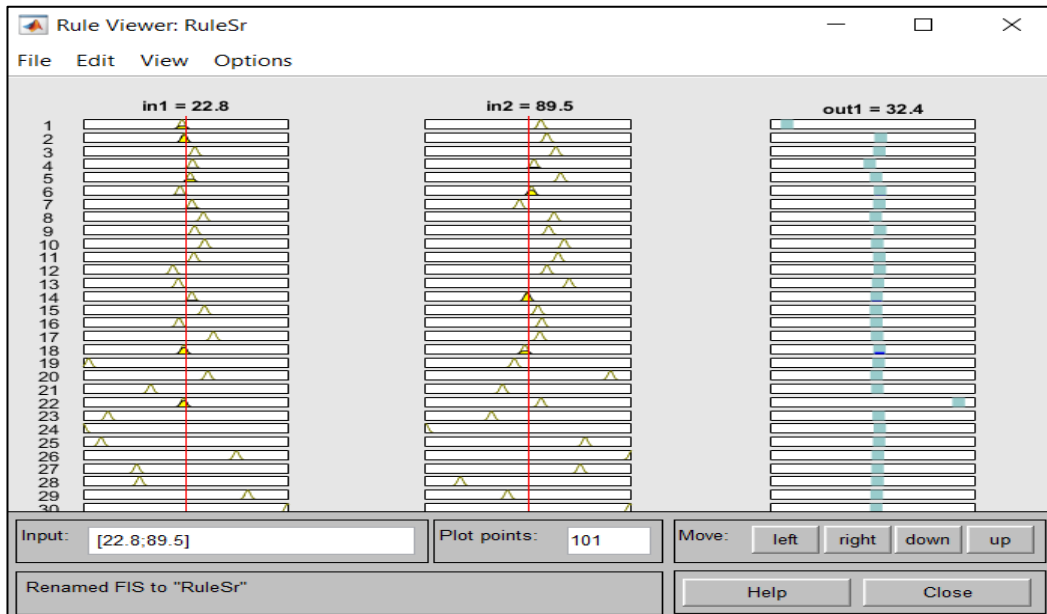
Figure A-10: ANFIS rule viewer for Sr (number of fuzzy rules: 64).
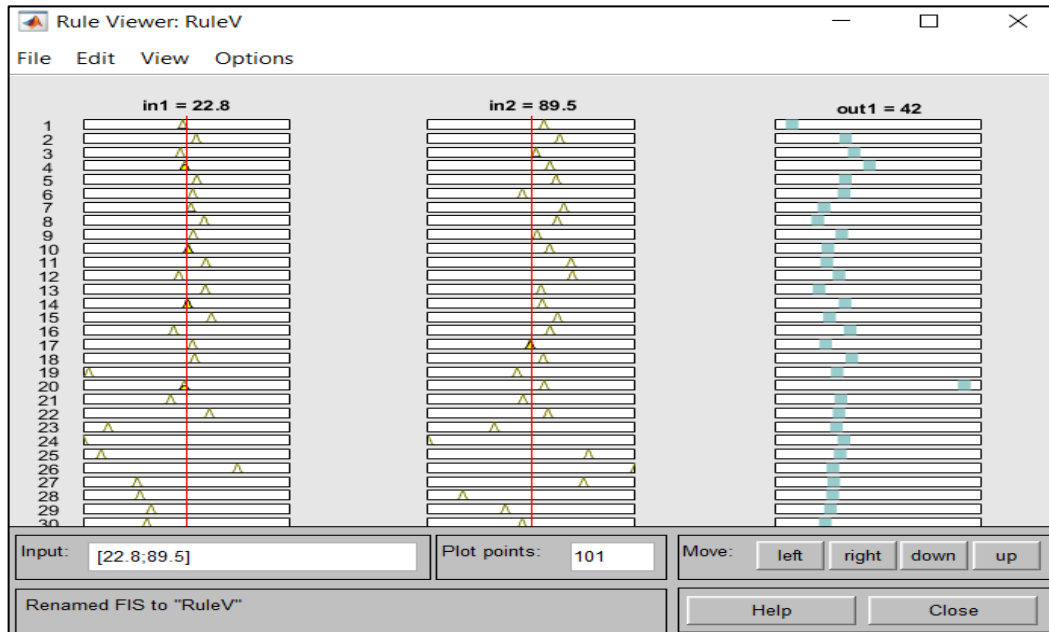
Figure A-11: ANFIS rule viewer for V (number of fuzzy rules: 66).
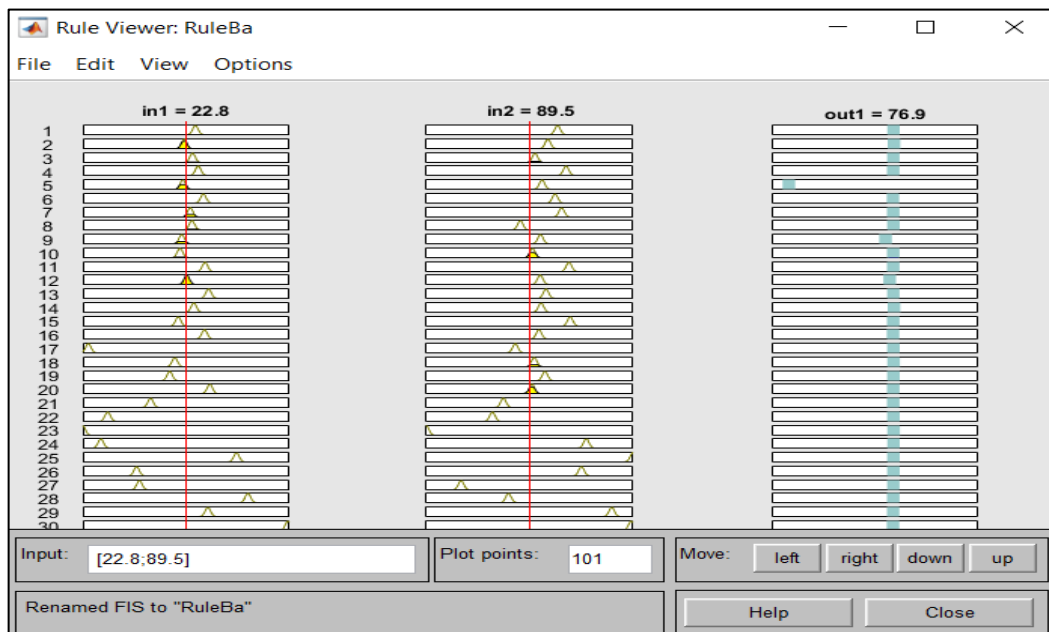
Figure A-12: ANFIS rule viewer for Ba (Number of fuzzy rules: 66).
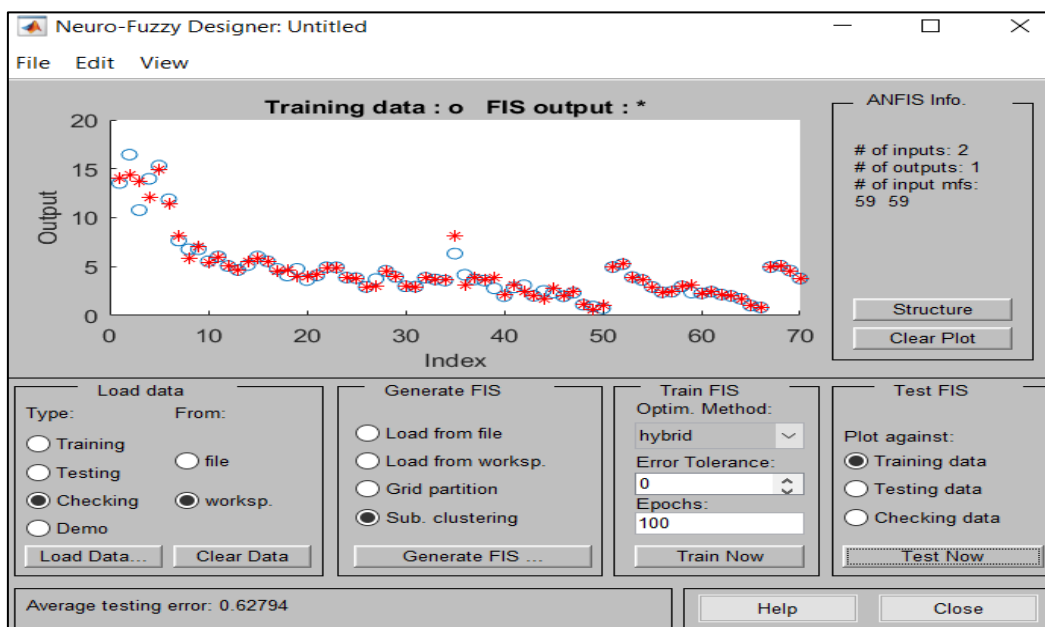
Figure A-13: FIS output of training data for Cu.
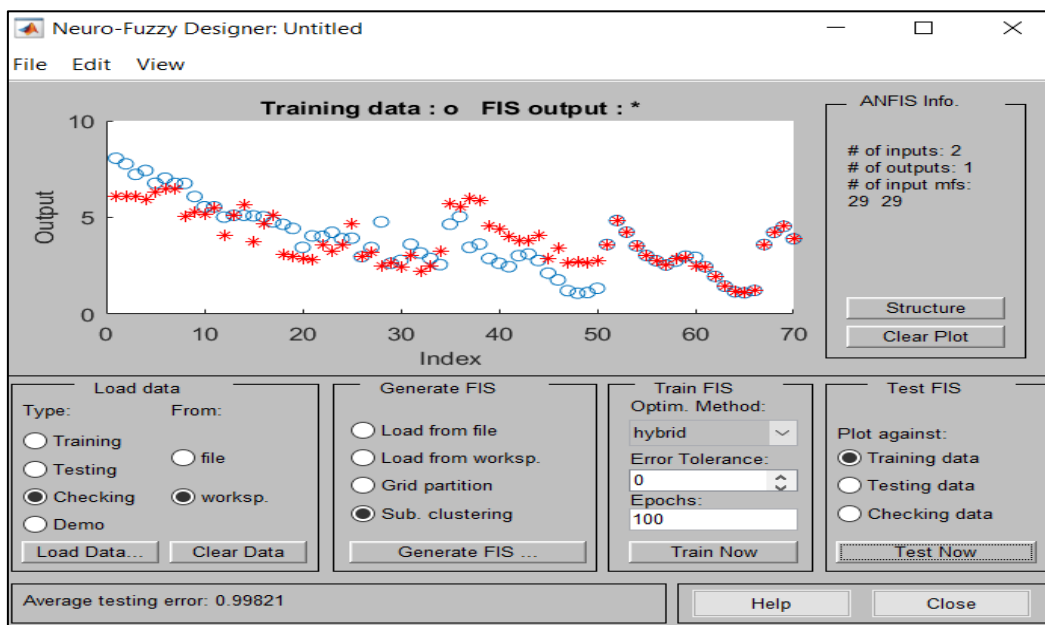
Figure A-14: FIS output of training data for Ni.
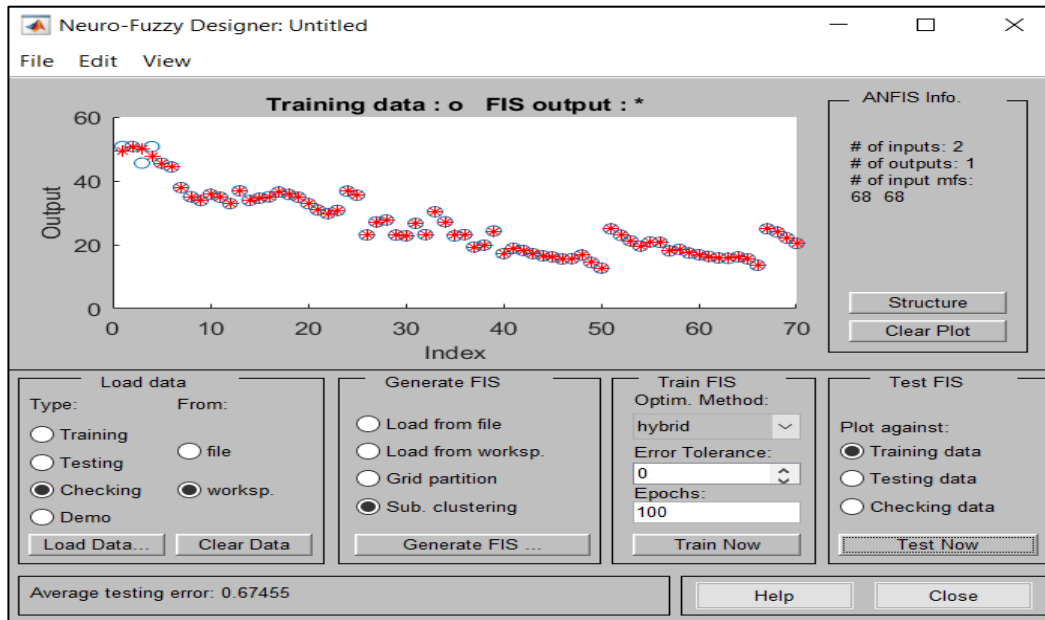
Figure A-15: FIS output of training data for Zn.
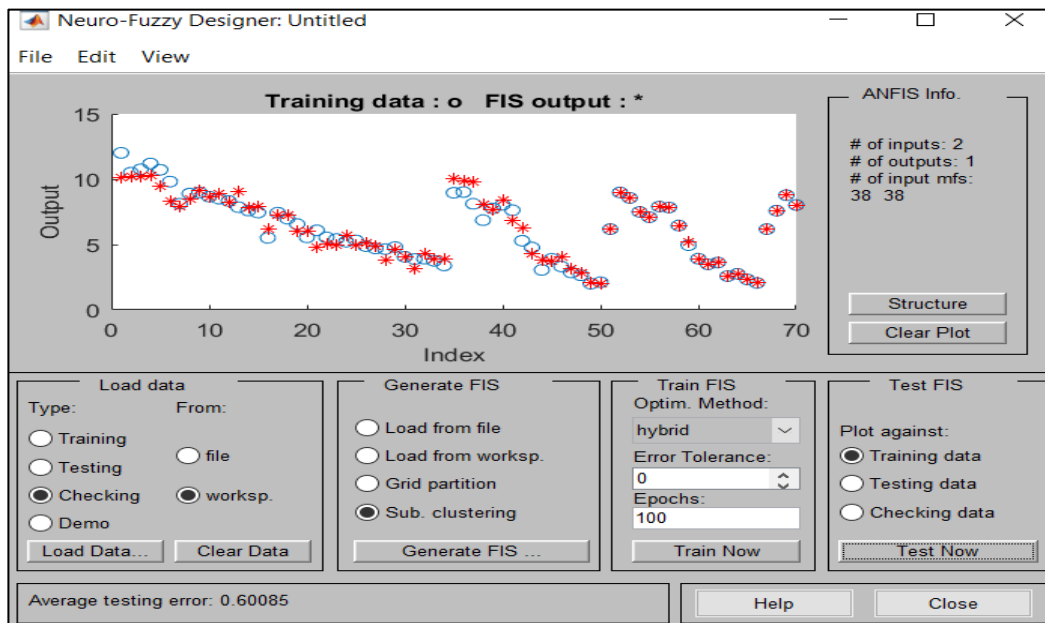
Figure A-16: FIS output of training data for Co.
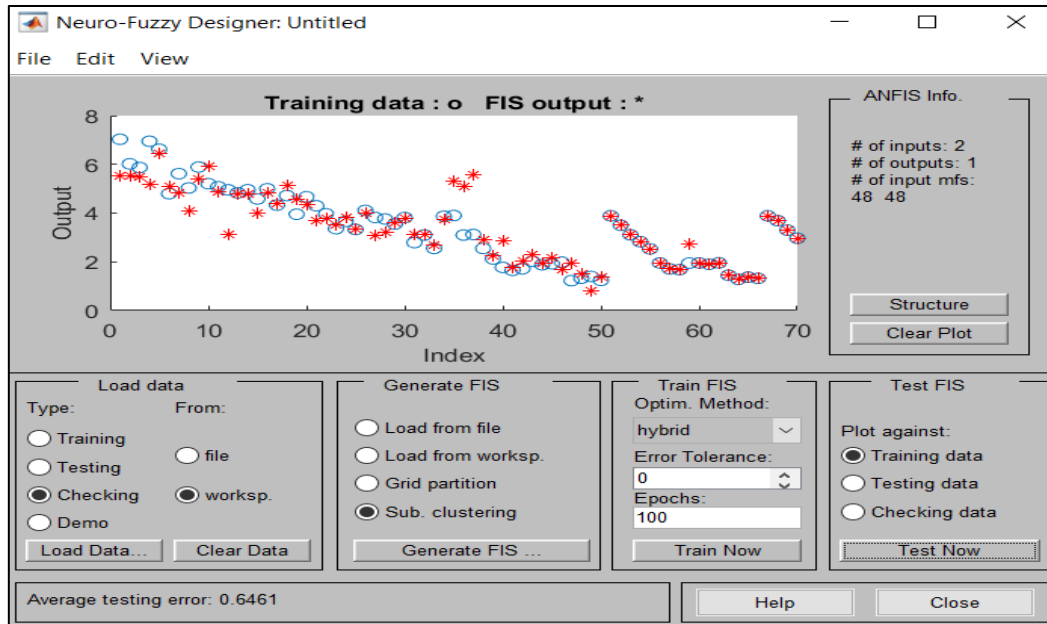
Figure A-17: FIS output of training data for Cd.
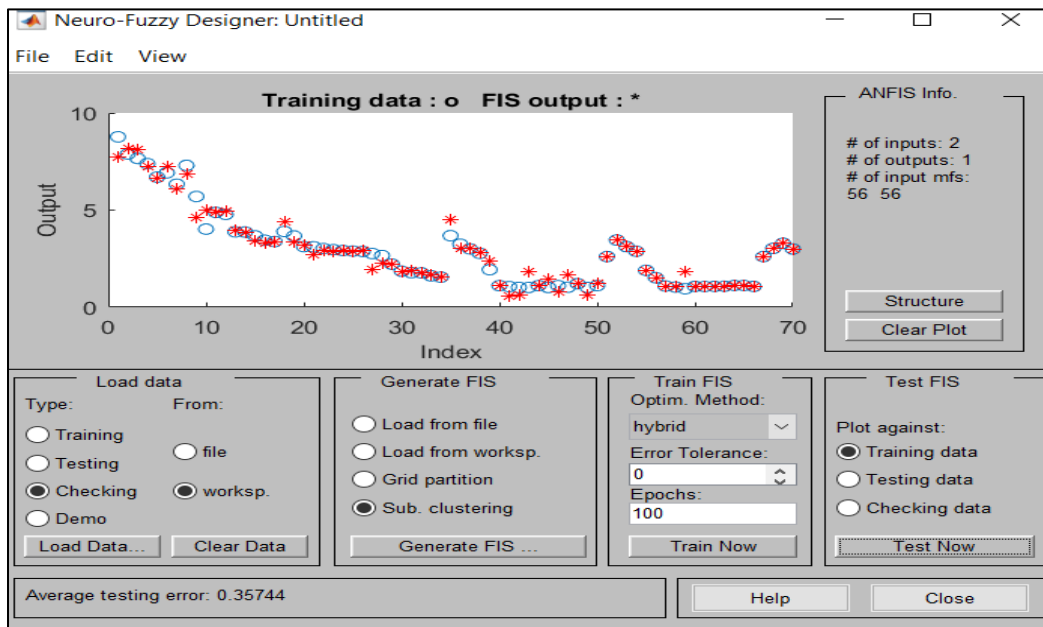
Figure A-18: FIS output of training data for As.
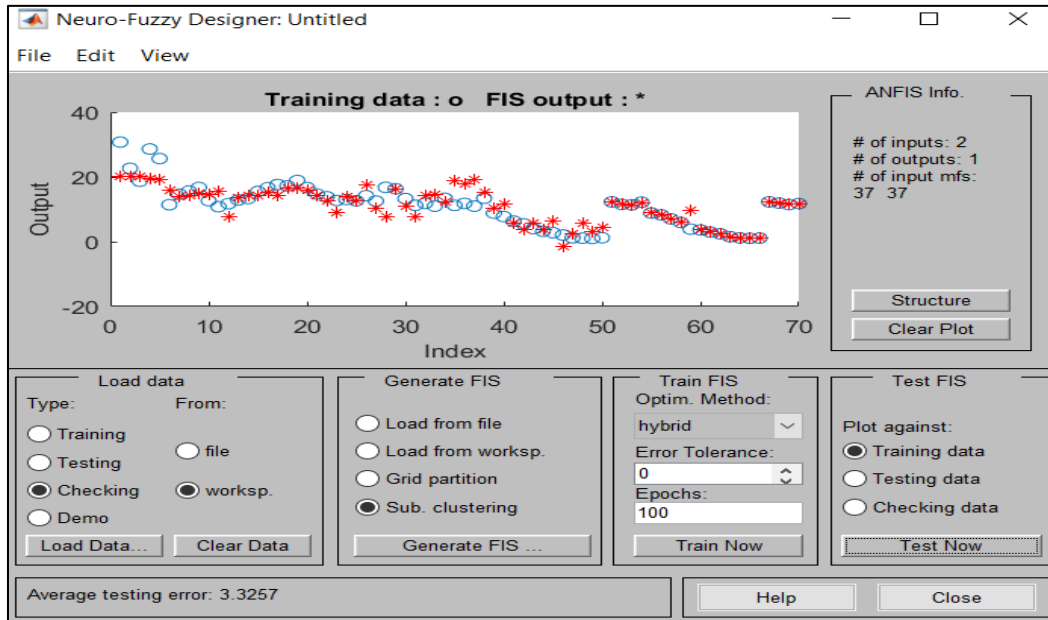
Figure A-19: FIS output of training data for Mn.
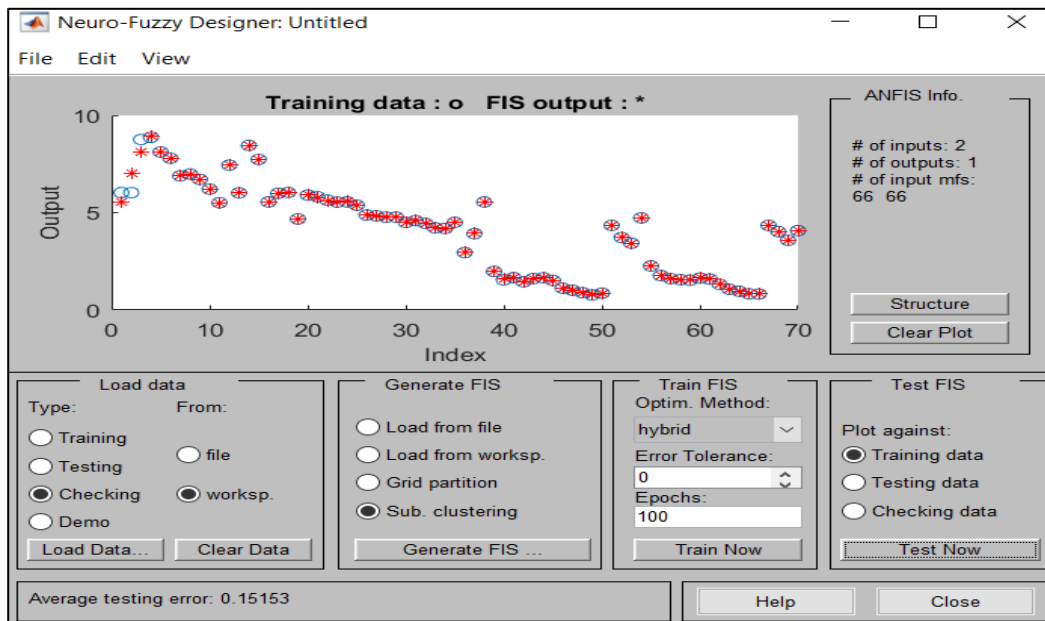
Figure A-20: FIS output of training data for Cr.
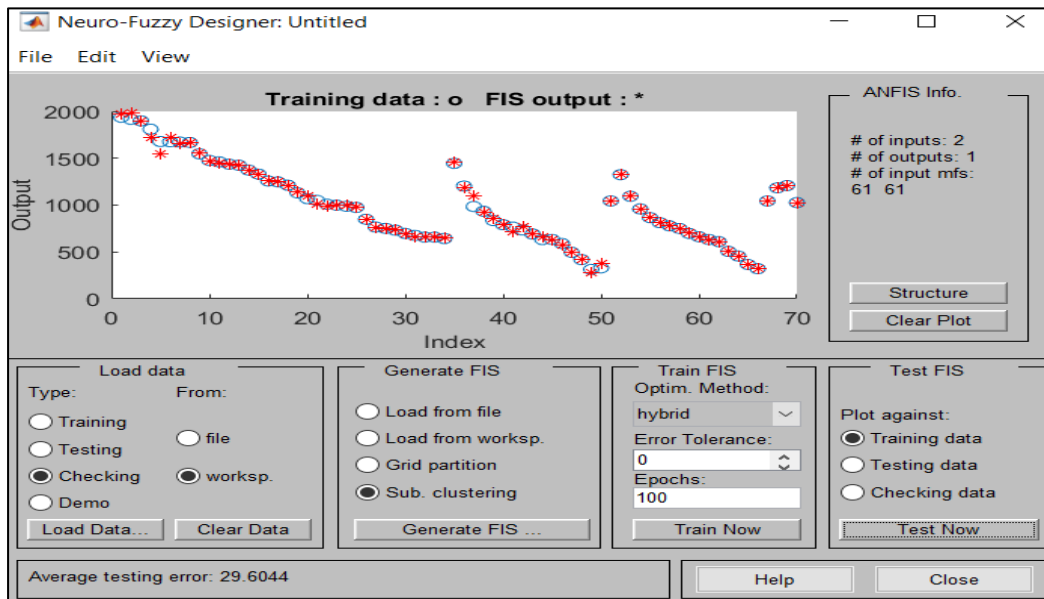
Figure A-21: FIS output of training data for Ti.
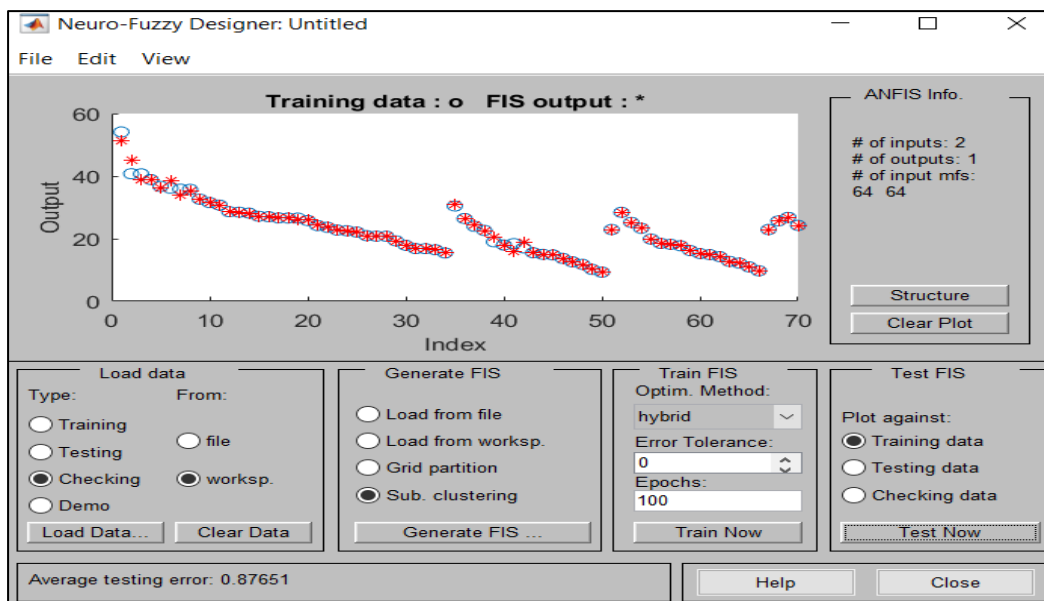
Figure A-22: FIS output of training data for Sr.
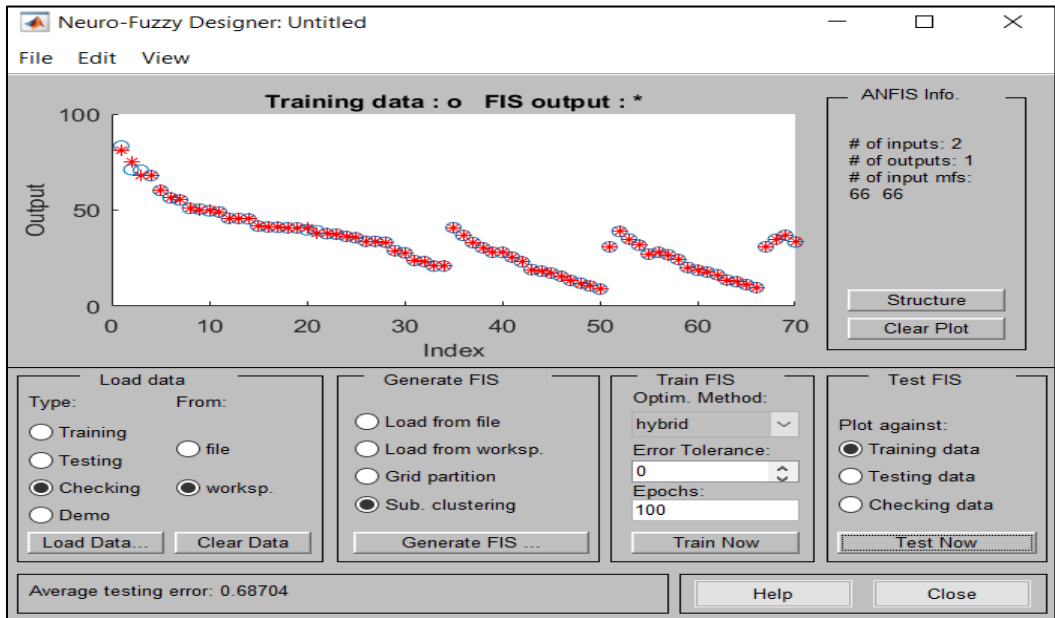
Figure A-23: FIS output of training data for V.
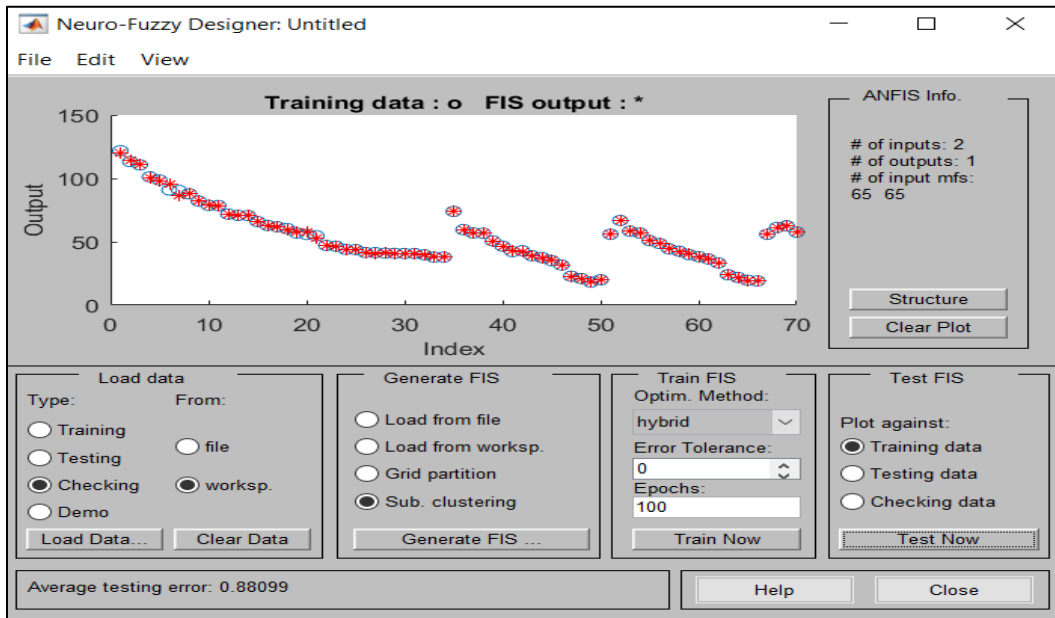
Figure A-24: FIS output of training data for Ba.
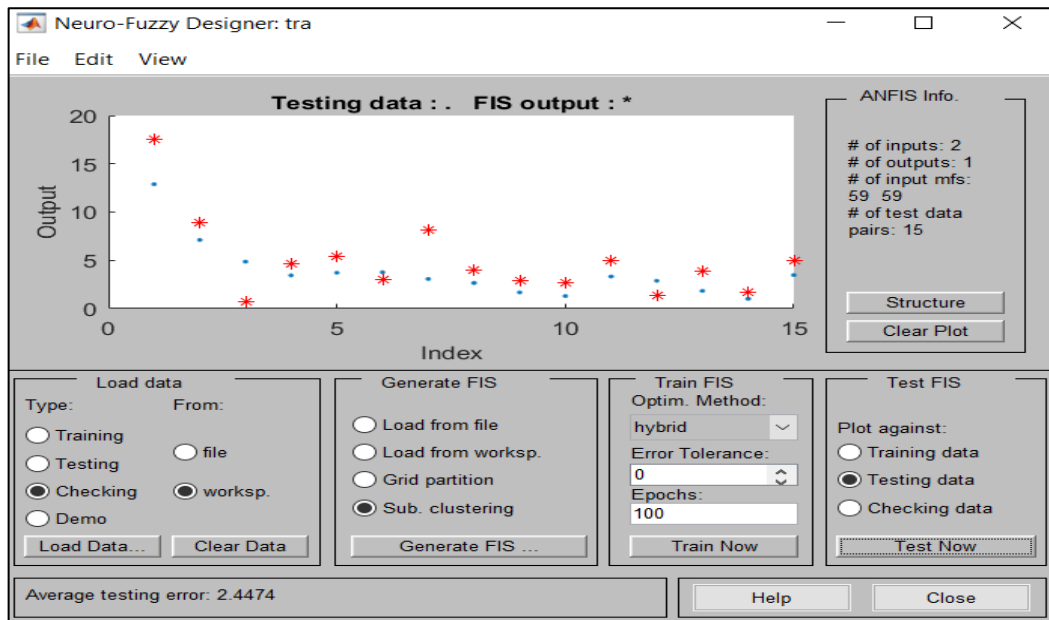
Figure A-25: FIS output of testing data for Cu.
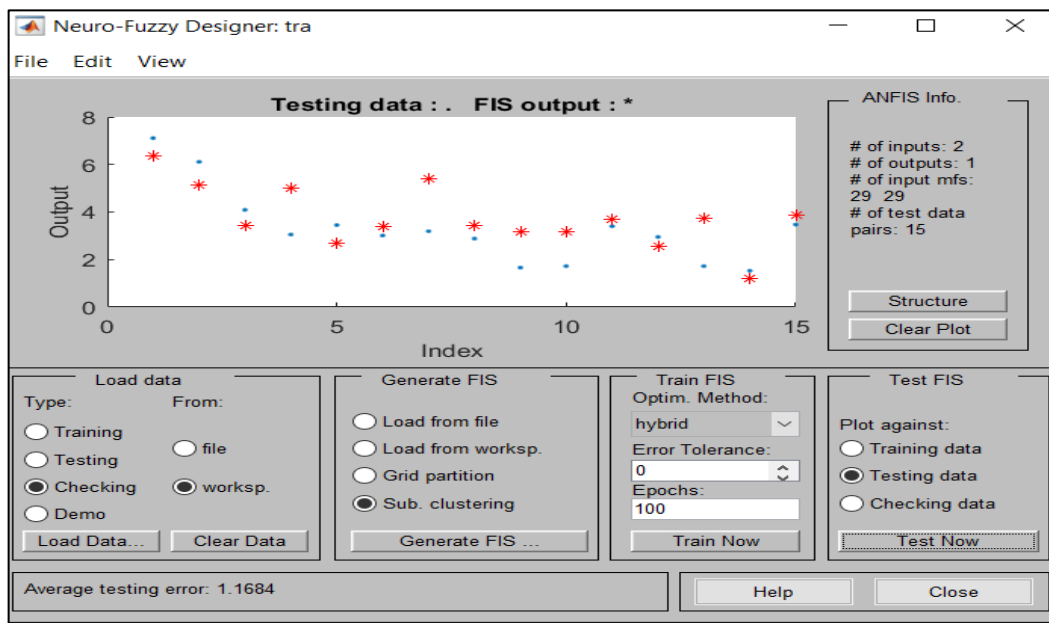
Figure A-26: FIS output of testing data for Ni.
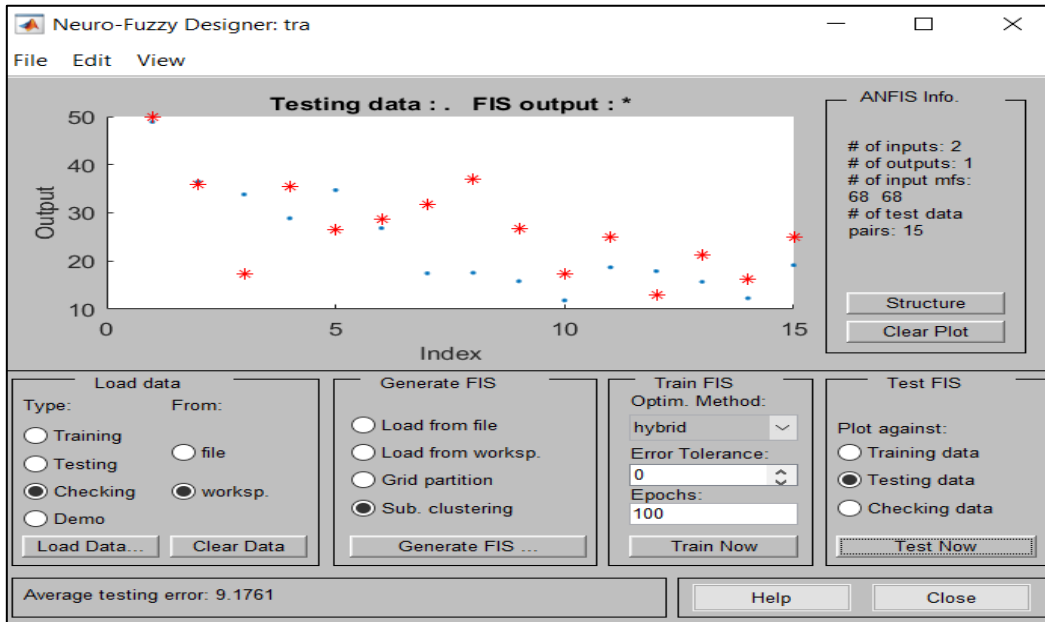
Figure A-27: FIS output of testing data for Zn.
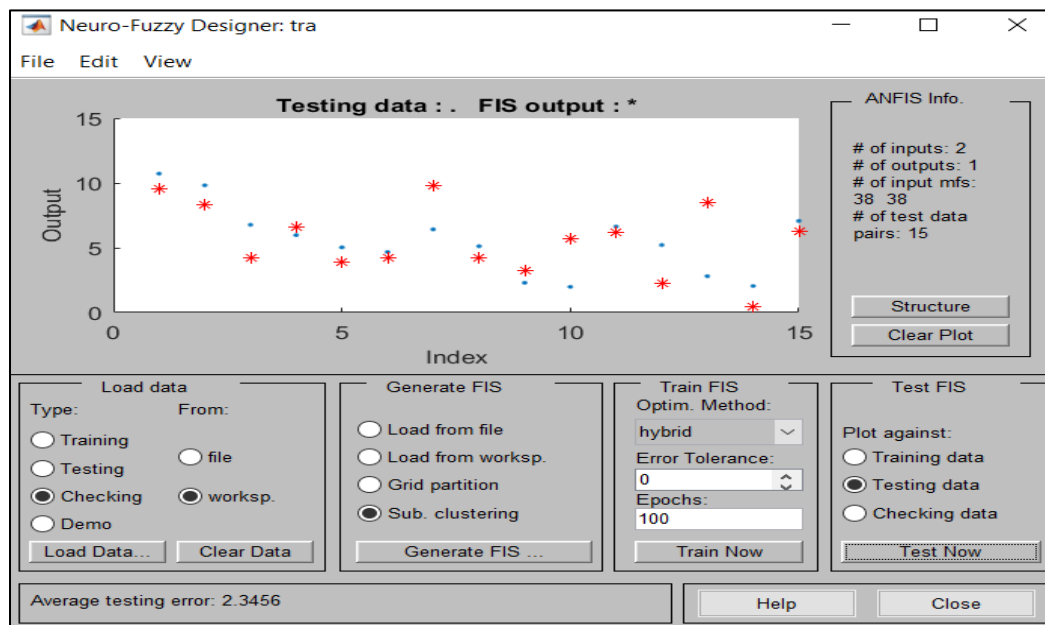
Figure A-28: FIS output of testing data for Co.
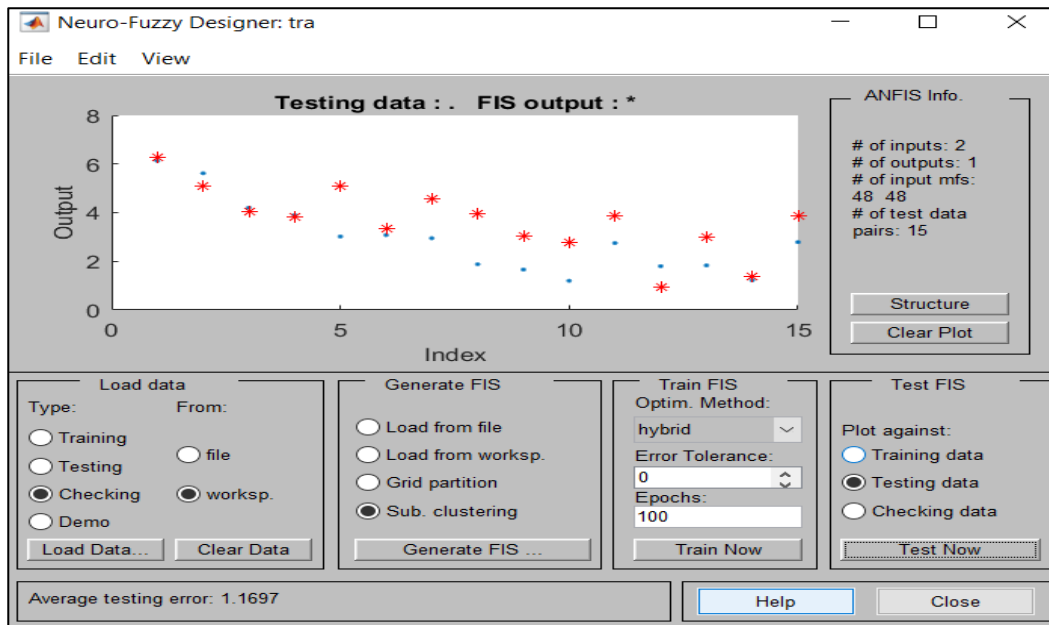
Figure A-29: FIS output of testing data for Cd.
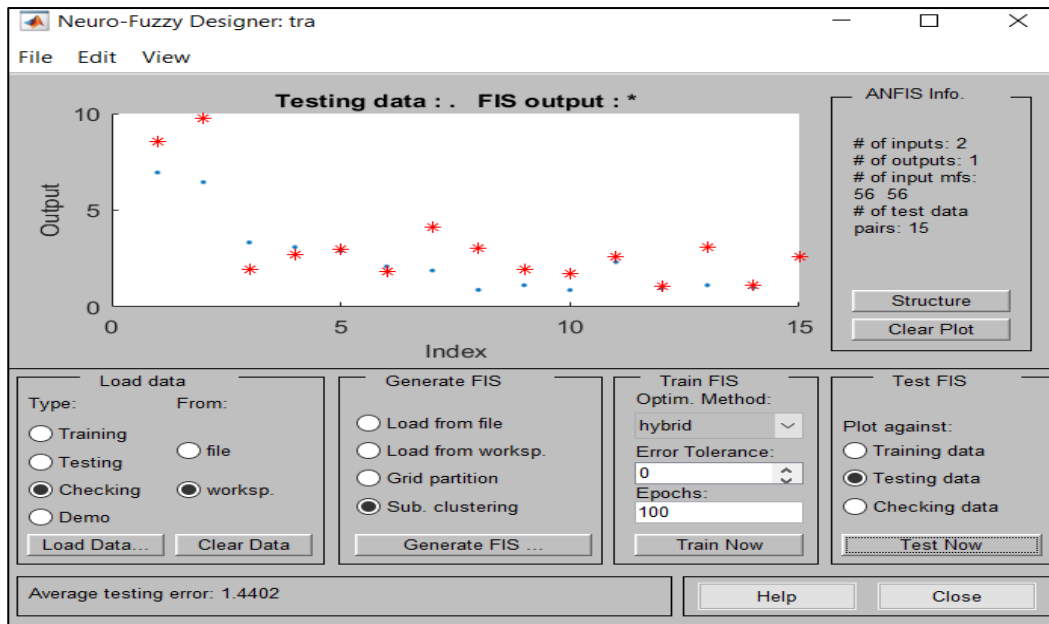
Figure A-30: FIS output of testing data for As.
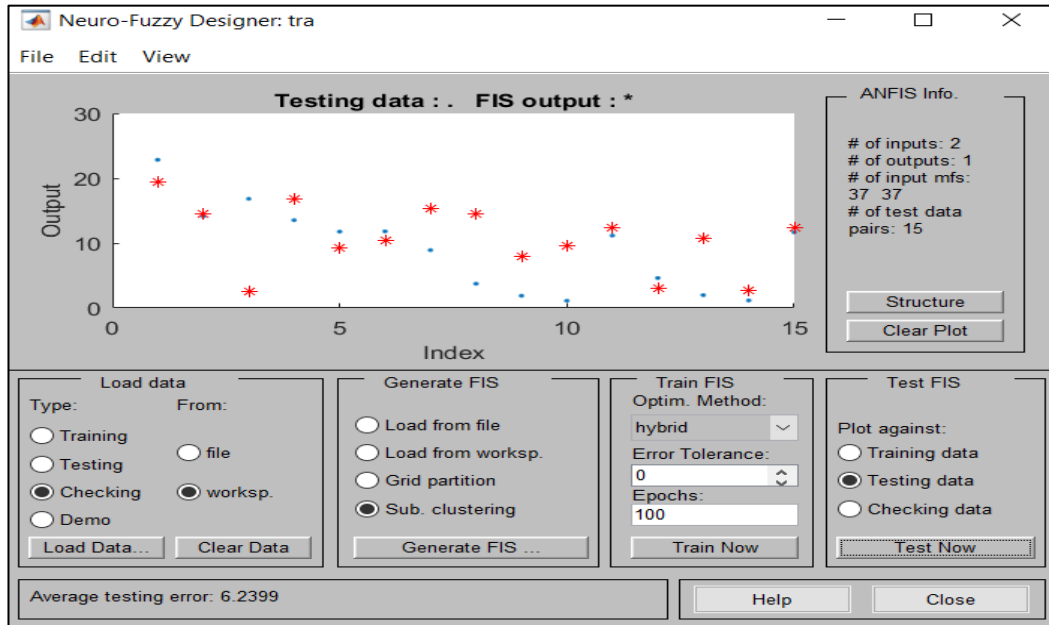
Figure A-31: FIS output of testing data for Mn.
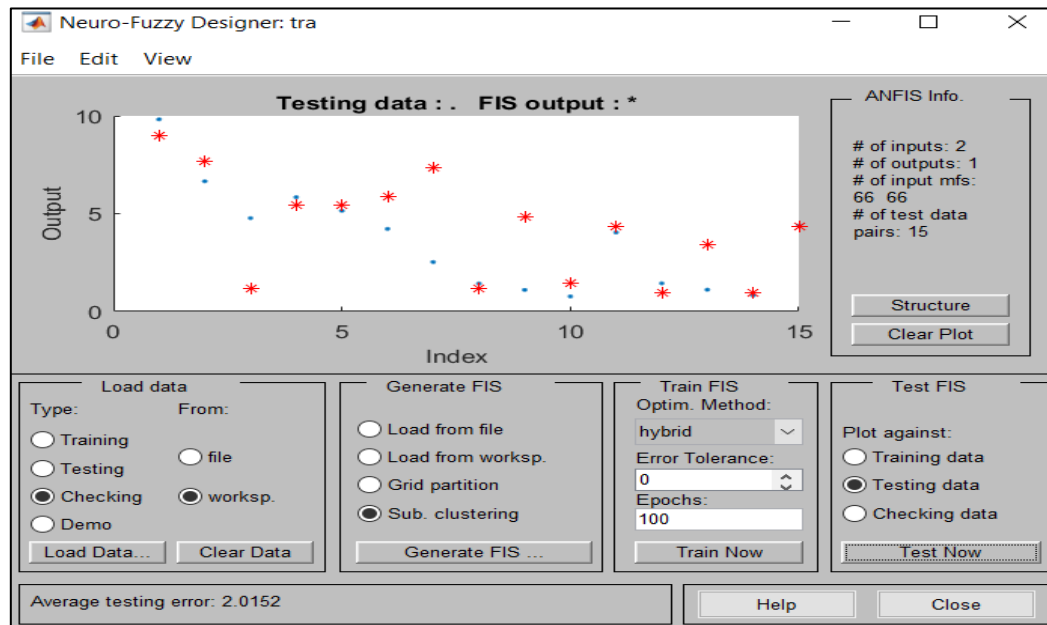
Figure A-32: FIS output of testing data for Cr.
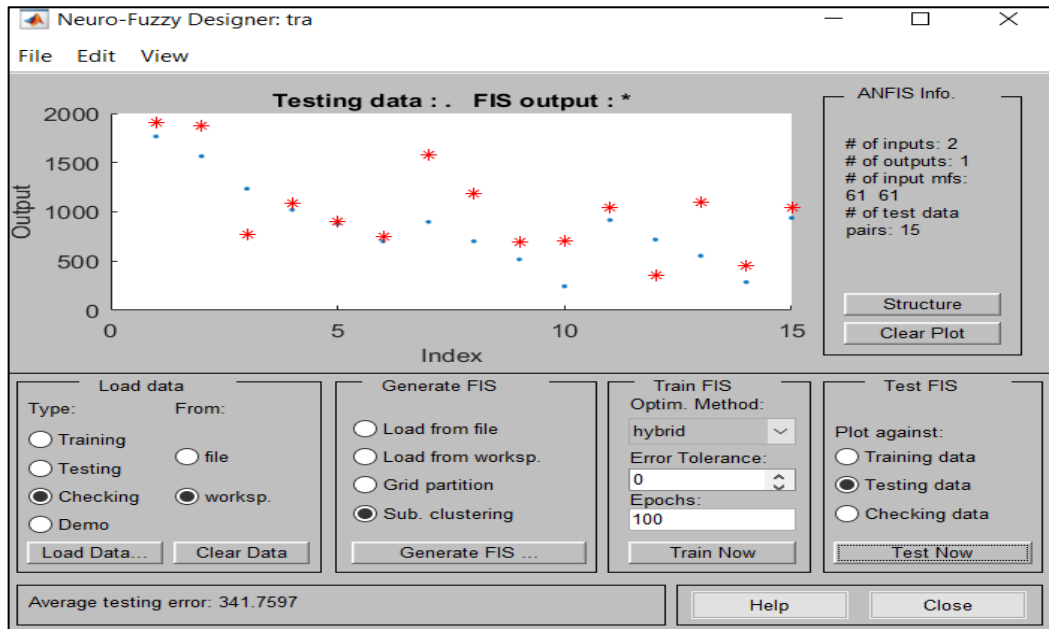
Figure A-33: FIS output of testing data for Ti.
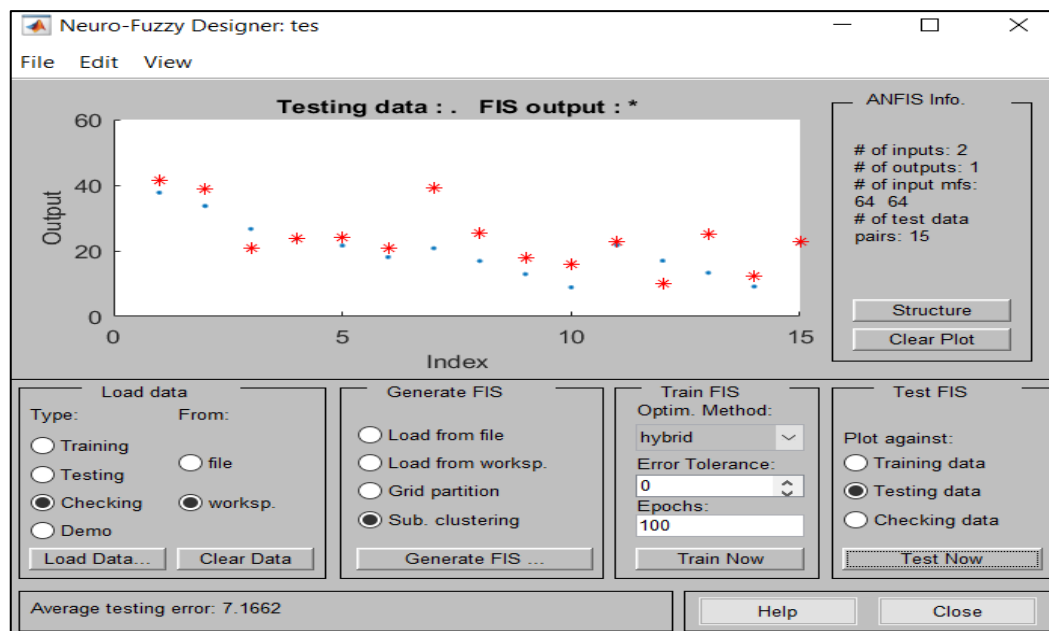
Figure A-34: FIS output of testing data for Sr.
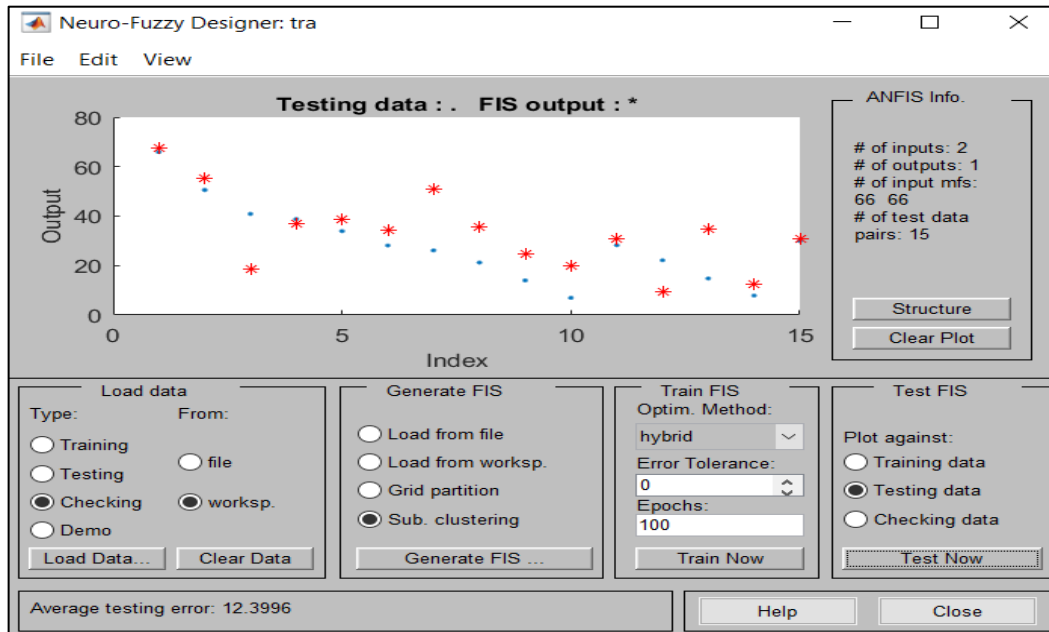
Figure A-35: FIS output of testing data for V.
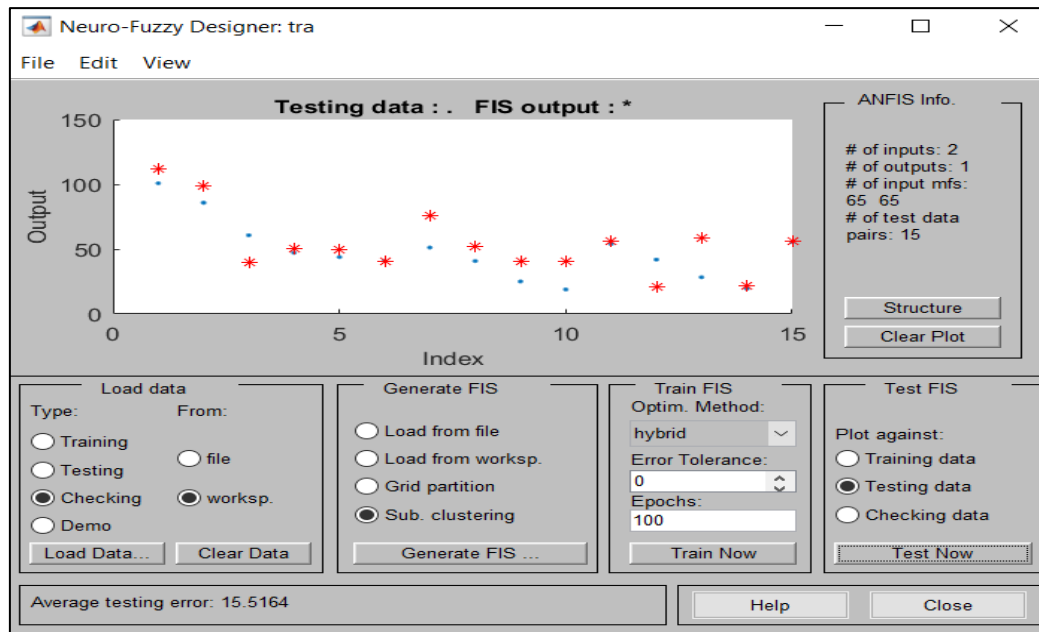
Figure A-36: FIS output of testing data for Ba.
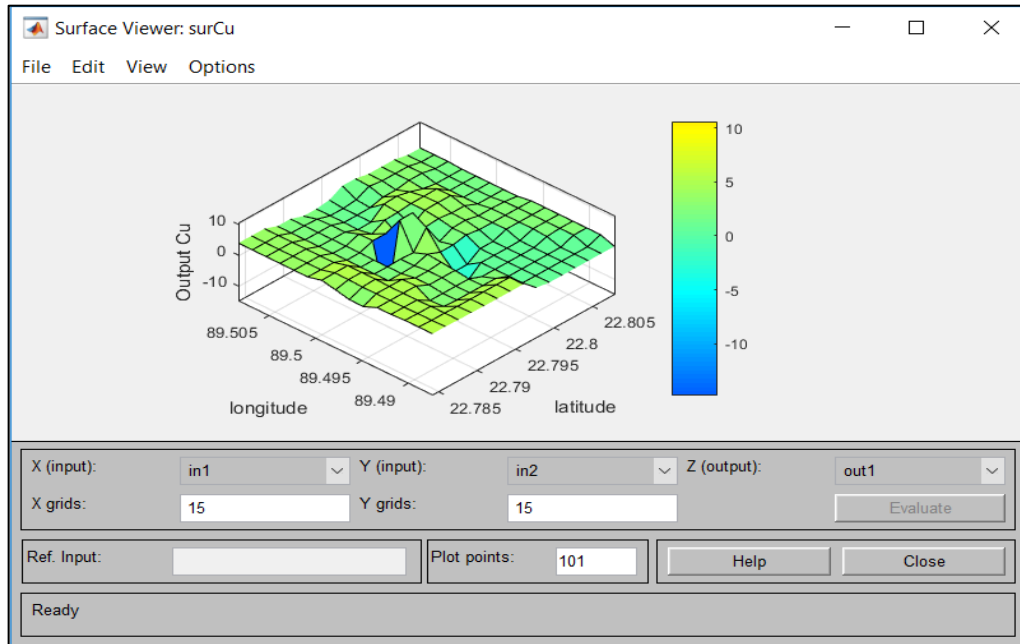
Figure A-37: Surface viewer of the outputs for Cu.
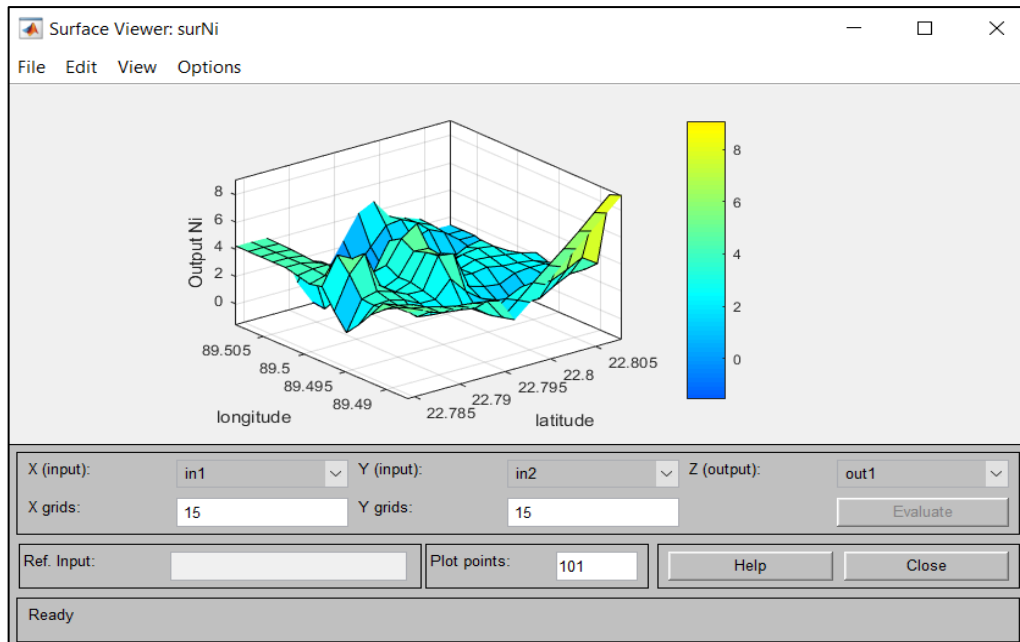
Figure A-38: Surface viewer of the outputs for Ni.

179
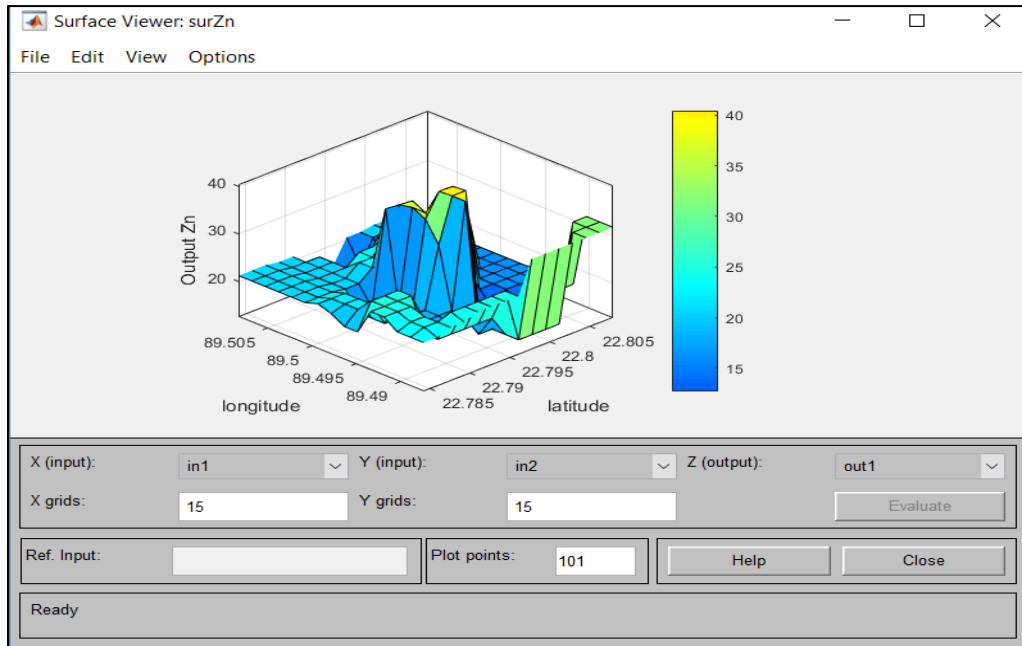
Figure A-39: Surface viewer of the outputs for Zn.
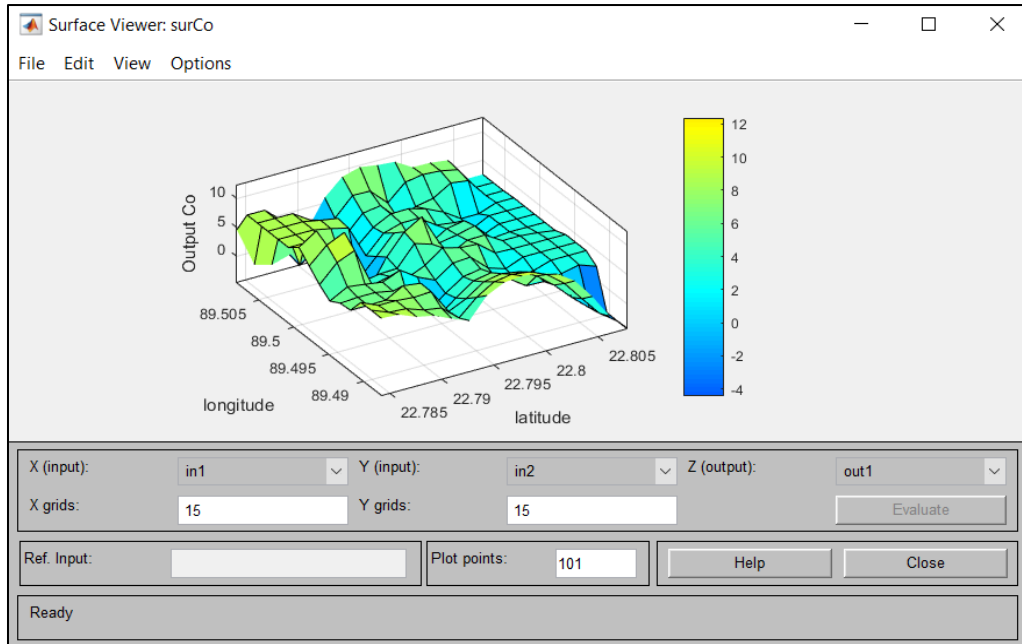
Figure A-40: Surface viewer of the outputs for Co.
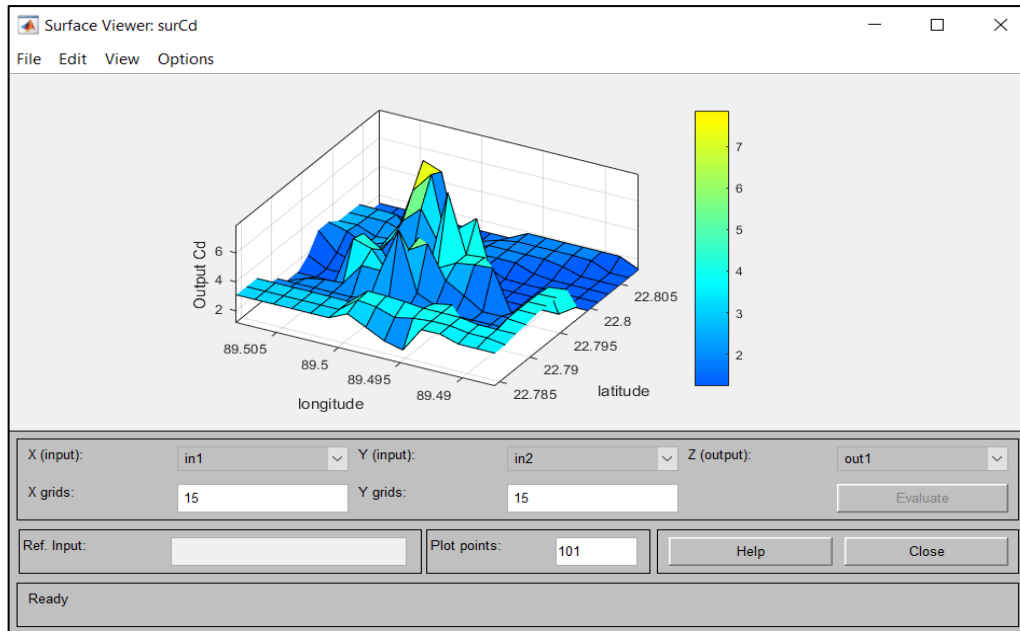
Figure A-41: Surface viewer of the outputs for Cd.
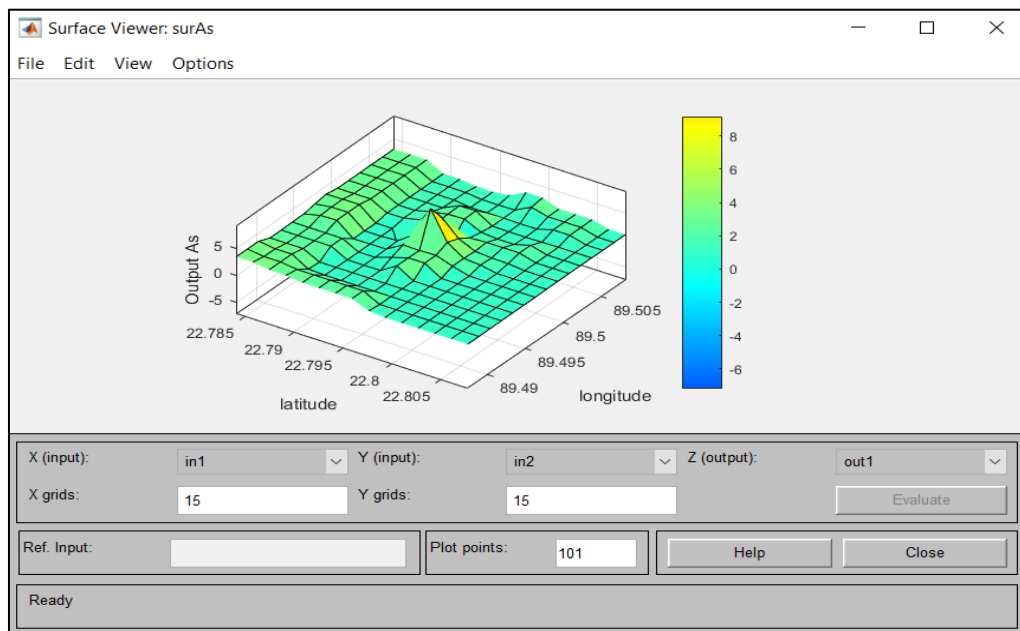
Figure A-42: Surface viewer of the outputs for As.
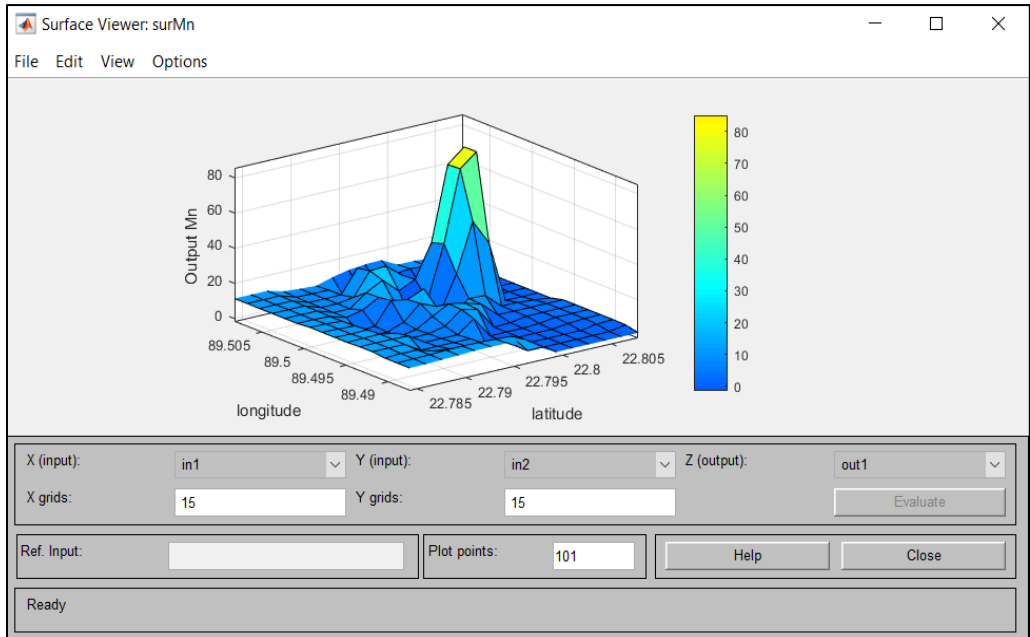
Figure A-43: Surface viewer of the outputs for Mn.
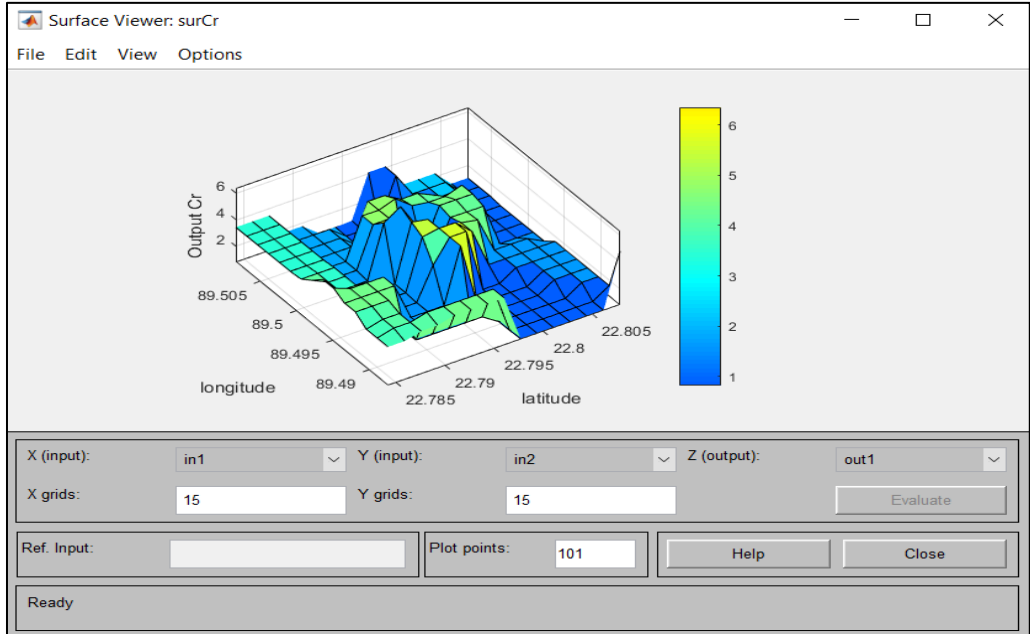
Figure A-44: Surface viewer of the outputs for Cr.
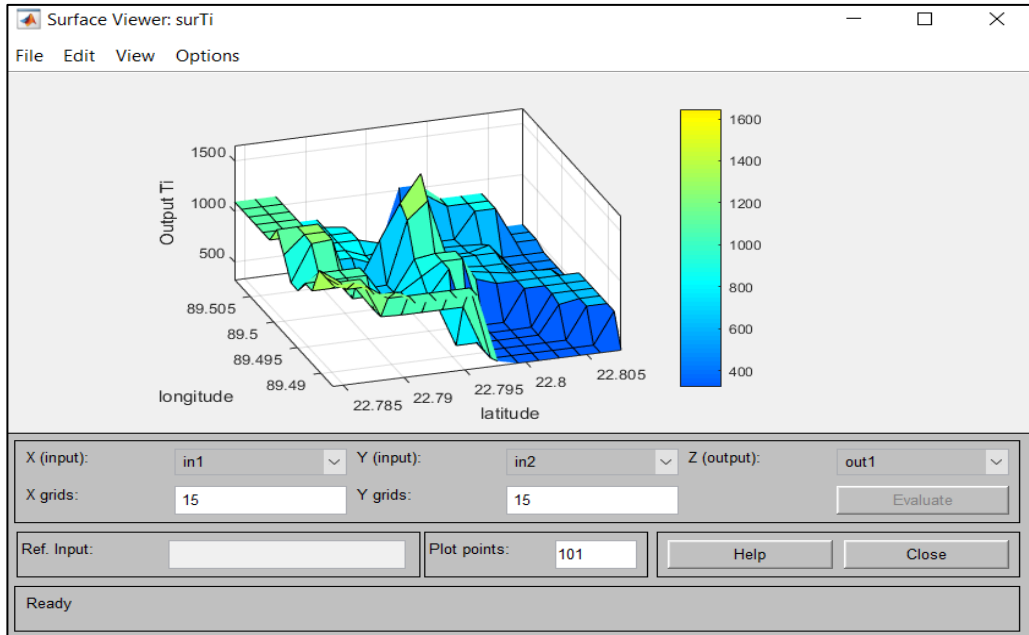
Figure A-45: Surface viewer of the outputs for Ti.
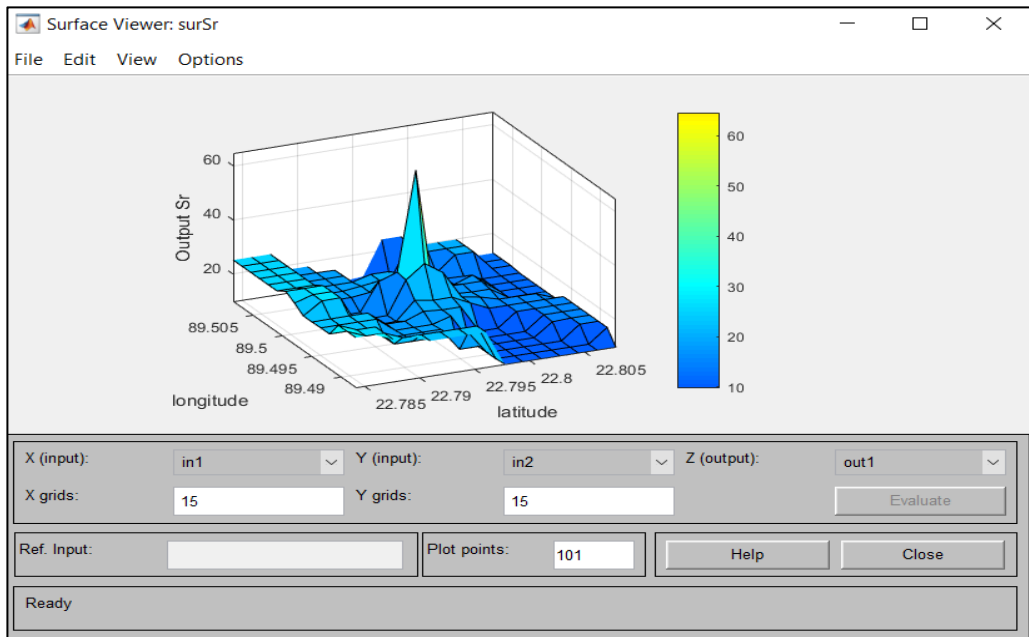
Figure A-46: Surface viewer of the outputs for Sr.
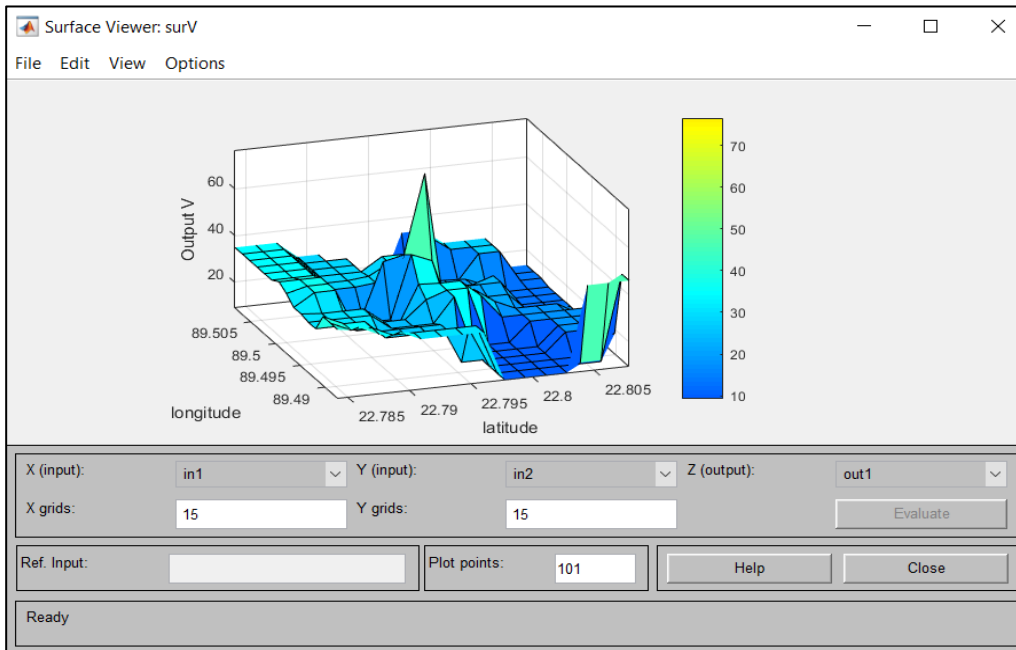
Figure A-47: Surface viewer of the outputs for V.
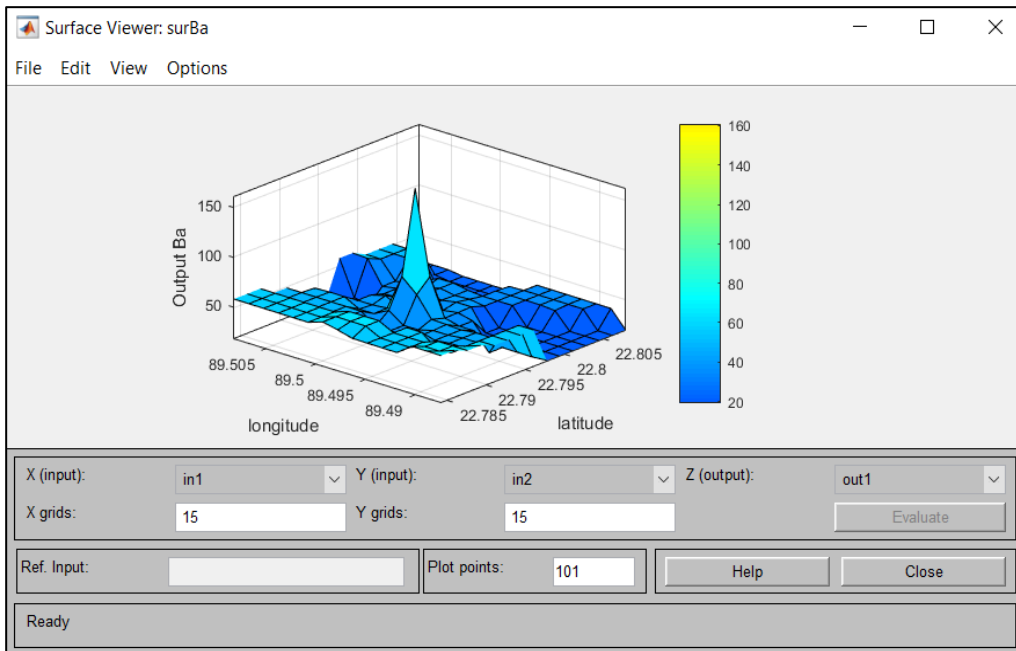
Figure A-48: Surface viewer of the outputs for Ba.
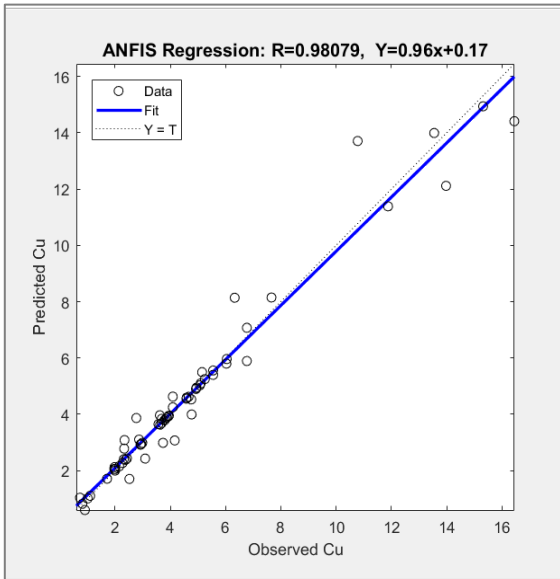
Figure A-49: Regression analysis for Cu in ANFIS (training).
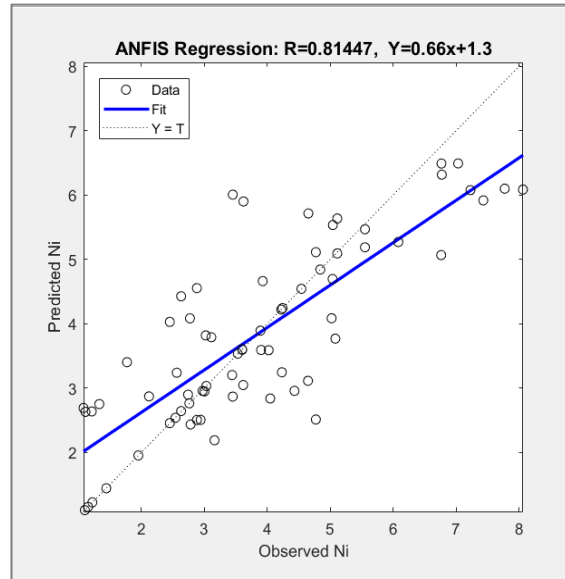
Figure A-50: Regression analysis for Ni in ANFIS (training).
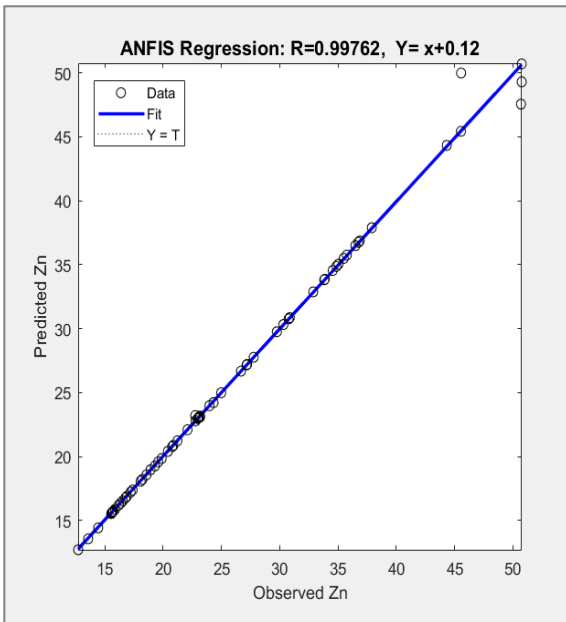
Figure A-51: Regression analysis for Zn in ANFIS (training).
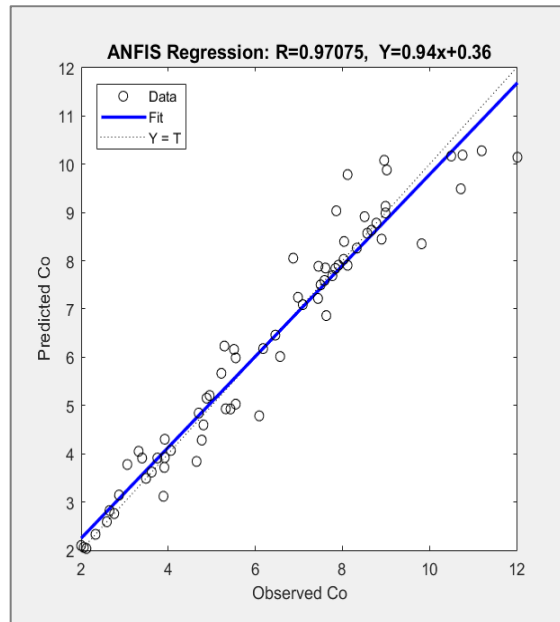
Figure A-52: Regression analysis for Co in ANFIS (training).
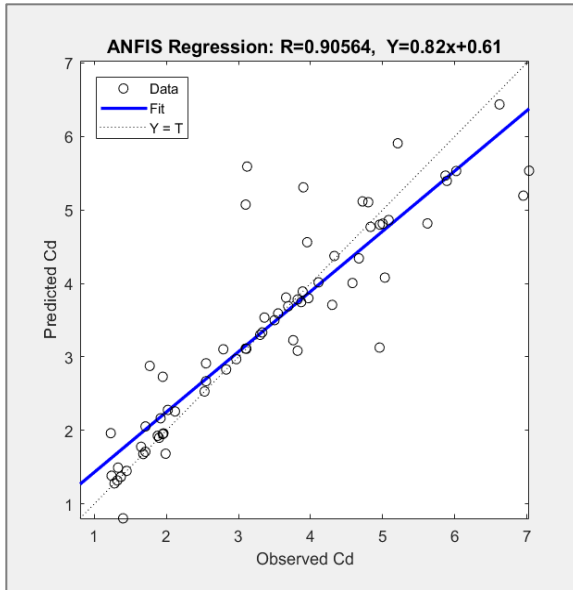
Figure A-53: Regression analysis for Cd in ANFIS (training).
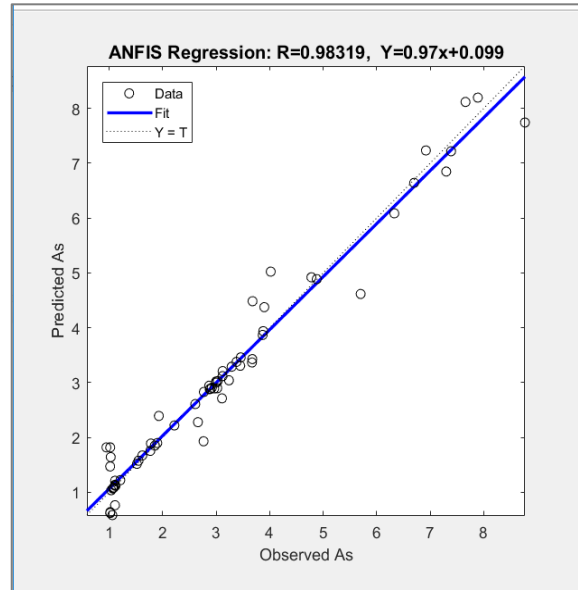
Figure A-54: Regression analysis for As in ANFIS (training).
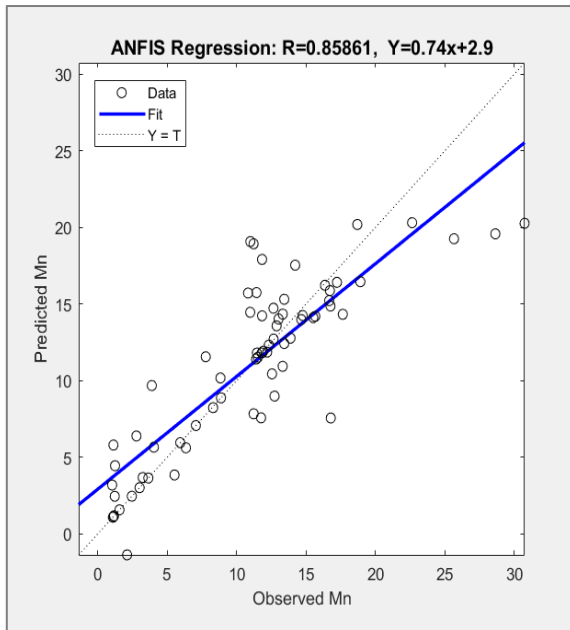
Figure A-55: Regression analysis for Mn in ANFIS (training).
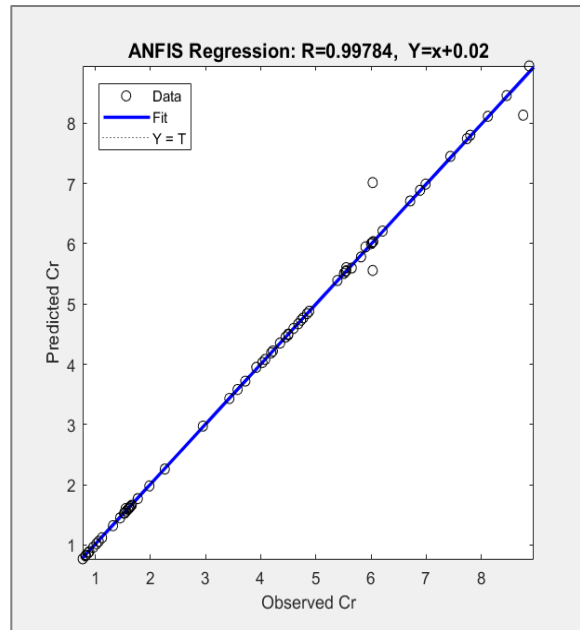
Figure A-56: Regression analysis for Cr in ANFIS (training).
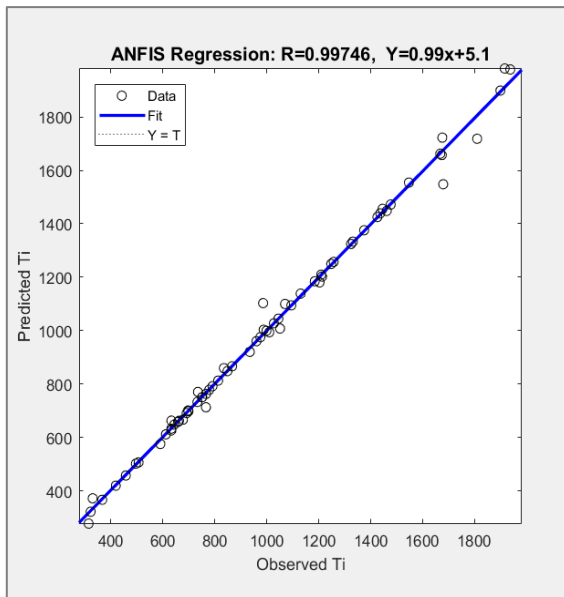
Figure A-57: Regression analysis for Ti in ANFIS (training).
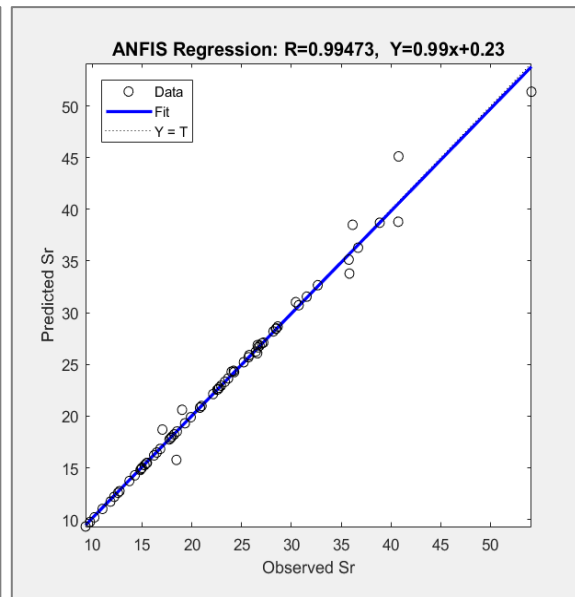
Figure A-58: Regression analysis for Sr in ANFIS (training).
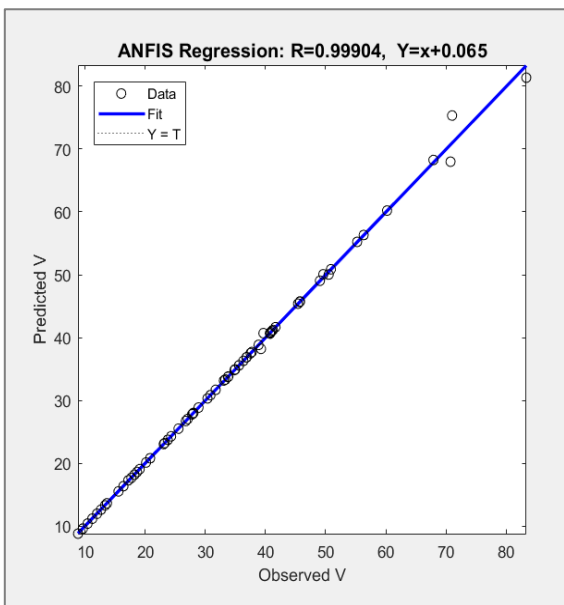
Figure A-69: Regression analysis for V in ANFIS (training).
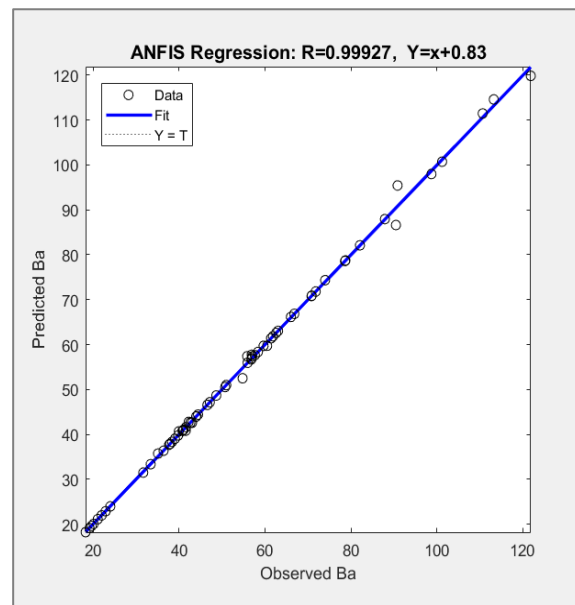
Figure A-60: Regression analysis for Ba in ANFIS (training).

187
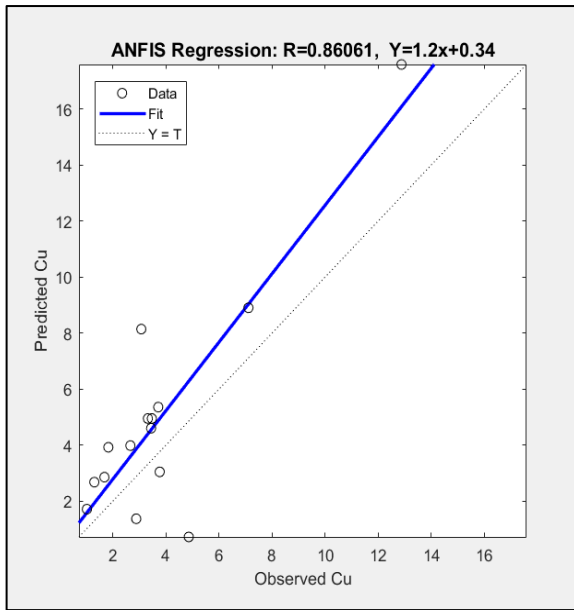
Figure A-61: Regression analysis for Cu in ANFIS (testing).
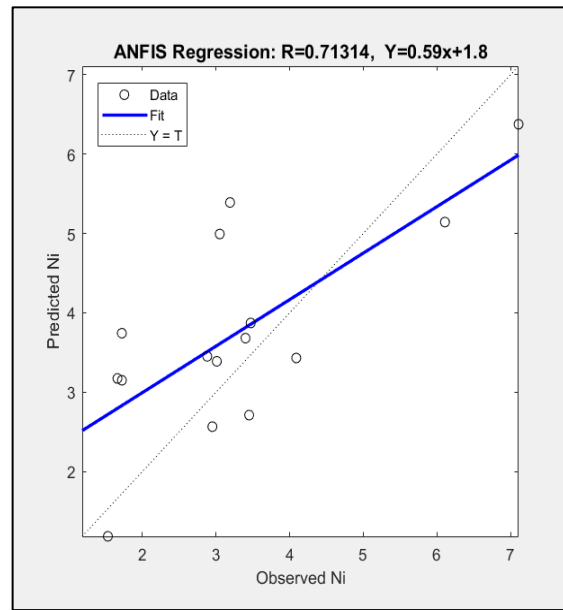
Figure A-62: Regression analysis for Ni in ANFIS (testing).
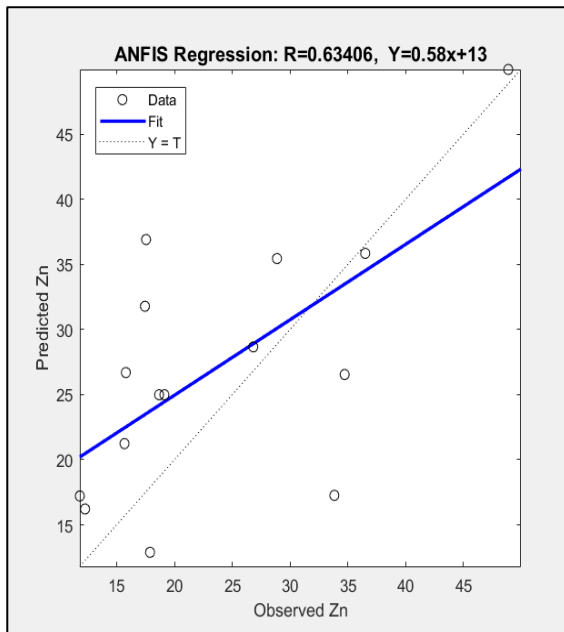
Figure A-63: Regression analysis for Zn in ANFIS (testing).
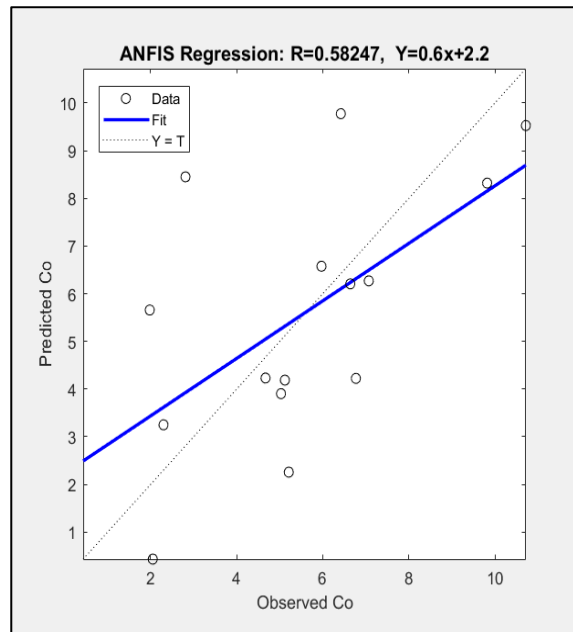
Figure A-64: Regression analysis for Co in ANFIS (testing).
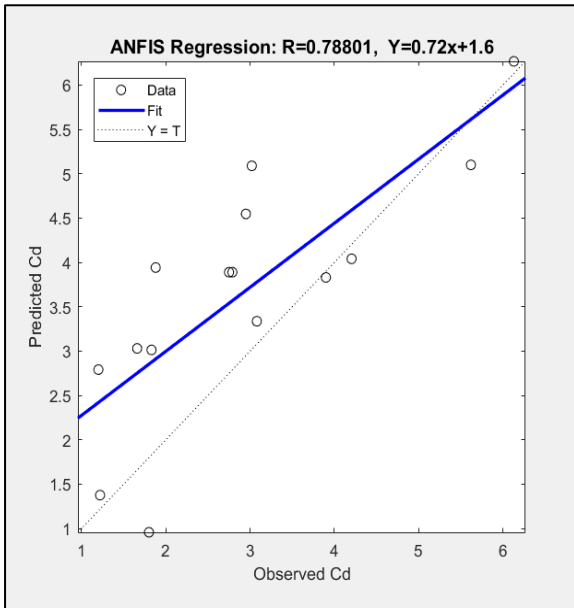
Figure A-65: Regression analysis for Cd in ANFIS (testing).
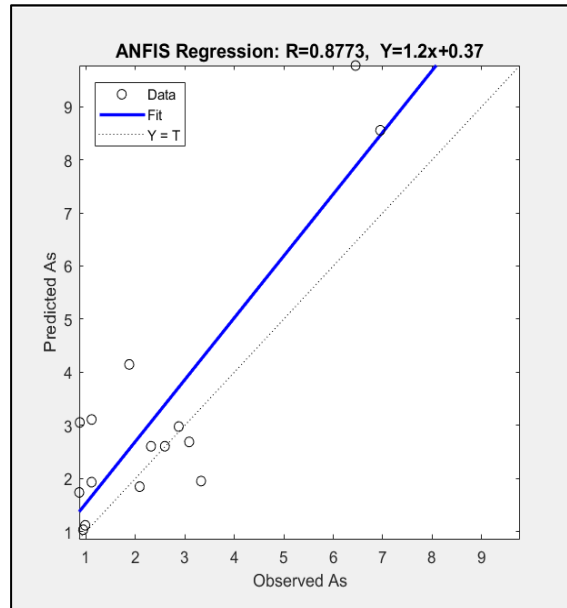
Figure A-66: Regression analysis for As in ANFIS (testing).
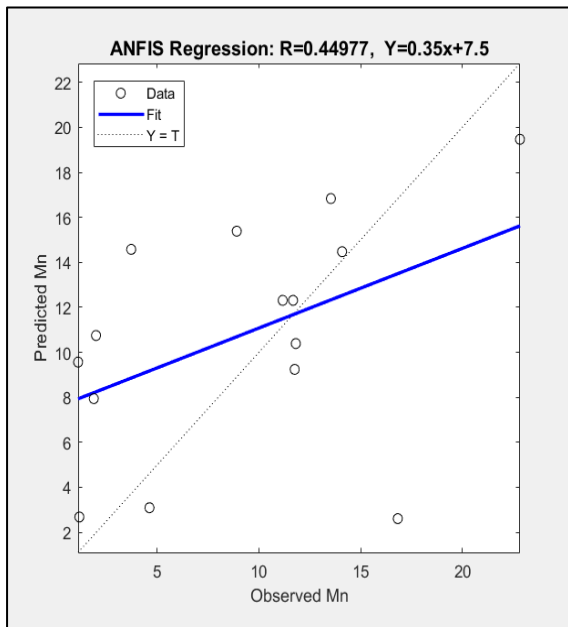
Figure A-67: Regression analysis for Mn in ANFIS (testing).
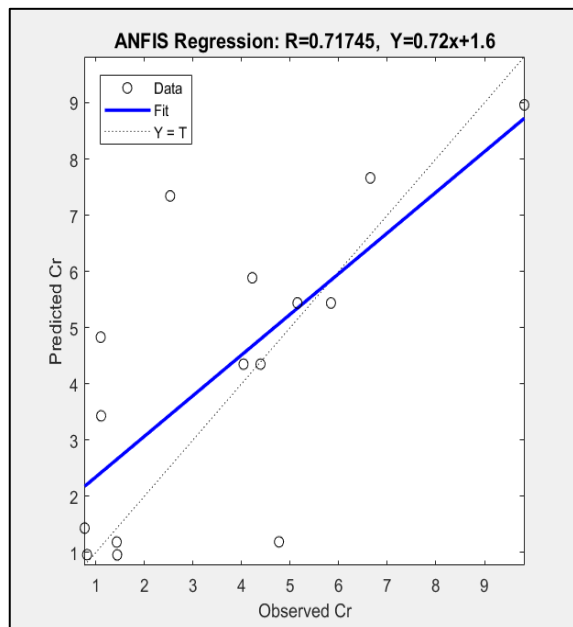
Figure A-68: Regression analysis for Cr in ANFIS (testing).
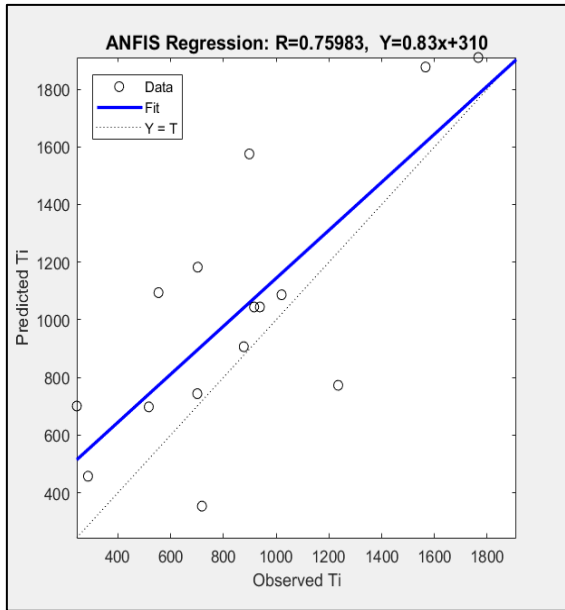
Figure A-69: Regression analysis for Ti in ANFIS (testing).
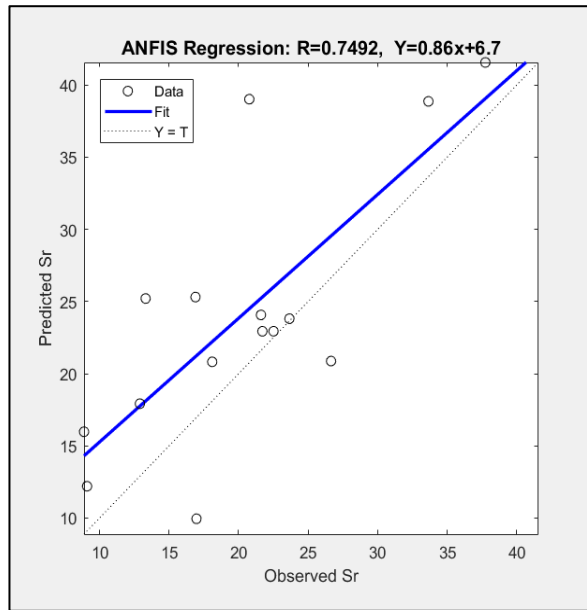
Figure A-70: Regression analysis for Sr in ANFIS (testing).
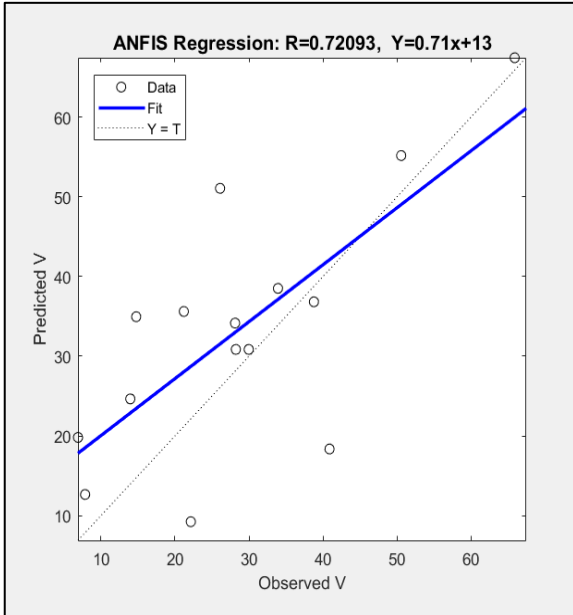
Figure A-71: Regression analysis for V in ANFIS (testing).
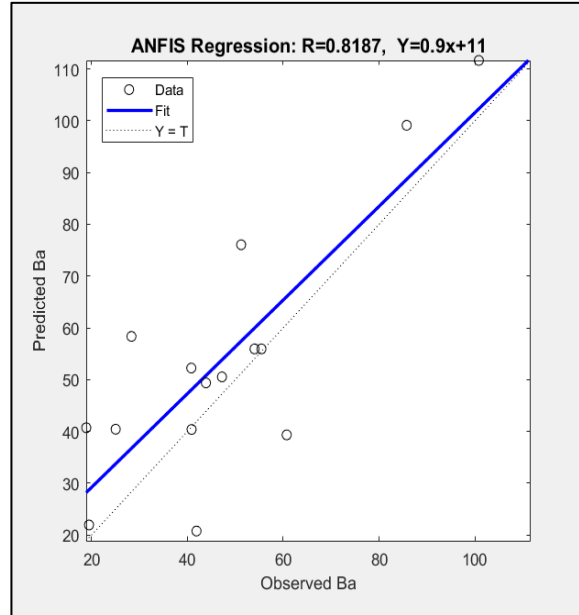
Figure A-72: Regression analysis for Ba in ANFIS (testing).

190
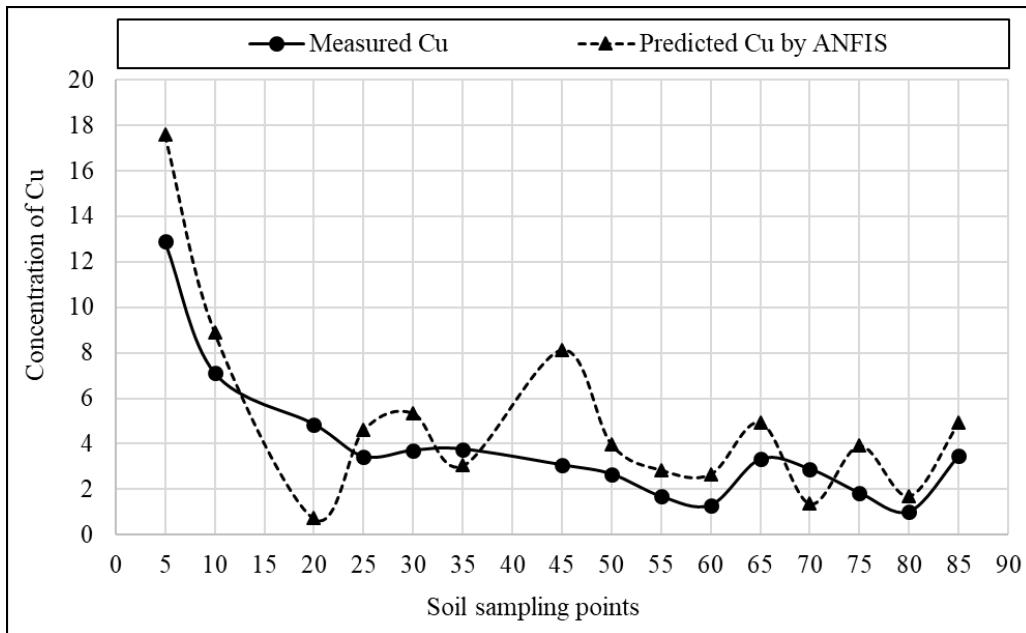
Figure A-73: Comparison of predicted and measured concentration of Cu from ANFIS (testing).
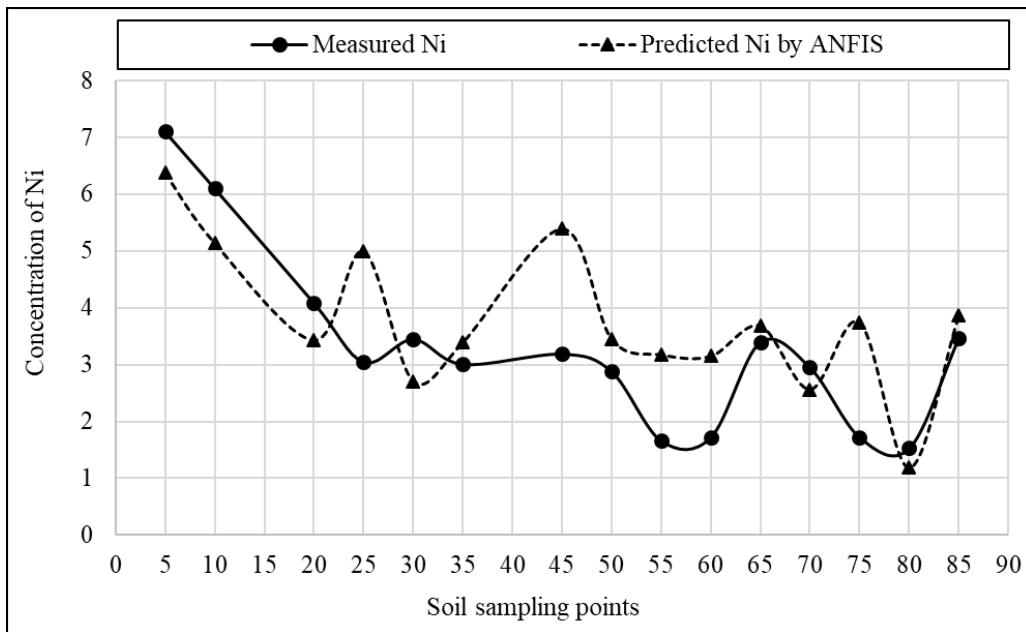
Figure A-74: Comparison of predicted and measured concentration of Ni from ANFIS (testing).
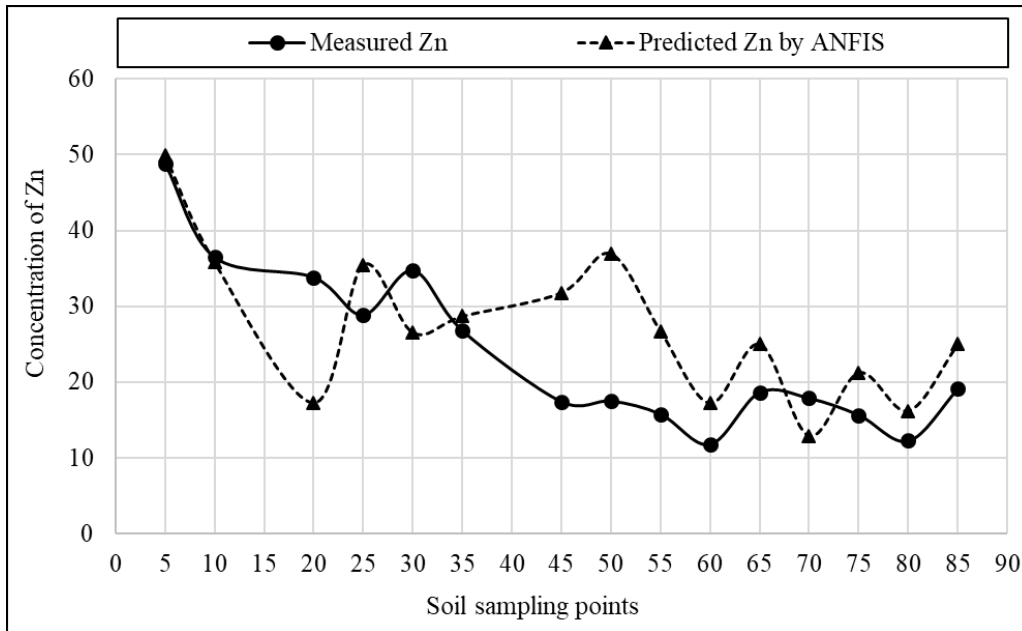
Figure A-75: Comparison of predicted and measured concentration of Zn from ANFIS (testing).
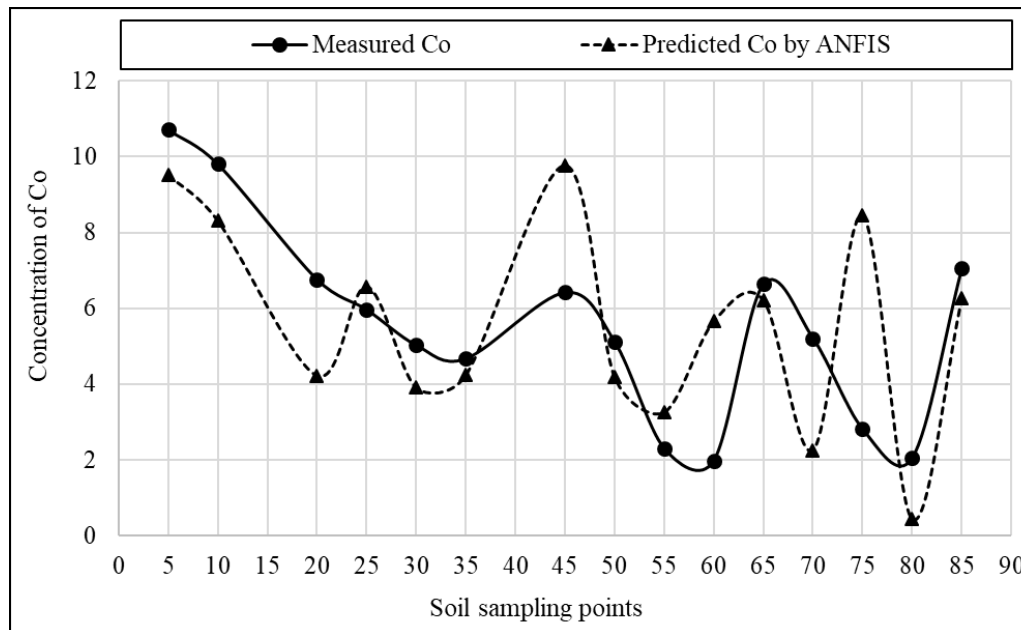
Figure A-76: Comparison of predicted and measured concentration of Co from ANFIS (testing).
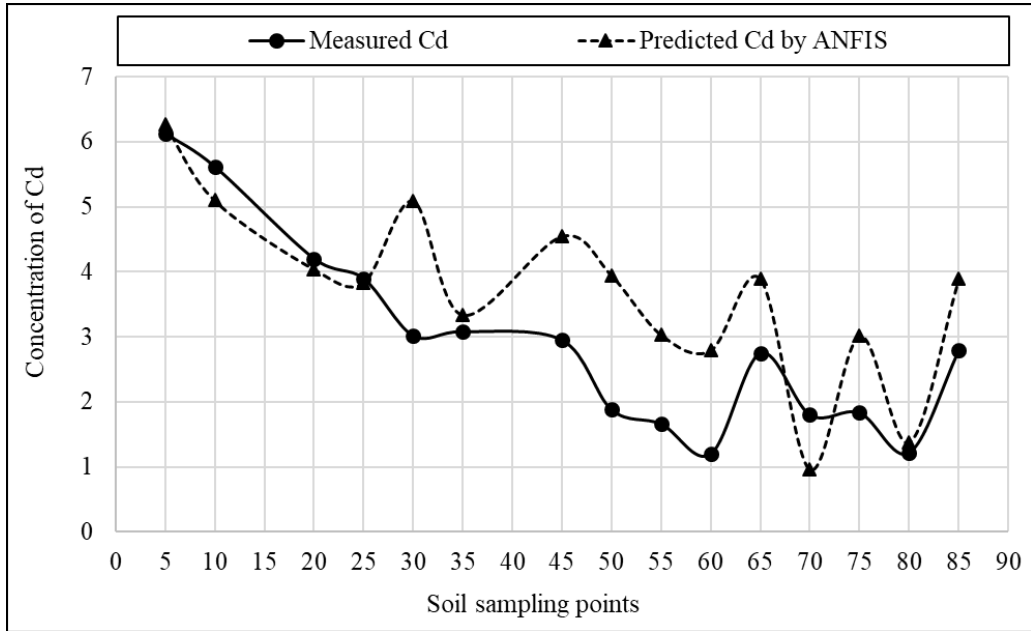
Figure A-77: Comparison of predicted and measured concentration of Cd from ANFIS (testing).
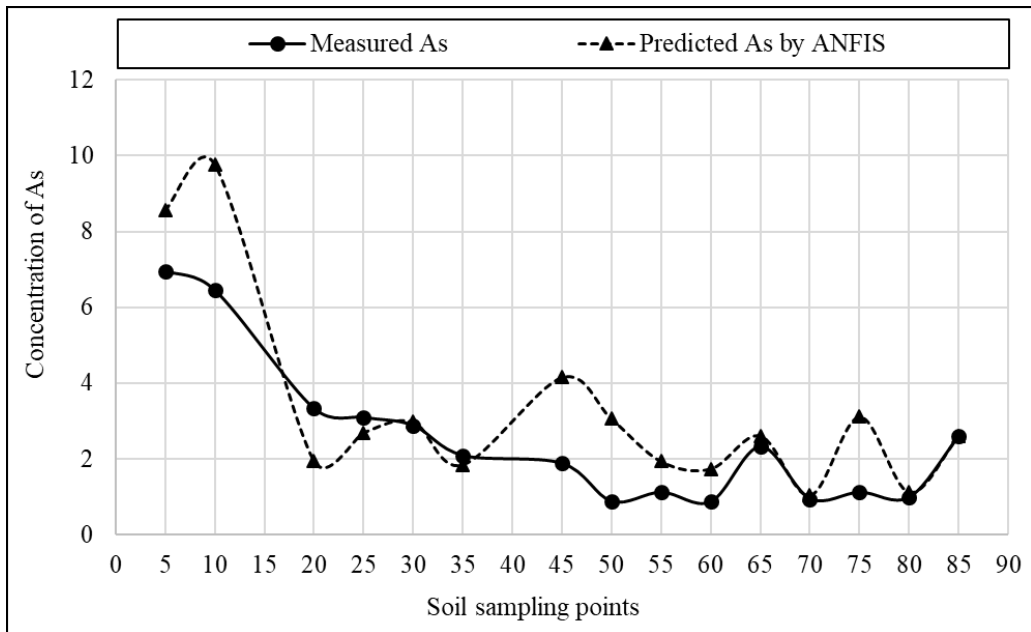
Figure A-78: Comparison of predicted and measured concentration of As from ANFIS (testing).

193
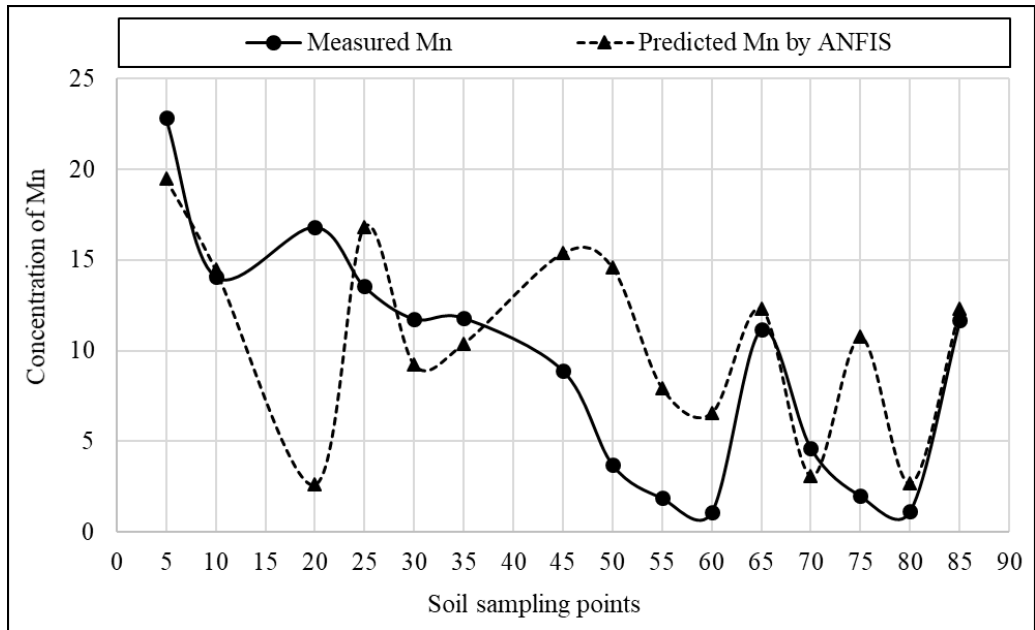
Figure A-79: Comparison of predicted and measured concentration of Mn from ANFIS (testing).
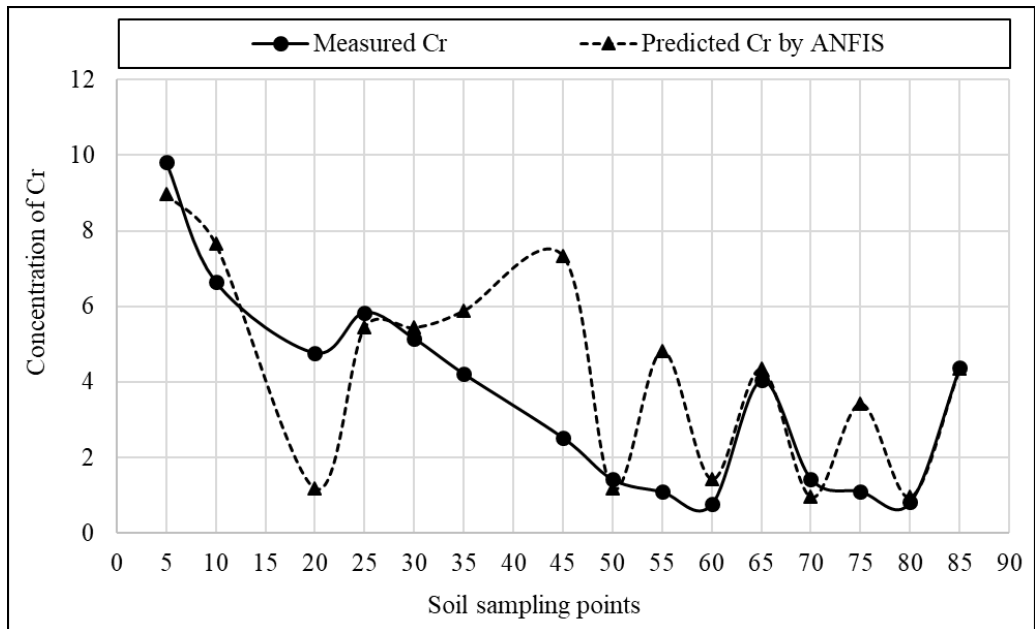
Figure A-80: Comparison of predicted and measured concentration of Cr from ANFIS (testing).
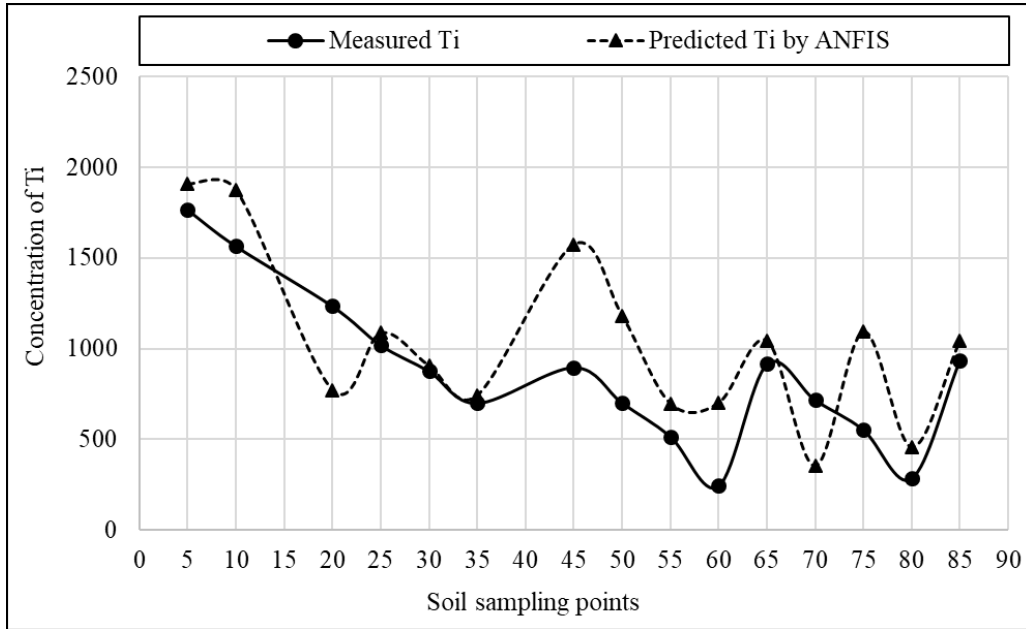
Figure A-81: Comparison of predicted and measured concentration of Ti from ANFIS (testing).
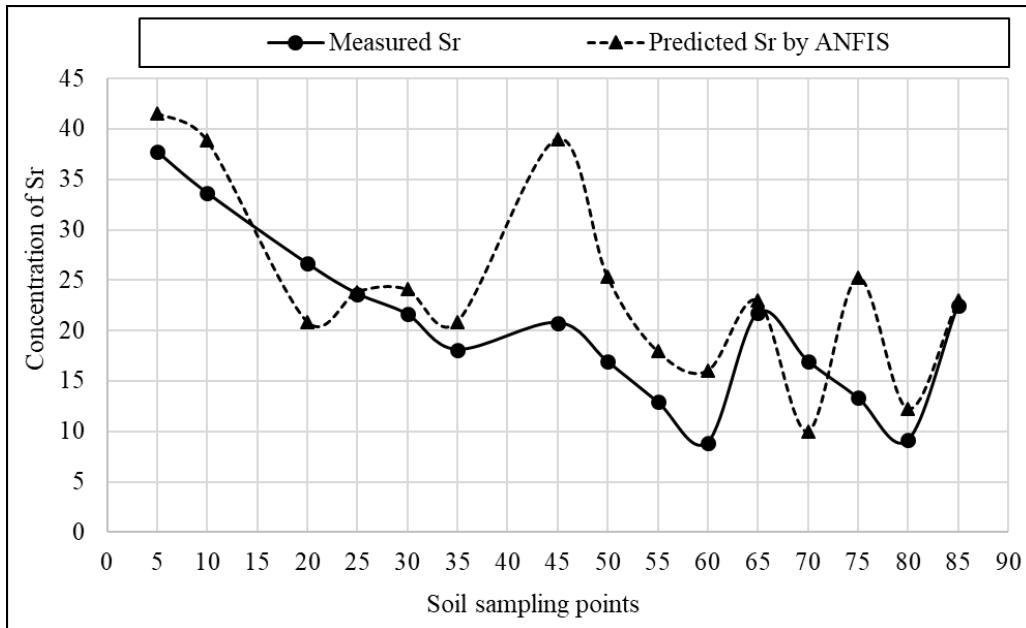
Figure A-82: Comparison of predicted and measured concentration of Sr from ANFIS (testing).

195
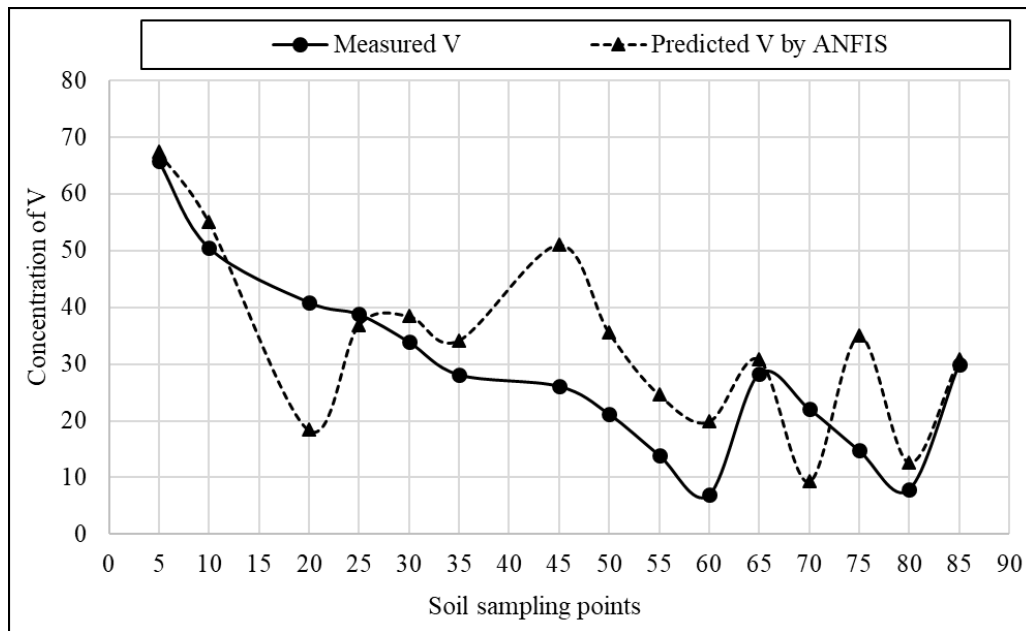
Figure A-83: Comparison of predicted and measured concentration of V from ANFIS model (testing).
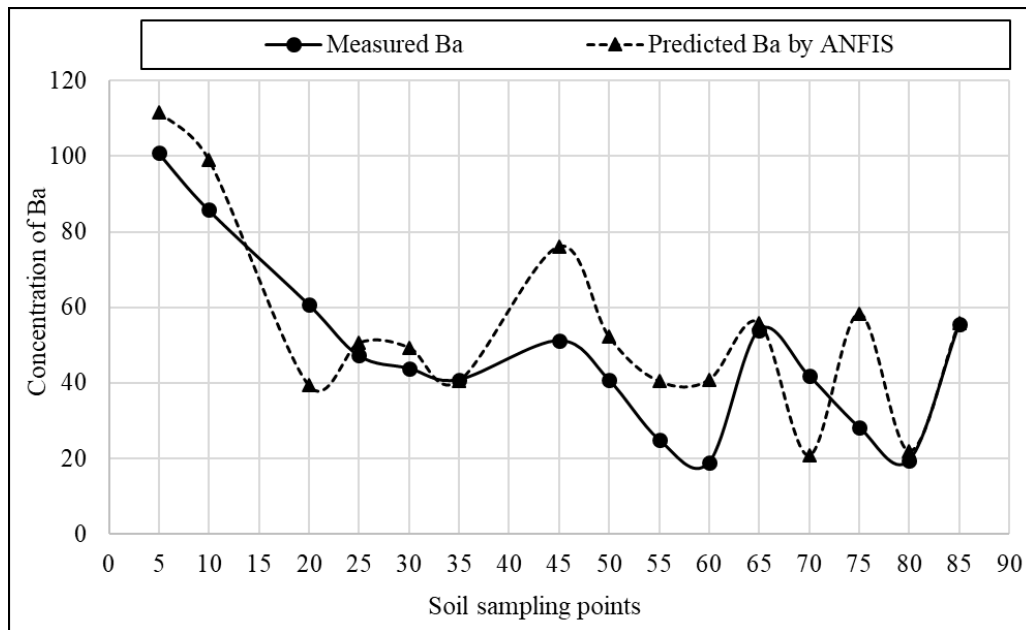
Figure A-84: Comparison of predicted and measured concentration of Ba from ANFIS model (testing).

196
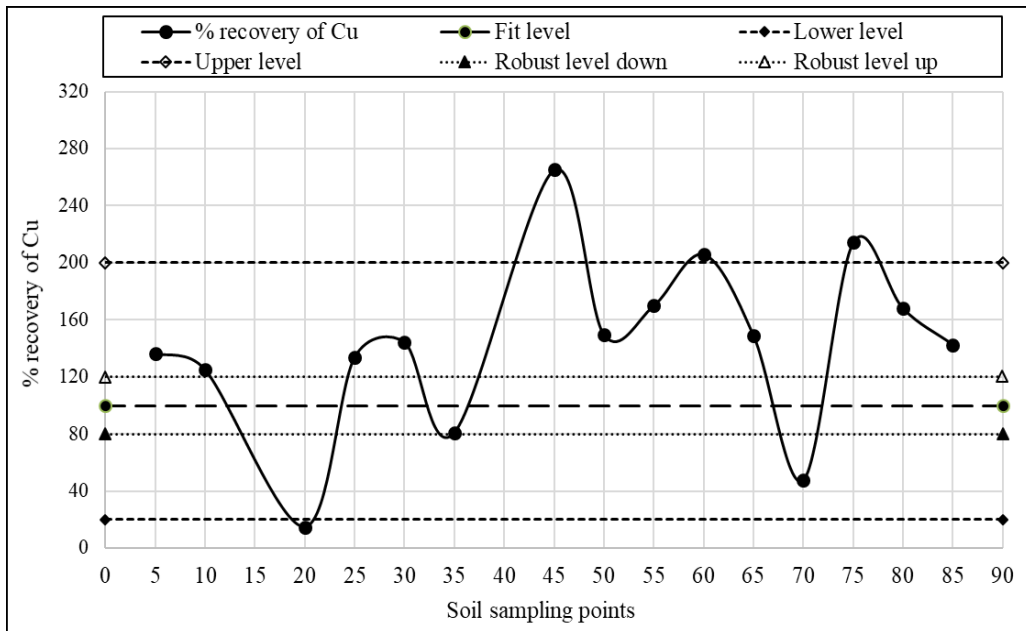
Figure A-85: Variation of recovery level for Cu in ANFIS.
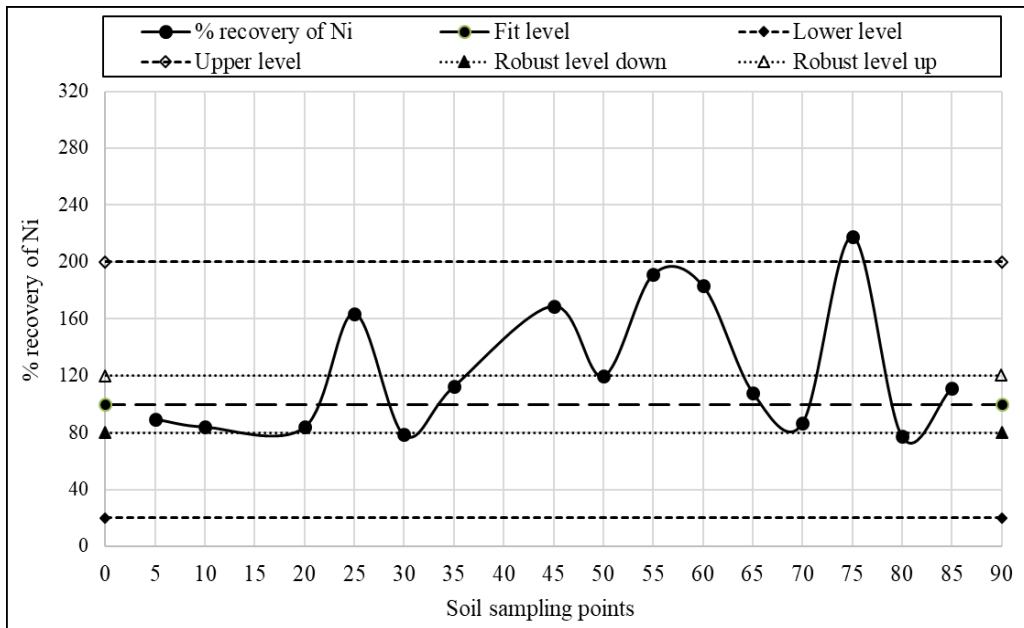
Figure A-86: Variation of recovery level for Ni in ANFIS.

197
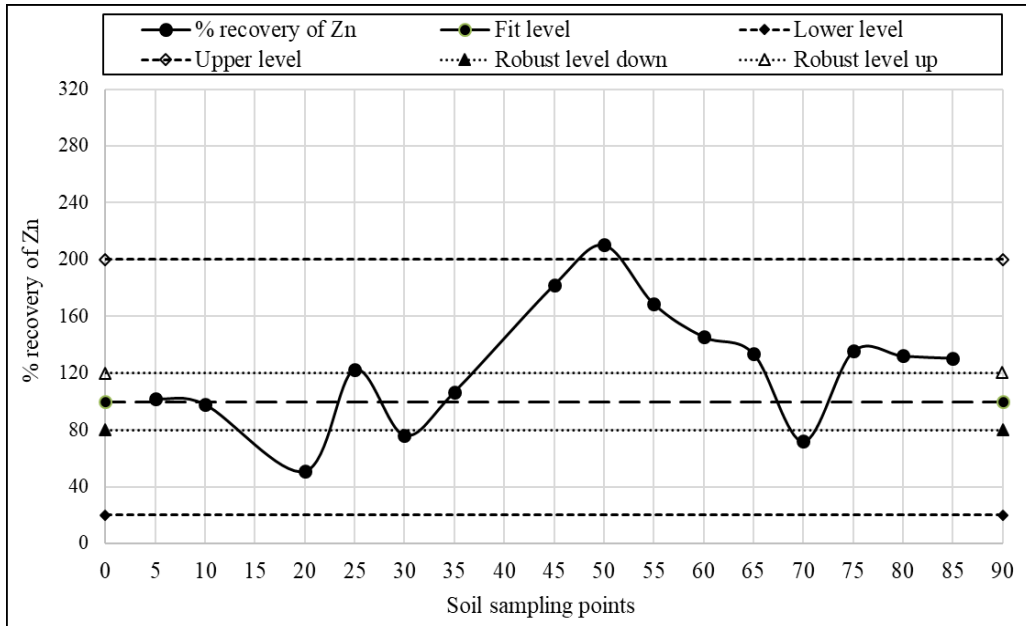
Figure A-87: Variation of recovery level for Zn in ANFIS.
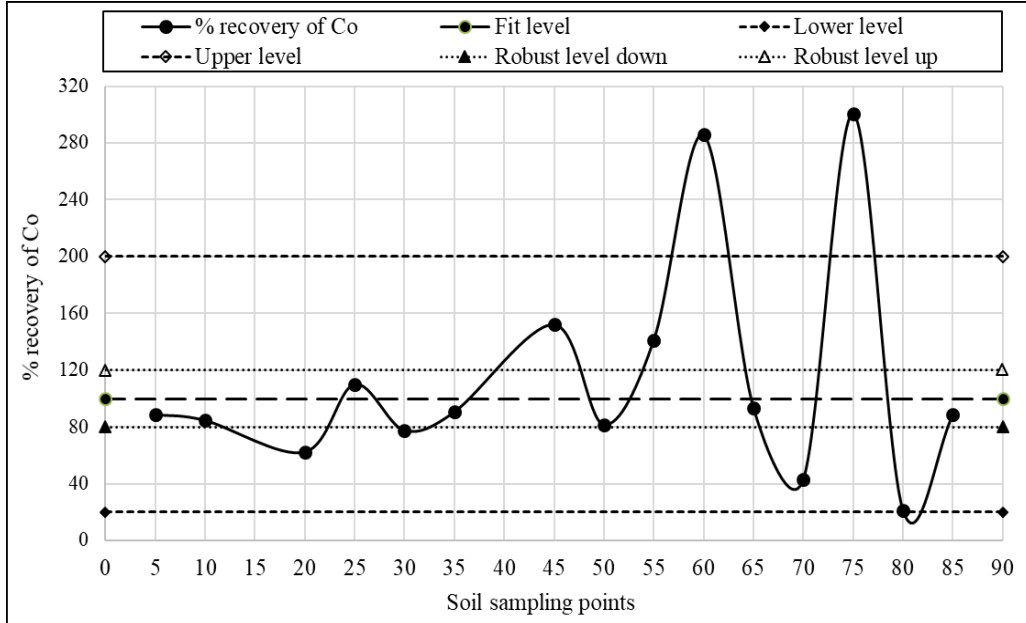
Figure A-88: Variation of recovery level for Co in ANFIS.
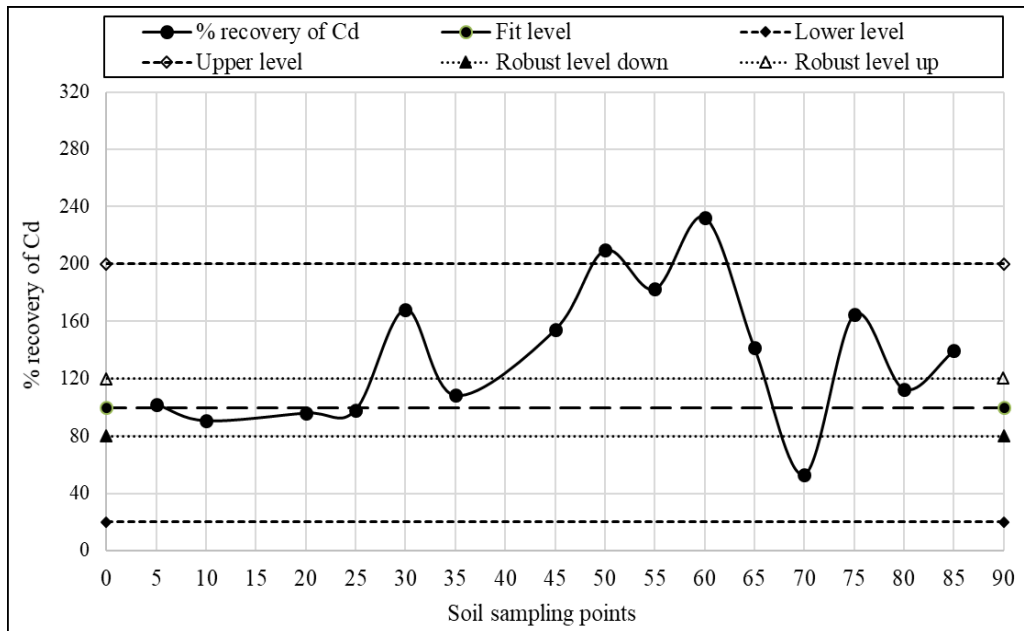
Figure A-89: Variation of recovery level for Cd in ANFIS.
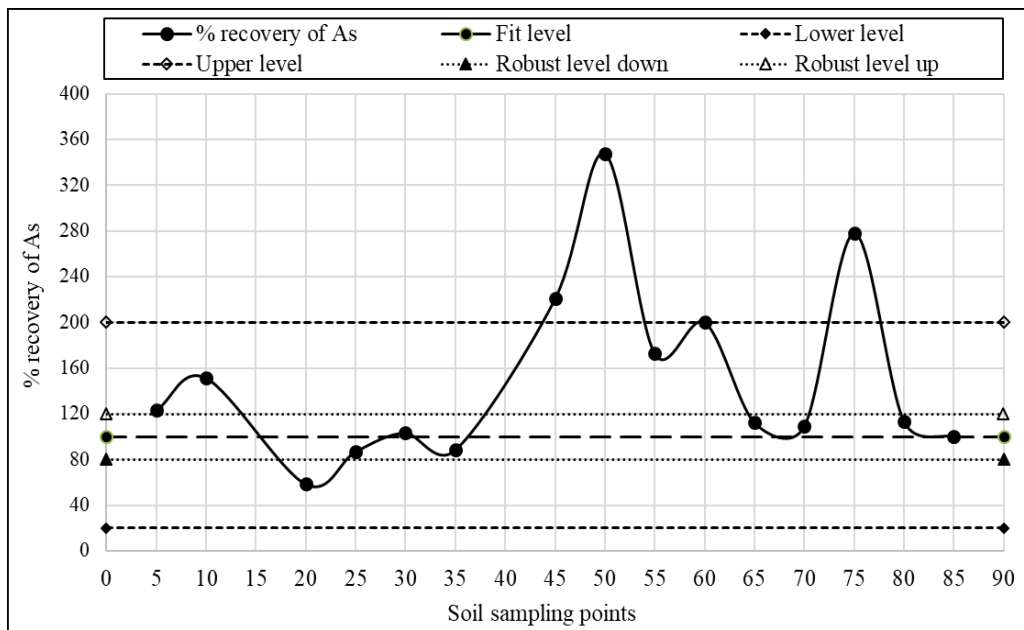
Figure A-90: Variation of recovery level for As in ANFIS.

199
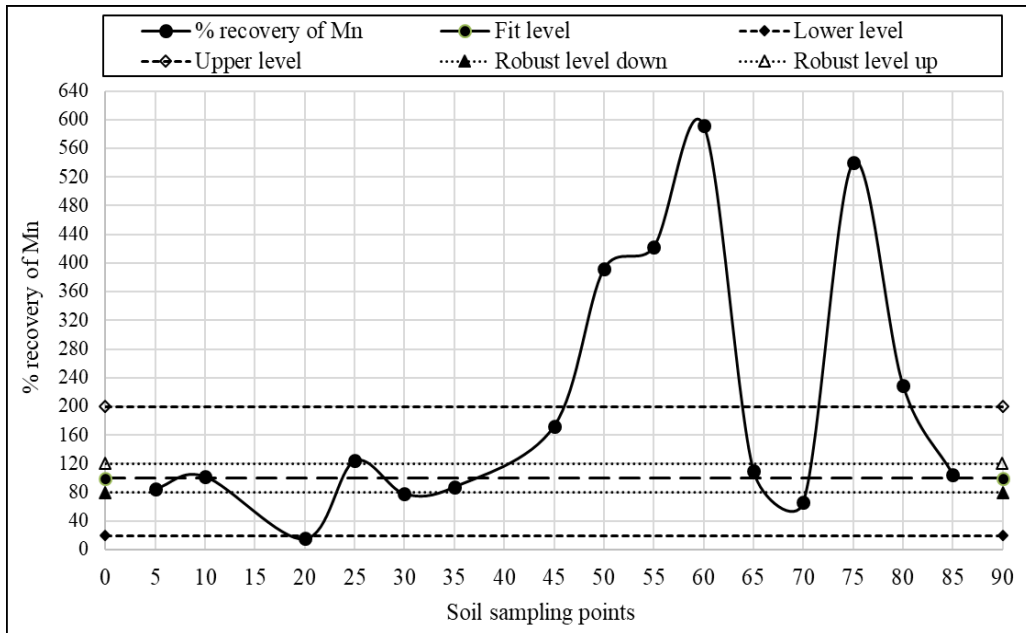
Figure A-91: Variation of recovery level for Mn in ANFIS.
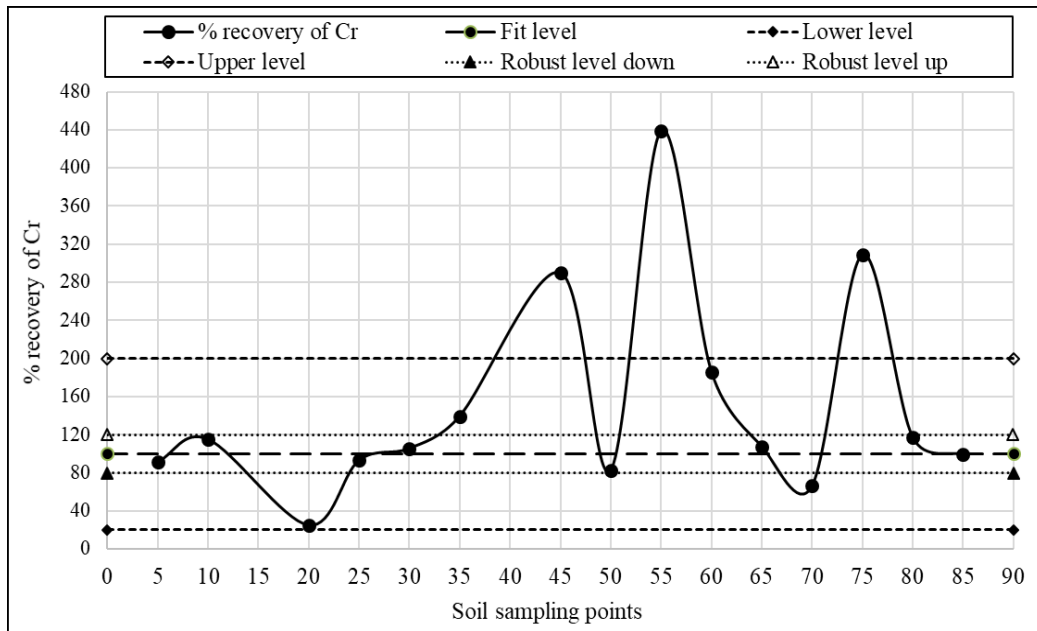
Figure A-92: Variation of recovery level for Cr in ANFIS.

200
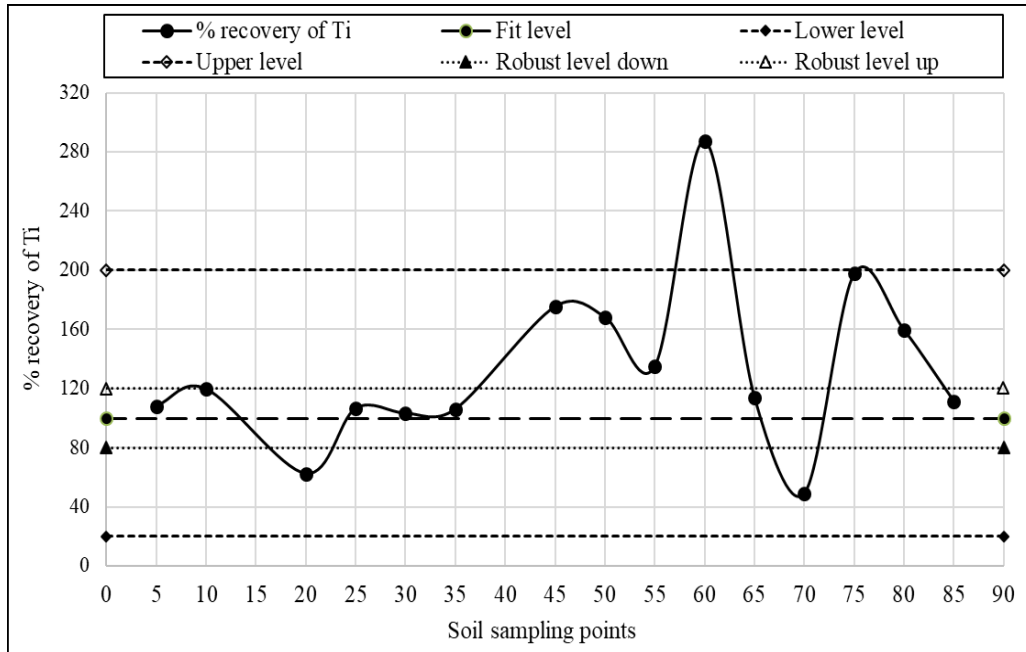
Figure A-93: Variation of recovery level for Ti in ANFIS.
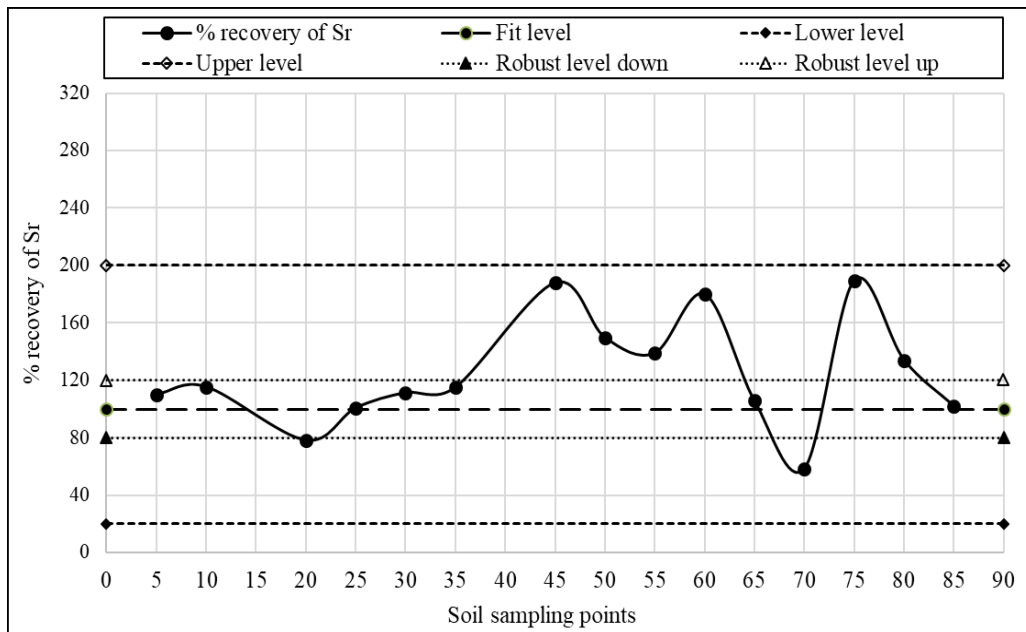
Figure A-94: Variation of recovery level for Sr in ANFIS.
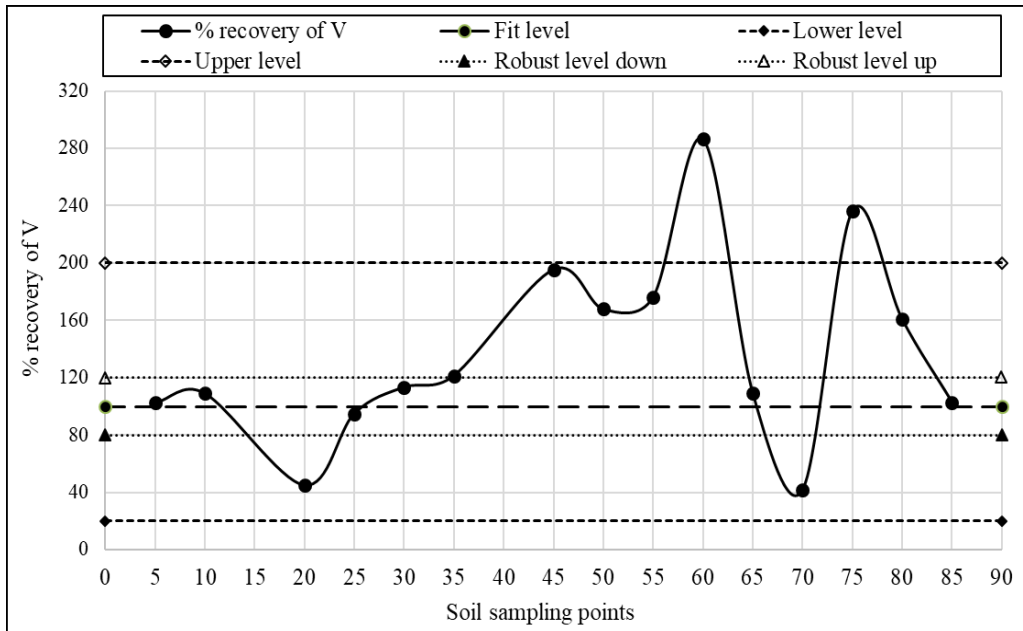
Figure A-95: Variation of recovery level for V in ANFIS.
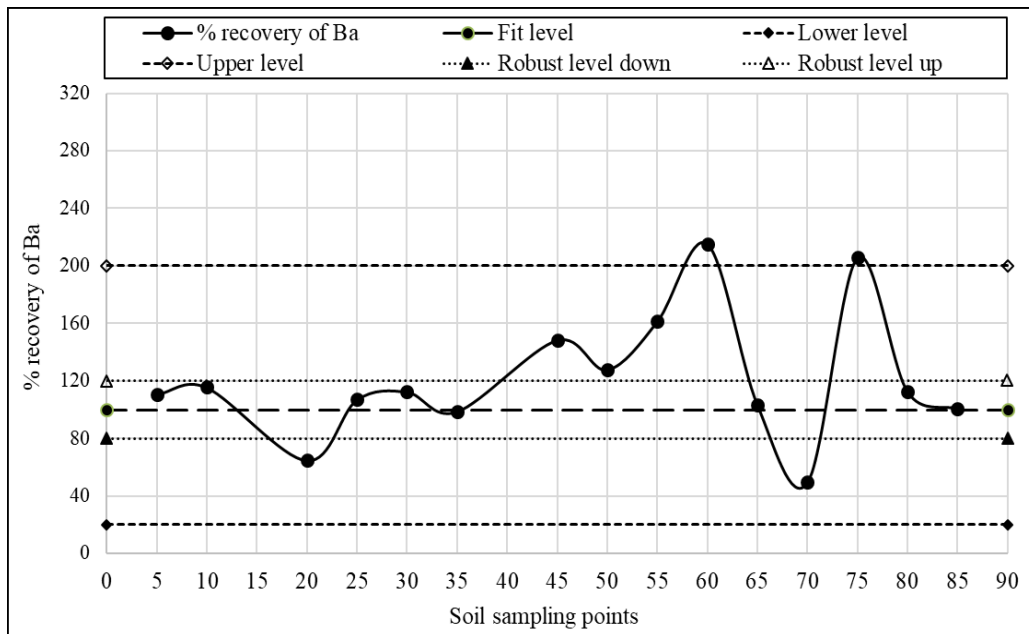
Figure A-96: Variation of recovery level for Ba in ANFIS.

202

Appendix-B

Recovery Level for Various Heavy Metals of ANFIS

Table B-1: Recovery level of Cu in ANFIS (testing)

Soils sampling points	Cu		
	Measured	Predicted	% recovery
5	12.88	17.59	136.57
10	7.11	8.90	125.22
20	4.86	0.72	14.86
25	3.44	4.60	133.79
30	3.71	5.36	144.40
35	3.77	3.05	80.88
45	3.07	8.14	265.21
50	2.66	3.98	149.71
55	1.68	2.86	169.99
60	1.3	2.68	205.98
65	3.32	4.95	149.09
70	2.88	1.37	47.58
75	1.83	3.92	214.38
80	1.02	1.71	168.09
85	3.47	4.95	142.65
% mean recovery			143.23
RMSE			2.45
MAPE			64.12
GRI			1.91

Table B-2: Recovery level of Ni in ANFIS (testing)

Soils sampling points	Ni		
	Measured	Predicted	% recovery
5	7.11	6.38	89.69
10	6.11	5.14	84.19
20	4.09	3.43	83.88
25	3.05	4.99	163.74
30	3.45	2.71	78.56
35	3.01	3.39	112.56
45	3.19	5.39	169.01
50	2.88	3.45	119.83
55	1.66	3.17	191.10
60	1.72	3.15	183.11
65	3.4	3.68	108.25
70	2.95	2.56	86.92
75	1.72	3.74	217.61
80	1.53	1.18	77.24
85	3.47	3.87	111.58
% mean recovery			125.15
RMSE			1.17
MAPE			38.42
GRI			1.46

Table B-3: Recovery level of Zn in ANFIS (testing)

Soils sampling points	Zn		
	Measured	Predicted	% recovery
5	48.87	49.96	102.23
10	36.5	35.84	98.20
20	33.83	17.27	51.04
25	28.87	35.45	122.78
30	34.72	26.55	76.46
35	26.82	28.67	106.90
45	17.45	31.78	182.13
50	17.55	36.91	210.29
55	15.8	26.69	168.95
60	11.82	17.22	145.66
65	18.67	24.99	133.85
70	17.89	12.89	72.07
75	15.68	21.24	135.46
80	12.27	16.22	132.19
85	19.13	24.99	130.63
% mean recovery			124.59
RMSE			9.18
MAPE			38.22
GRI			1.48

Table B-4: Recovery level of Co in ANFIS (testing)

Soils sampling points	Co		
	Measured	Predicted	% recovery
5	10.72	9.53	88.88
10	9.82	8.32	84.72
20	6.77	4.22	62.39
25	5.97	6.58	110.15
30	5.03	3.90	77.61
35	4.67	4.23	90.61
45	6.42	9.78	152.27
50	5.12	4.19	81.79
55	2.3	3.25	141.35
60	1.98	5.66	285.74
65	6.64	6.21	93.47
70	5.21	2.26	43.34
75	2.81	8.45	300.76
80	2.05	0.43	21.15
85	7.07	6.27	88.67
% mean recovery			114.86
RMSE			2.35
MAPE			50.51
GRI			1.80

Soils sampling points	Cd		
	Measured	Predicted	% recovery
5	6.13	6.27	102.23
10	5.62	5.10	90.75
20	4.20	4.04	96.11
25	3.9	3.83	98.22
30	3.02	5.09	168.49
35	3.08	3.34	108.34
45	2.95	4.55	154.14
50	1.88	3.94	209.69
55	1.66	3.03	182.55
60	1.20	2.79	232.65
65	2.75	3.89	141.45
70	1.8	0.96	53.21
75	1.83	3.01	164.69
80	1.22	1.38	112.80
85	2.79	3.89	139.41
% mean recovery			136.98
RMSE			1.17
MAPE			45.21
GRI			1.55

Table B-5: Recovery level of Cd in ANFIS (testing)

Soils sampling points	As		
	Measured	Predicted	% recovery
5	6.95	8.55	123.06
10	6.45	9.77	151.50
20	3.33	1.95	58.70
25	3.09	2.69	87.11
30	2.88	2.98	103.47
35	2.09	1.85	88.49
45	1.88	4.15	220.65
50	0.88	3.06	347.49
55	1.12	1.93	172.74
60	0.87	1.74	200.19
65	2.32	2.61	112.50
70	0.95	1.04	109.39
75	1.12	3.11	277.93
80	0.99	1.12	113.43
85	2.6	2.61	100.39
% mean recovery			151.14
RMSE			1.44
MAPE			59.90
GRI			1.69

Table B-6: Recovery level of As in ANFIS (testing)

Table B-7: Recovery level of Mn in ANFIS (testing)

Soils sampling points	Mn		
	Measured	Predicted	% recovery
5	22.84	19.47	85.24
10	14.09	14.47	102.71
20	16.83	2.61	15.51
25	13.54	16.83	124.33
30	11.76	9.24	78.60
35	11.815	10.39	87.94
45	8.91	15.38	172.66
50	3.72	14.58	391.84
55	1.88	7.94	422.56
60	1.11	6.57	591.81
65	11.17	12.31	110.22
70	4.62	3.09	66.97
75	1.99	10.75	540.12
80	1.17	2.68	229.37
85	11.68	12.31	105.40
% mean recovery			208.35
RMSE			6.24
MAPE			130.45
GRI			2.40

Table B-8: Recovery level of Cr in ANFIS (testing)

Soils sampling points	Cr		
	Measured	Predicted	% recovery
5	9.82	8.96	91.27
10	6.65	7.66	115.24
20	4.77	1.19	24.91
25	5.84	5.44	93.12
30	5.15	5.44	105.60
35	4.22	5.89	139.46
45	2.53	7.34	290.19
50	1.43	1.18	82.82
55	1.1	4.83	439.12
60	0.77	1.43	186.14
65	4.04	4.35	107.67
70	1.44	0.96	66.42
75	1.11	3.43	309.01
80	0.82	0.96	117.08
85	4.39	4.35	99.09
% mean recovery			151.14
RMSE			2.02
MAPE			70.13
GRI			1.90

Table B-9: Recovery level of Ti in ANFIS (testing)

Soils sampling points	Ti		% recovery
	Measured	Predicted	
5	1765.91	1910.06	108.16
10	1566.169	1877.04	119.85
20	1234.98	772.68	62.57
25	1020.28	1086.96	106.54
30	876.93	906.49	103.37
35	700.83	743.63	106.11
45	898.11	1575.55	175.43
50	702.3	1182.59	168.39
55	516.89	697.98	135.04
60	243.88	700.78	287.35
65	916.17	1044.41	114.00
70	718.54	353.38	49.18
75	553.55	1094.22	197.67
80	286.55	457.14	159.53
85	938.02	1044.46	111.35
% mean recovery			133.63
RMSE			341.76
MAPE			45.40
GRI			1.57

Table B-10: Recovery level of Sr in ANFIS (testing)

Soils sampling points	Sr		% recovery
	Measured	Predicted	
5	37.75	41.57	110.12
10	33.66	38.87	115.49
20	26.65	20.87	78.32
25	23.65	23.81	100.68
30	21.61	24.07	111.41
35	18.10	20.82	115.03
45	20.77	39.02	187.88
50	16.9	25.31	149.78
55	12.9	17.92	138.90
60	8.88	15.97	179.88
65	21.71	22.94	105.66
70	16.97	9.95	58.61
75	13.31	25.21	189.41
80	9.1	12.20	134.07
85	22.51	22.94	101.91
% mean recovery			125.14
RMSE			7.17
MAPE			33.55
GRI			1.42

Table B-11: Recovery level of V in ANFIS (testing)

Soils sampling points	V		
	Measured	Predicted	% recovery
5	65.88	67.43	102.35
10	50.55	55.16	109.12
20	40.88	18.37	44.95
25	38.77	36.82	94.98
30	33.92	38.51	113.54
35	28.13	34.16	121.47
45	26.1	51.05	195.60
50	21.21	35.62	167.93
55	13.98	24.66	176.36
60	6.92	19.83	286.53
65	28.23	30.86	109.31
70	22.15	9.26	41.80
75	14.78	34.96	236.54
80	7.88	12.67	160.79
85	29.97	30.86	102.97
% mean recovery			137.61
RMSE			12.40
MAPE			53.38
GRI			1.71

Table B-12: Recovery level of Ba in ANFIS (testing)

Soils sampling points	Ba		
	Measured	Predicted	% recovery
5	100.82	111.64	110.73
10	85.76	99.12	115.58
20	60.73	39.35	64.79
25	47.22	50.54	107.02
30	43.88	49.37	112.52
35	40.88	40.39	98.81
45	51.22	76.05	148.48
50	40.82	52.25	127.99
55	25.04	40.43	161.45
60	18.9	40.71	215.42
65	54.01	55.94	103.57
70	41.91	20.77	49.57
75	28.34	58.35	205.89
80	19.48	21.95	112.68
85	55.46	55.94	100.87
% mean recovery			122.36
RMSE			15.52
MAPE			33.94
GRI			1.47

Appendix-C

Results of Heavy Metal Analysis and Assessment of SVM
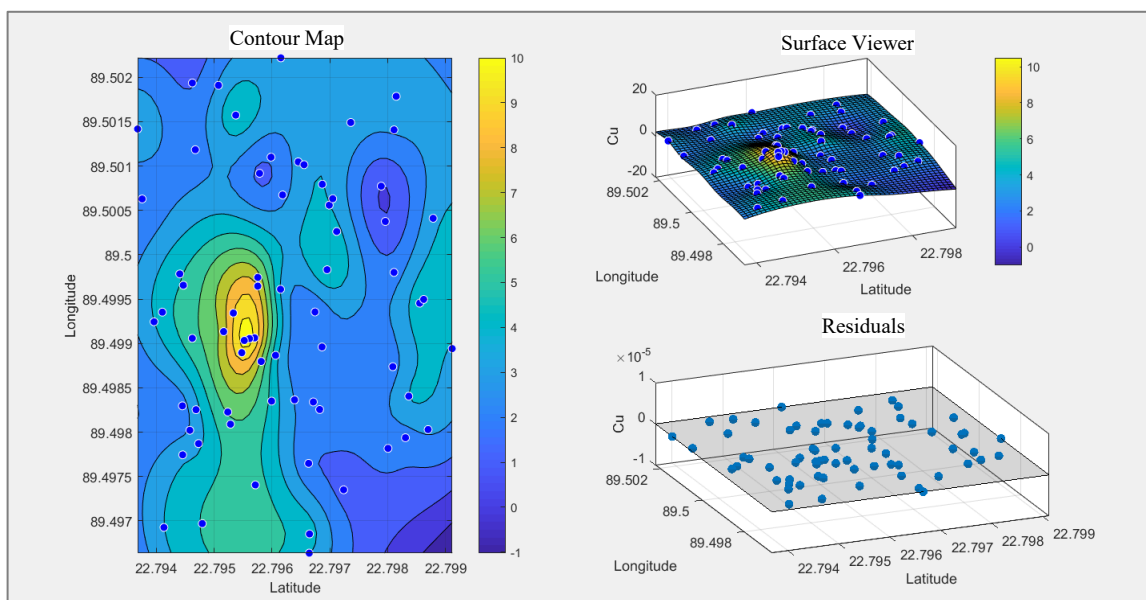
Figure C-1: Graphical representation of outputs for Cu.
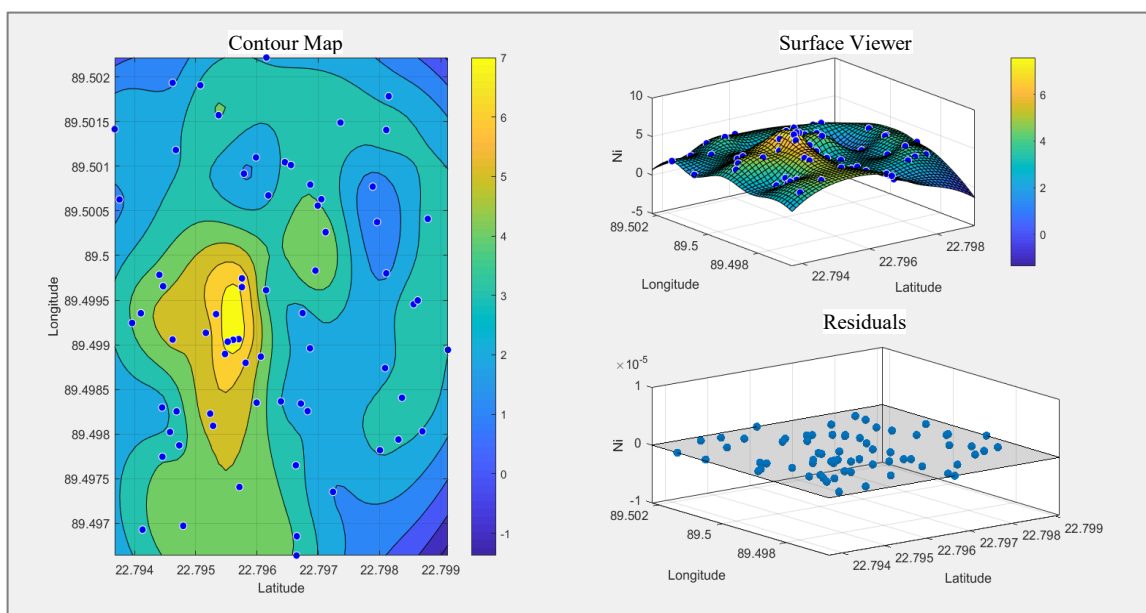
Figure C-2: Graphical representation of outputs for Ni.
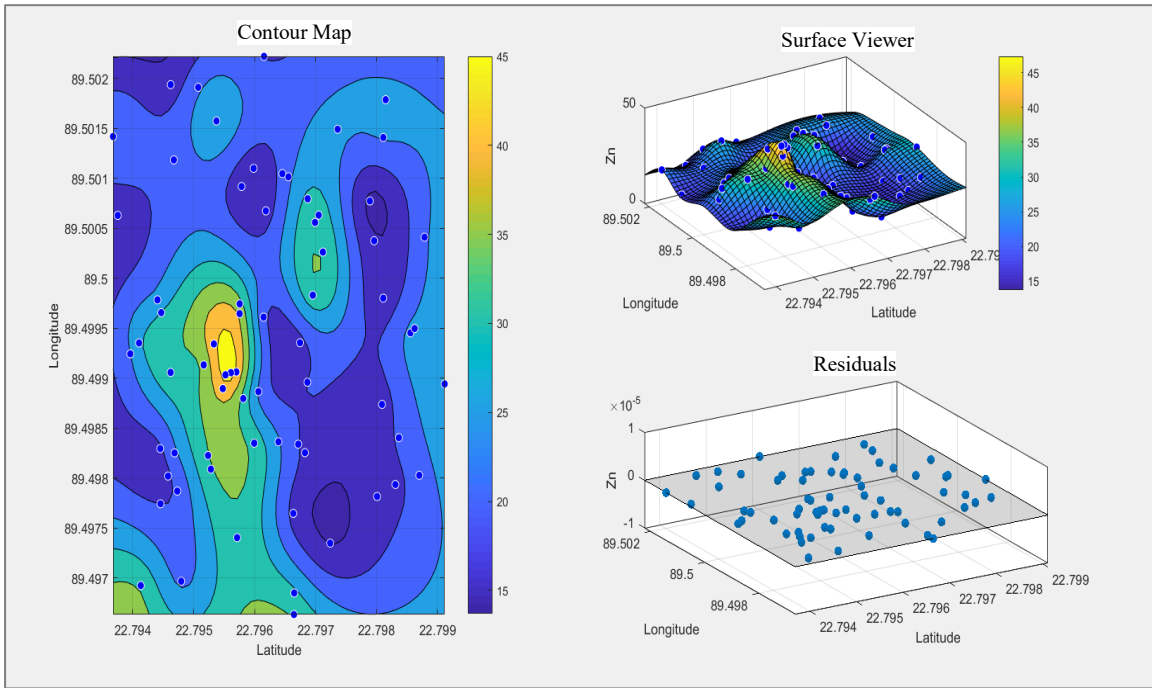
Figure C-3: Graphical representation of outputs for Zn.
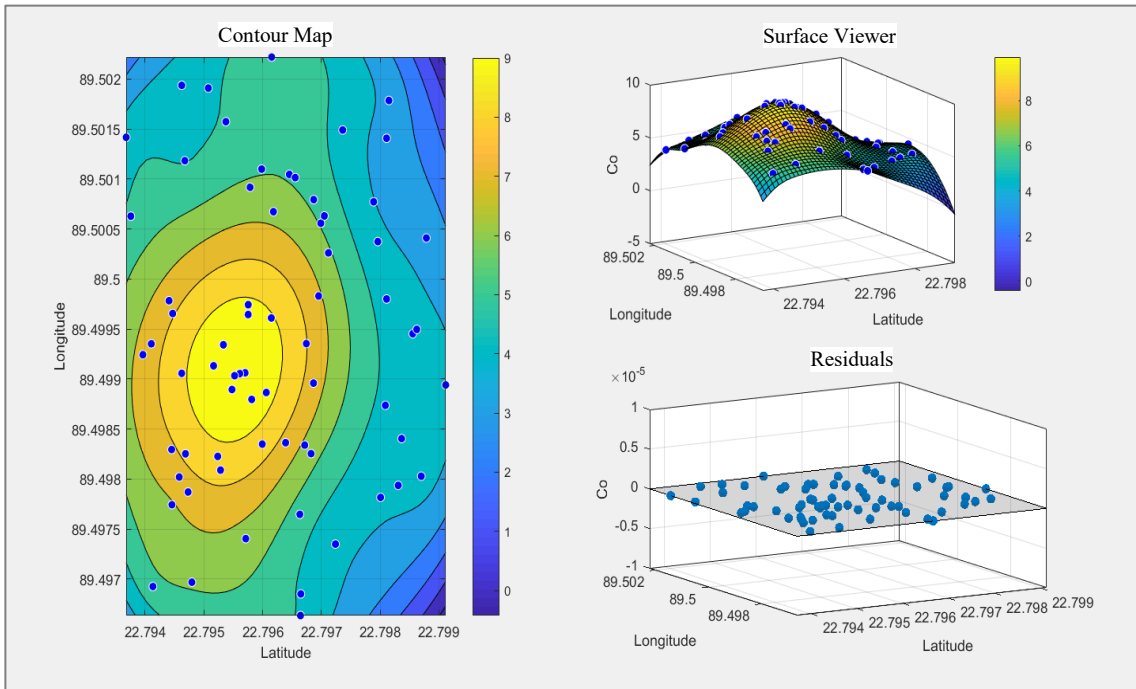
Figure C-4: Graphical representation of outputs for Co.
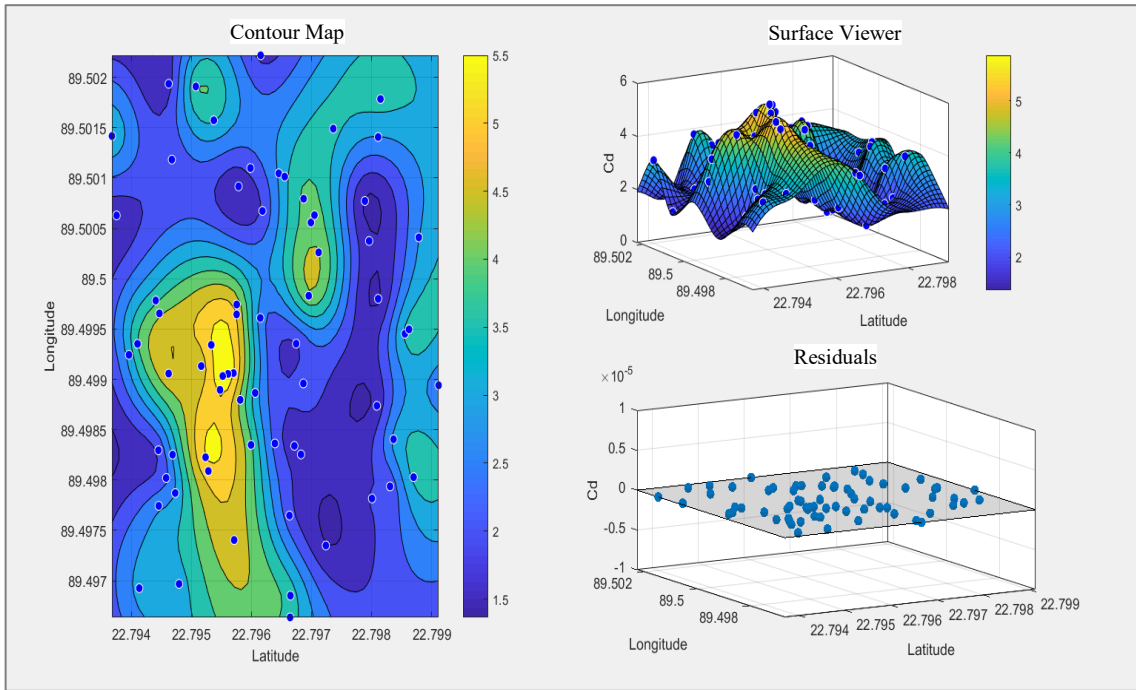
Figure C-5: Graphical representation of outputs for Cd.
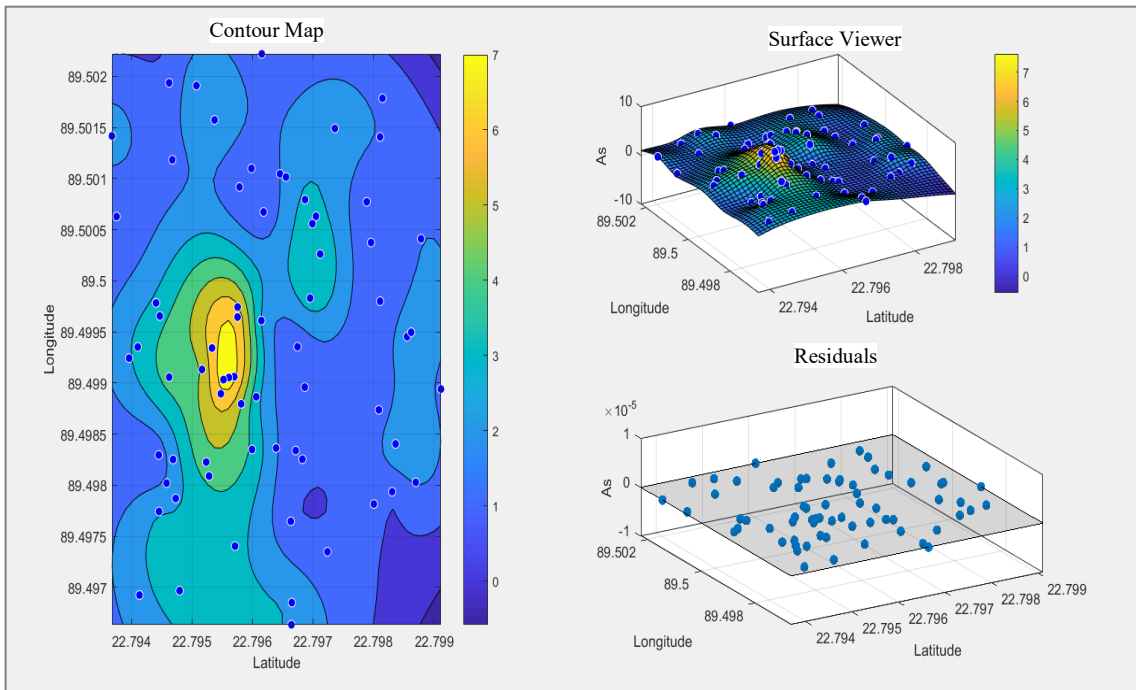
Figure C-6: Graphical representation of outputs for As.
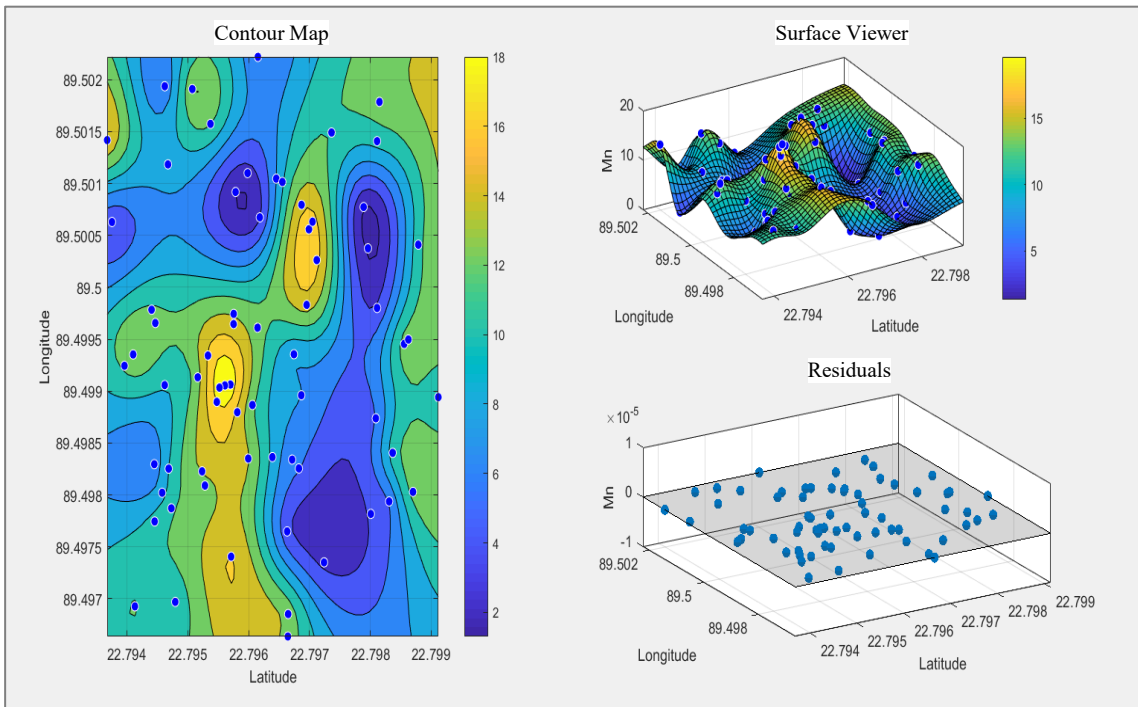
Figure C-7: Graphical representation of outputs for Mn.
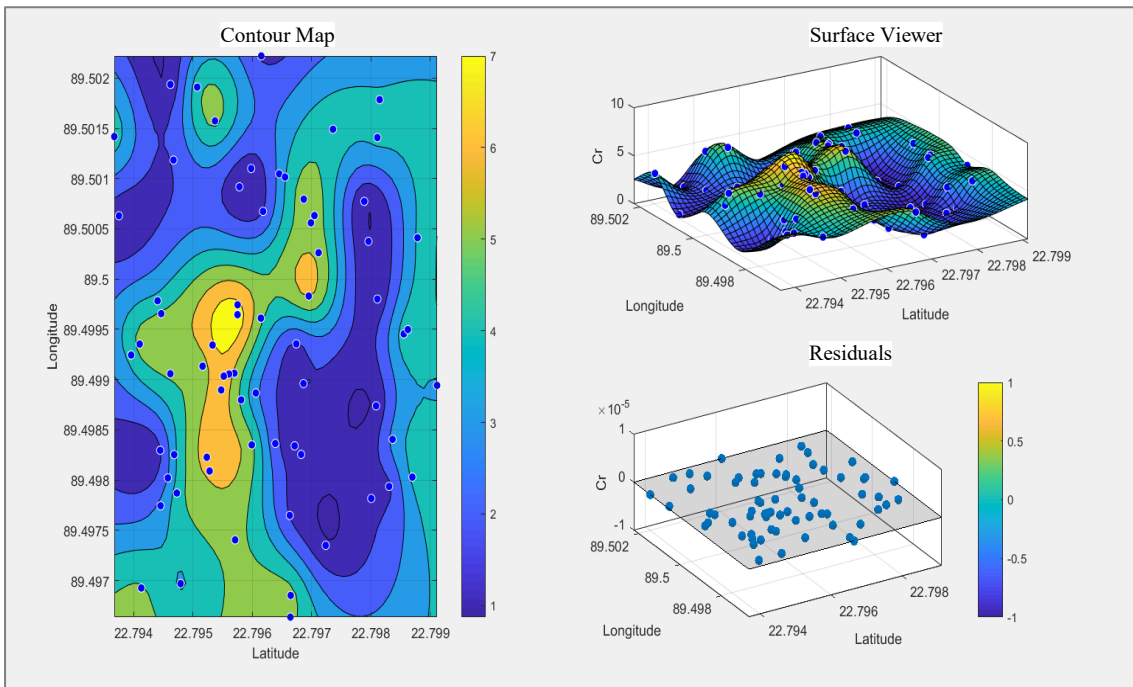
Figure C-8: Graphical representation of outputs for Cr.
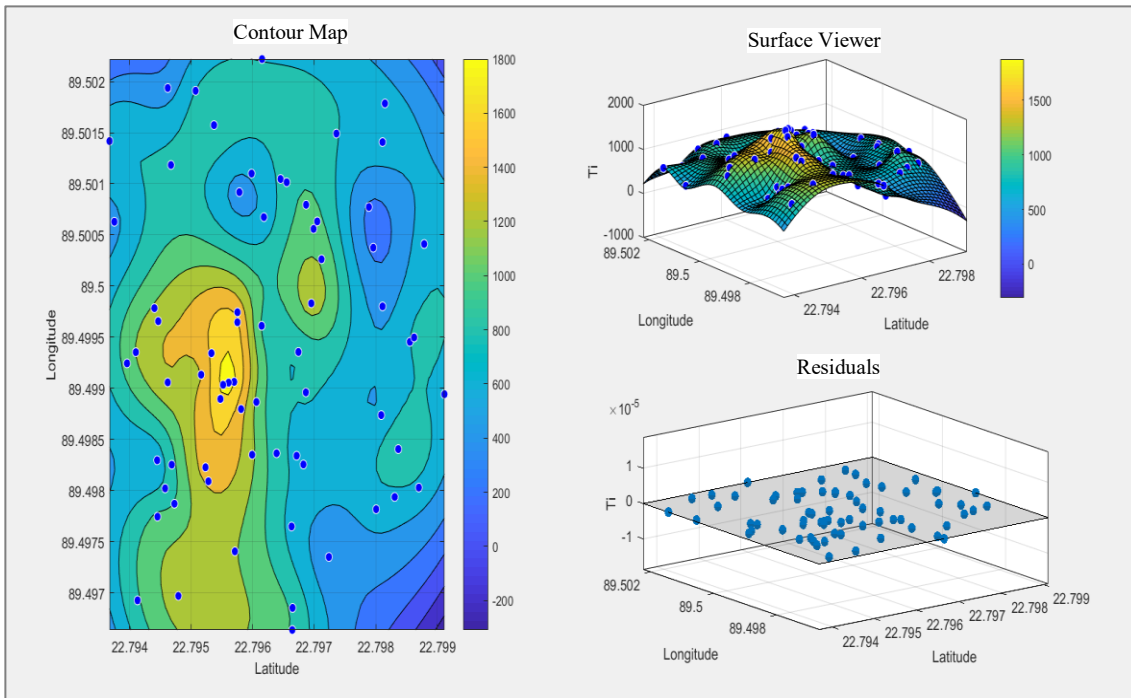
Figure C-9: Graphical representation of outputs for Ti.
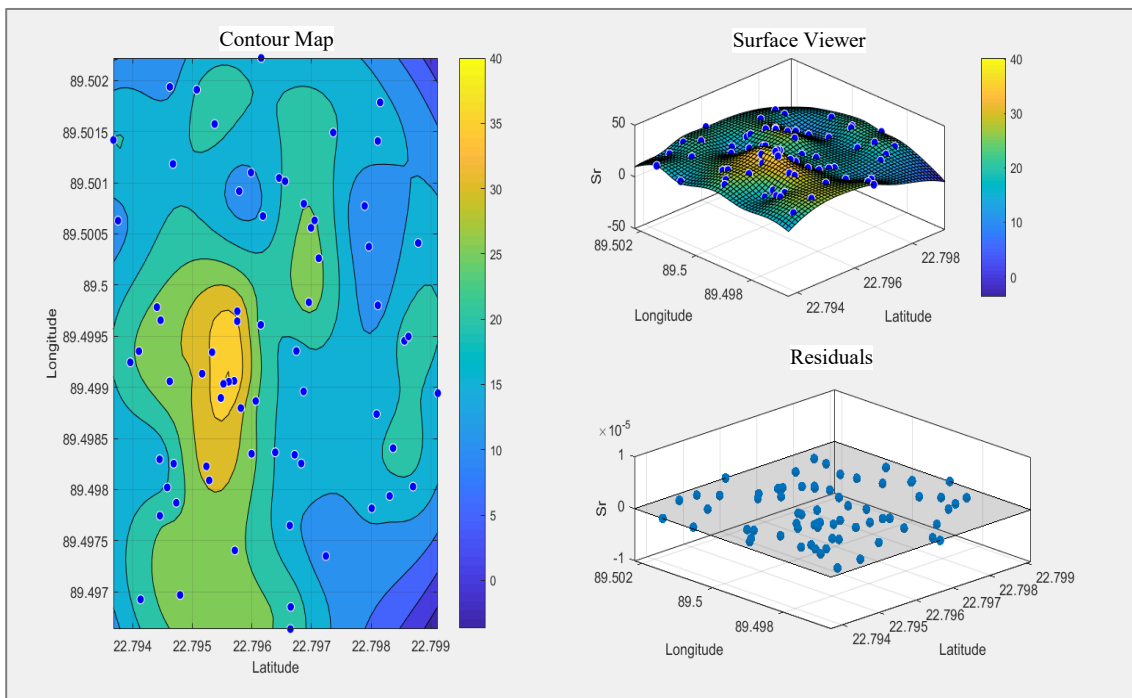
Figure C-10: Graphical representation of outputs for Sr.
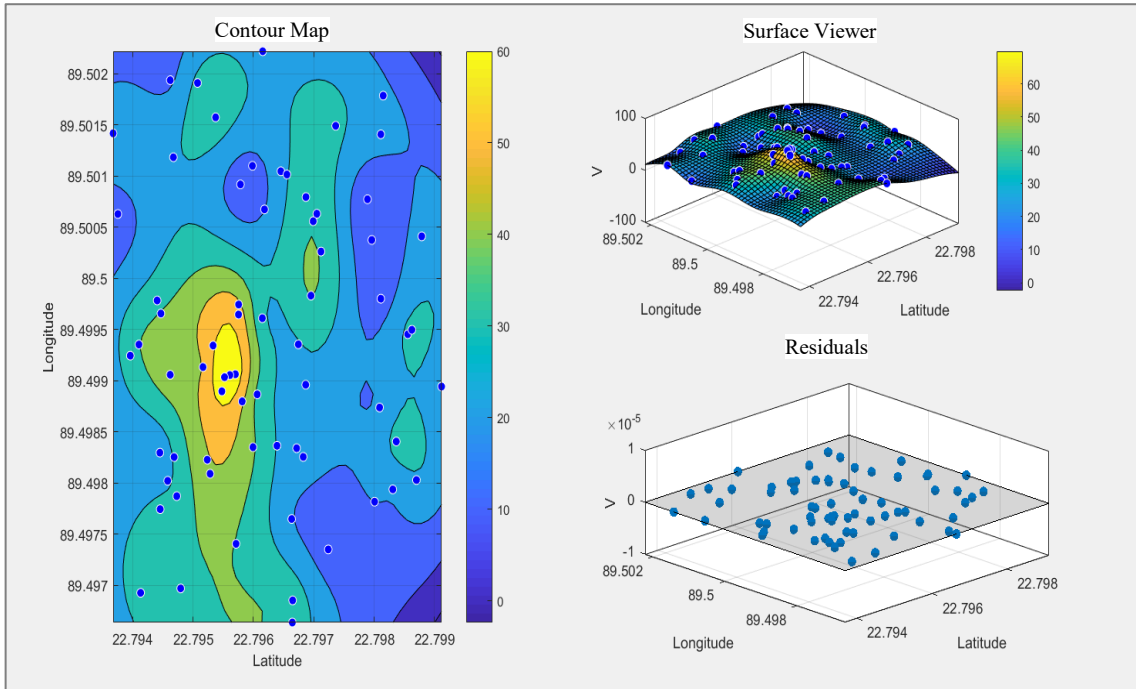
Figure C-11: Graphical representation of outputs for V.
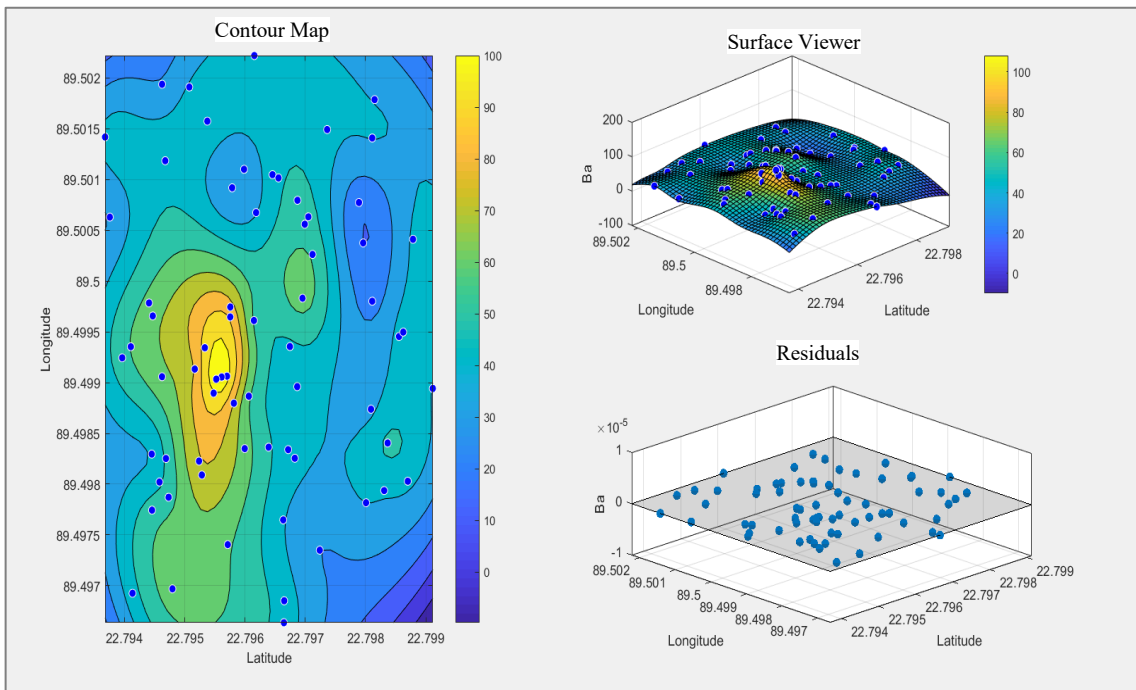
Figure C-12: Graphical representation of outputs for Ba.

214
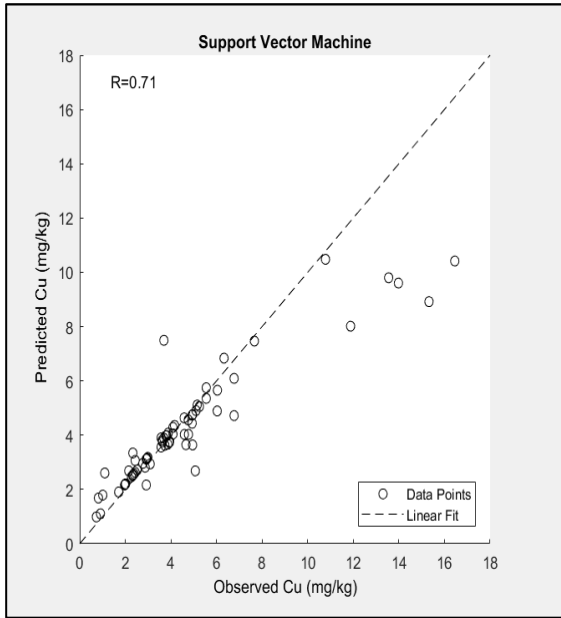
Figure C-13: Regression analysis for Cu in SVM (training).
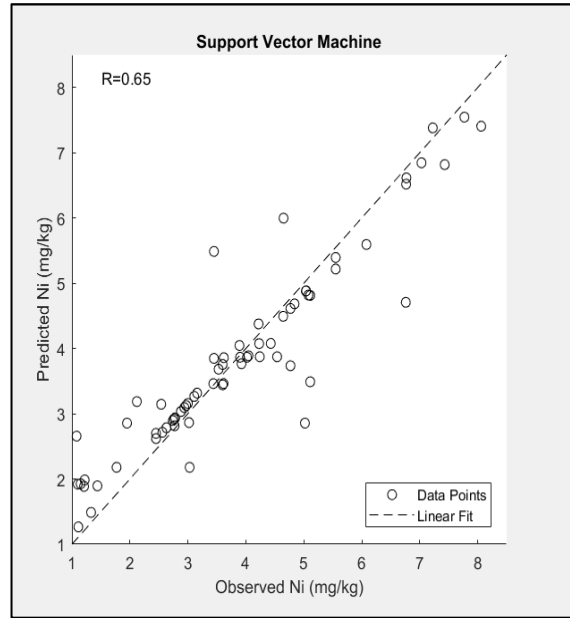
Figure C-14: Regression analysis for Ni in SVM (training).
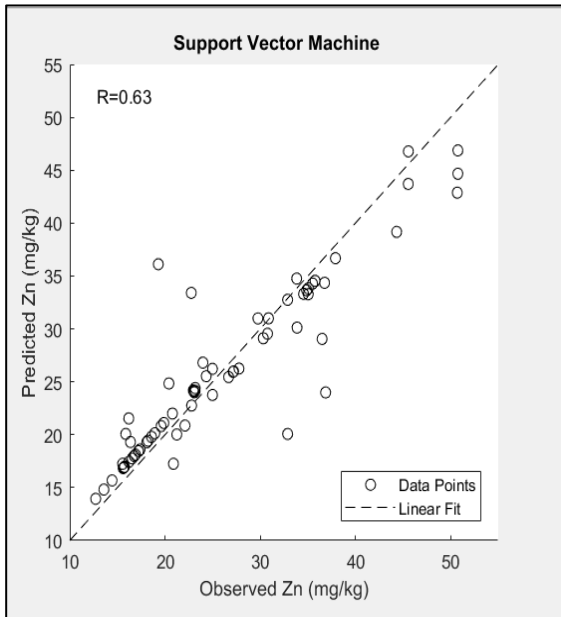
Figure C-15: Regression analysis for Zn in SVM (training).
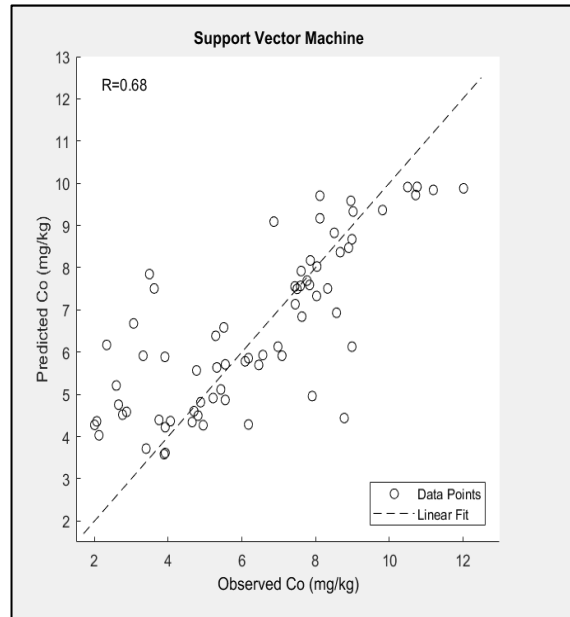
Figure C-16: Regression analysis for Co in SVM (training).
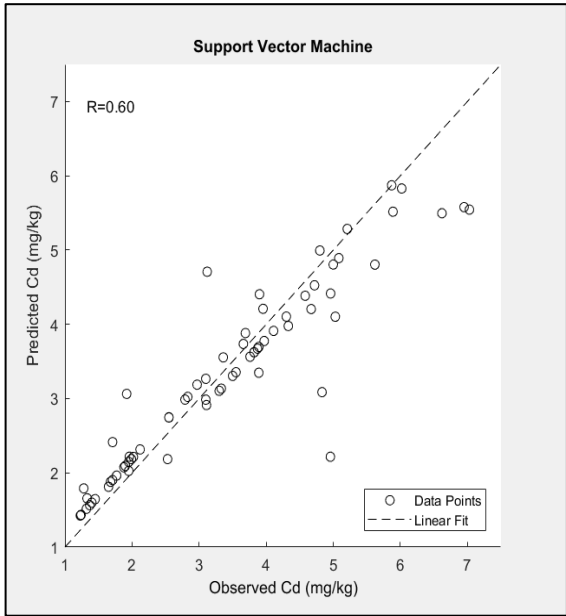
Figure C-17: Regression analysis for Cd in SVM (training).
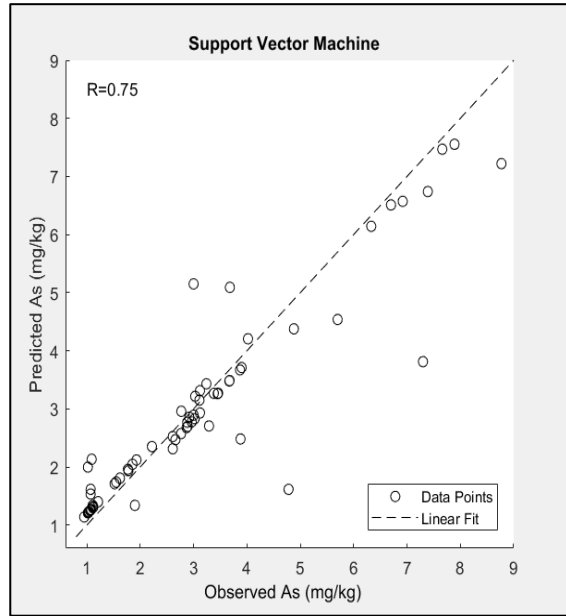
Figure C-18: Regression analysis for As in SVM (training).
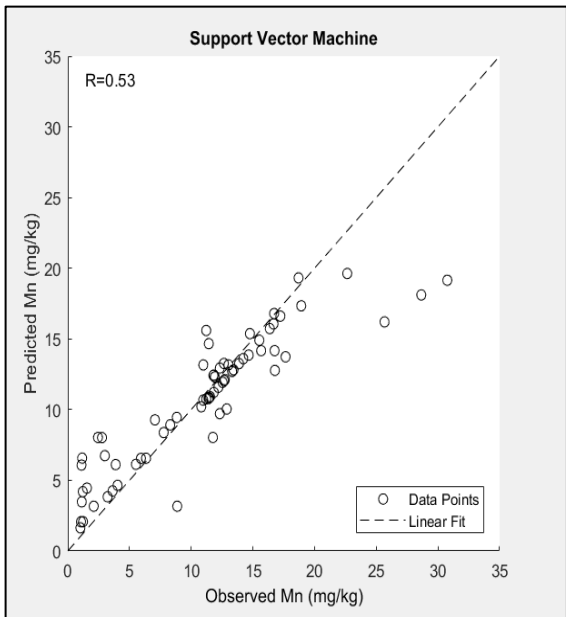
Figure C-19: Regression analysis for Mn in SVM (training).
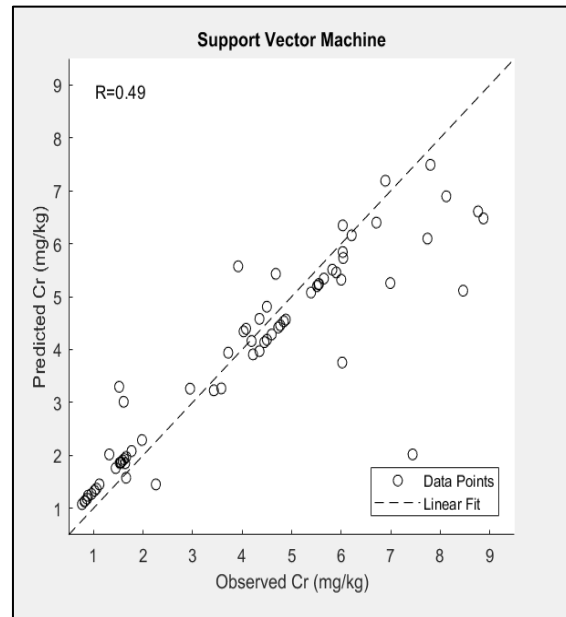
Figure C-20: Regression analysis for Cr in SVM (training).
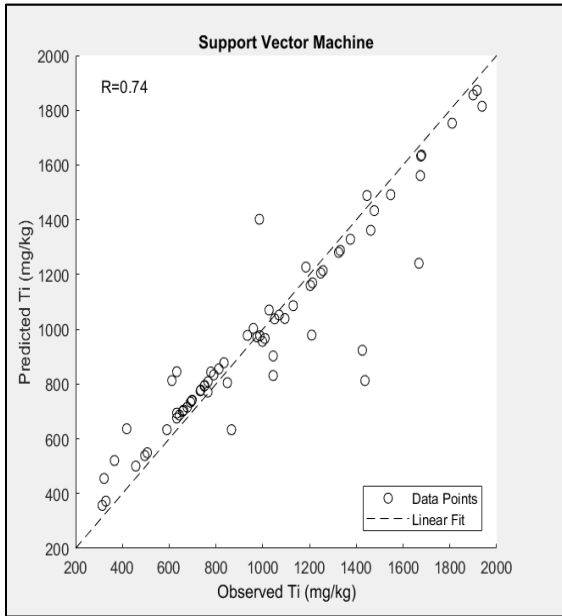
Figure C-21: Regression analysis for Ti in SVM (training).
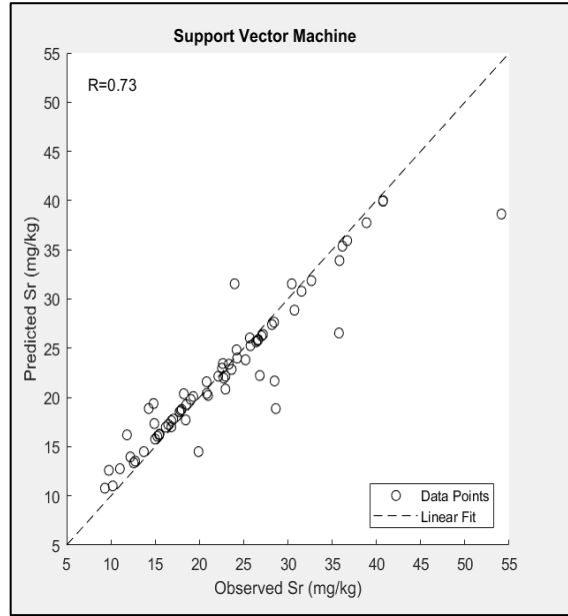
Figure C-22: Regression analysis for Sr in SVM (training).
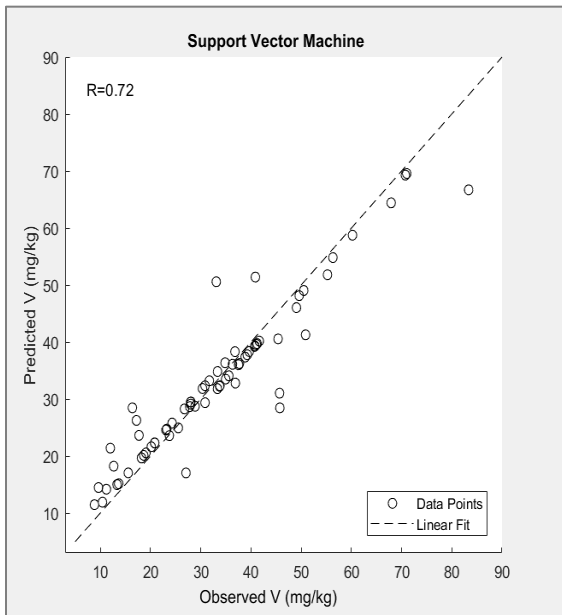
Figure C-23: Regression analysis for V in SVM (training).
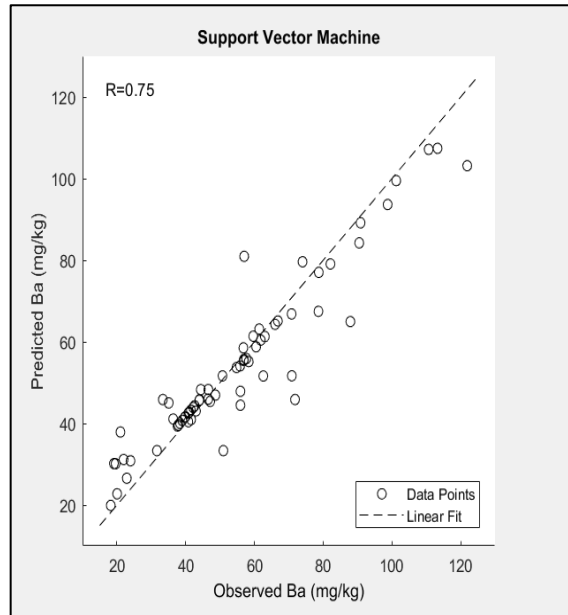
Figure C-24: Regression analysis for Ba in SVM (training).
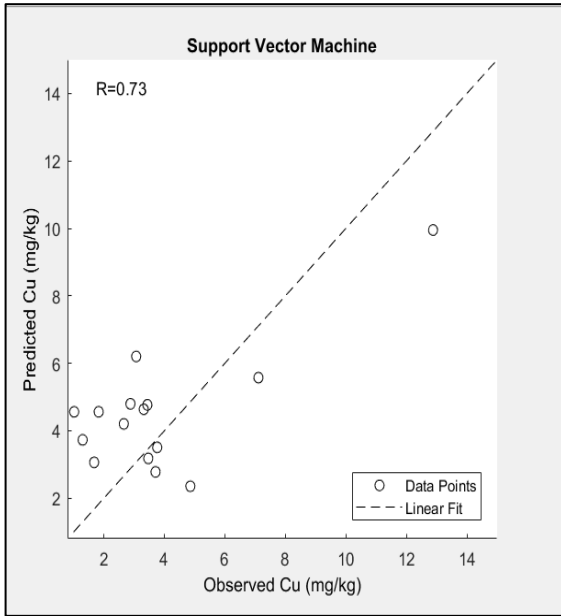
Figure C-25: Regression analysis for Cu in SVM (testing).
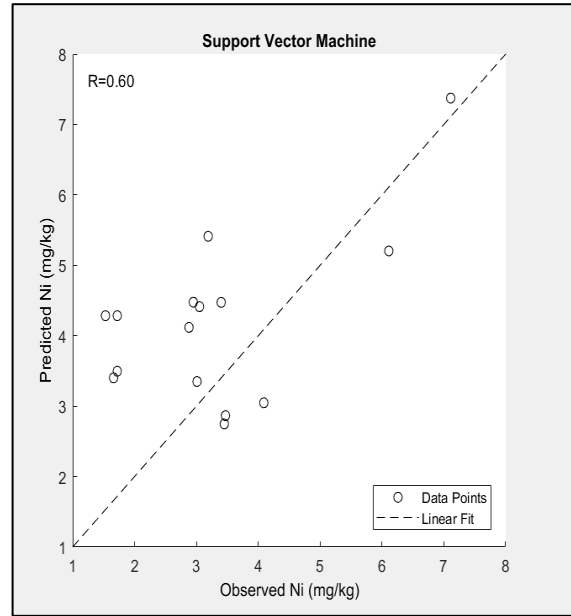
Figure C-26: Regression analysis for Ni in SVM (testing).
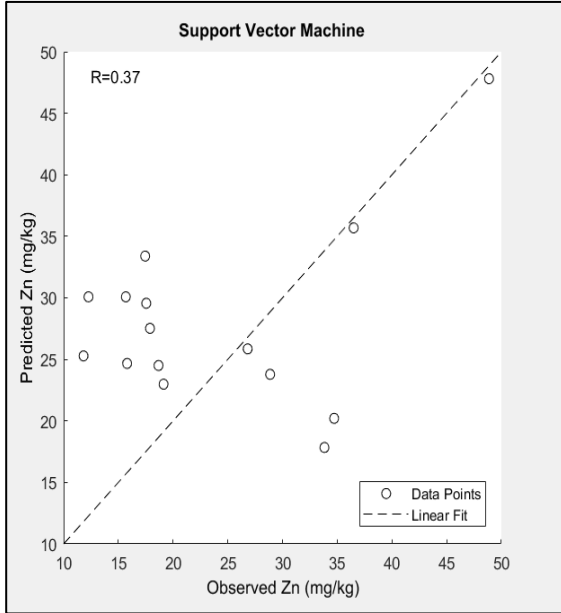
Figure C-27: Regression analysis for Zn in SVM (testing).
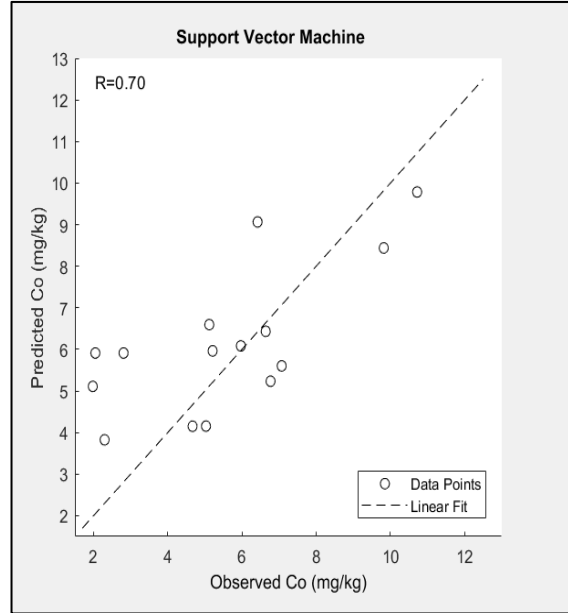
Figure C-28: Regression analysis for Co in SVM (testing).
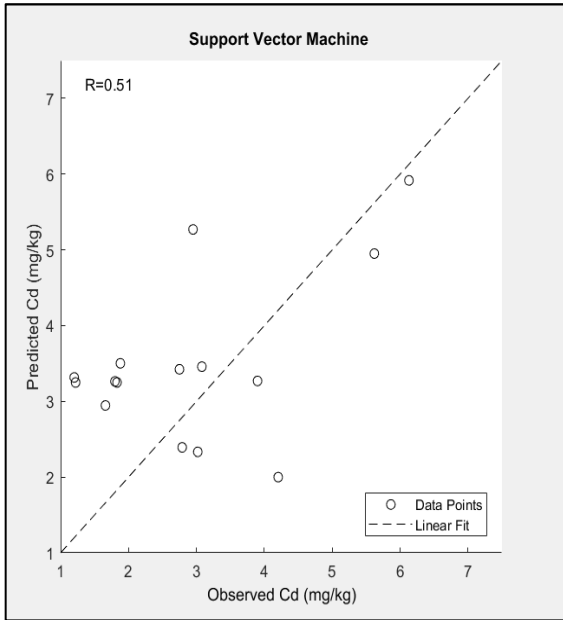
Figure C-29: Regression analysis for Cd in SVM (testing).
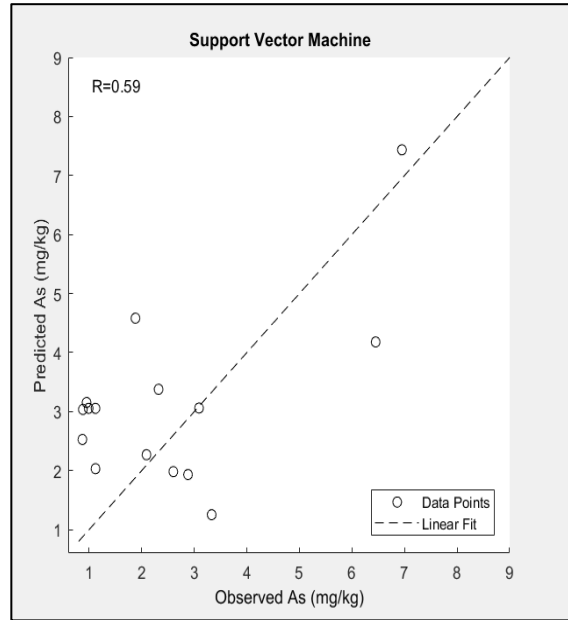
Figure C-30: Regression analysis for As in SVM (testing).
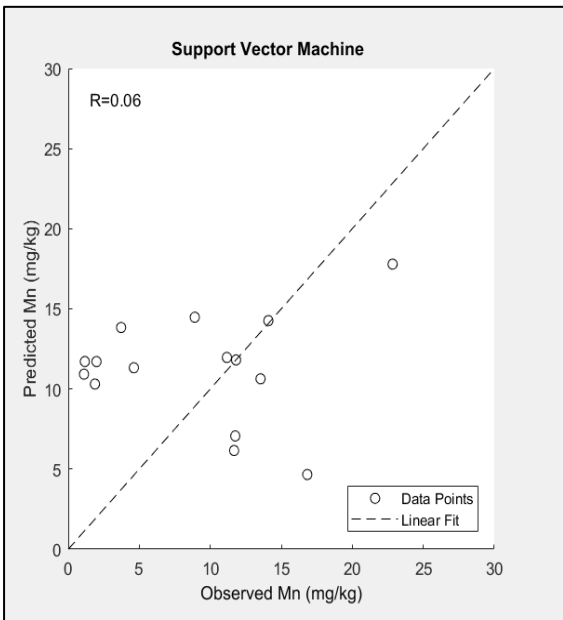
Figure C-31: Regression analysis for Mn in SVM (testing).
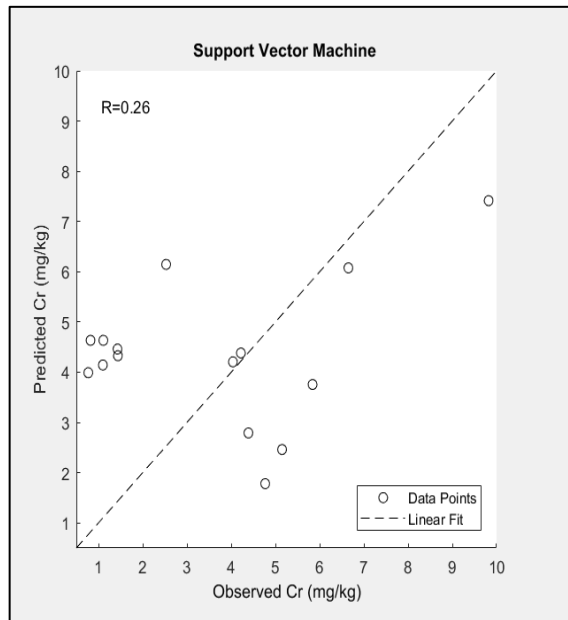
Figure C-32: Regression analysis for Cr in SVM (testing).
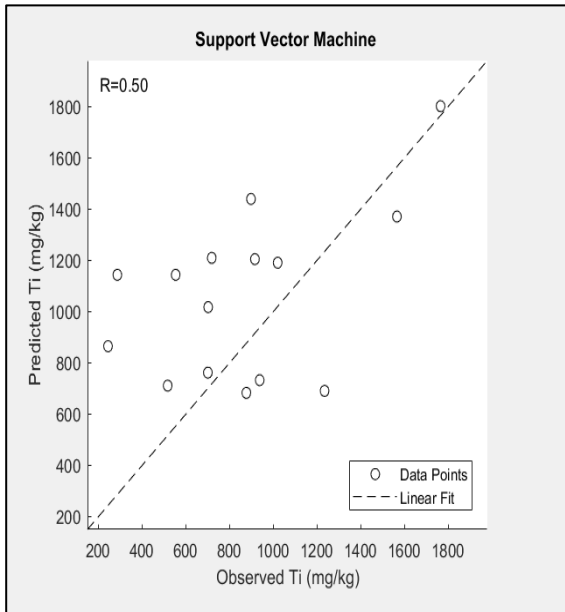
Figure C-33: Regression analysis for Ti in SVM (testing).
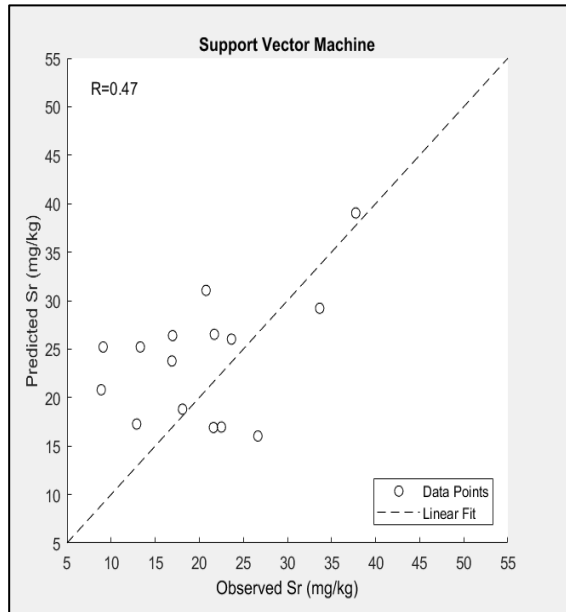
Figure C-34: Regression analysis for Sr in SVM (testing).
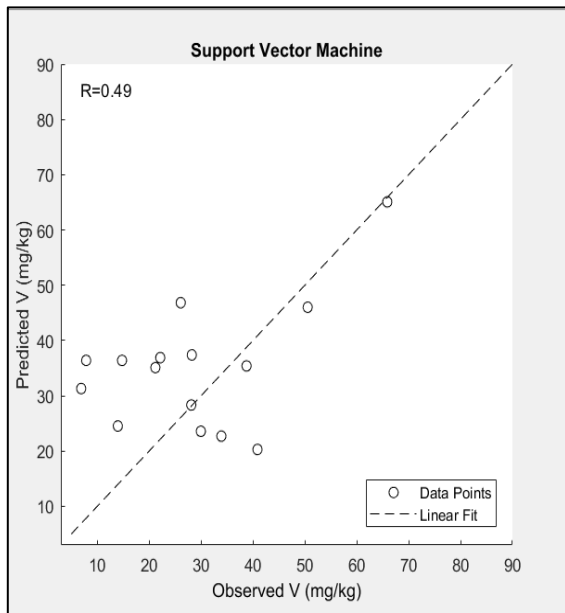
Figure C-35: Regression analysis for V in SVM (testing).
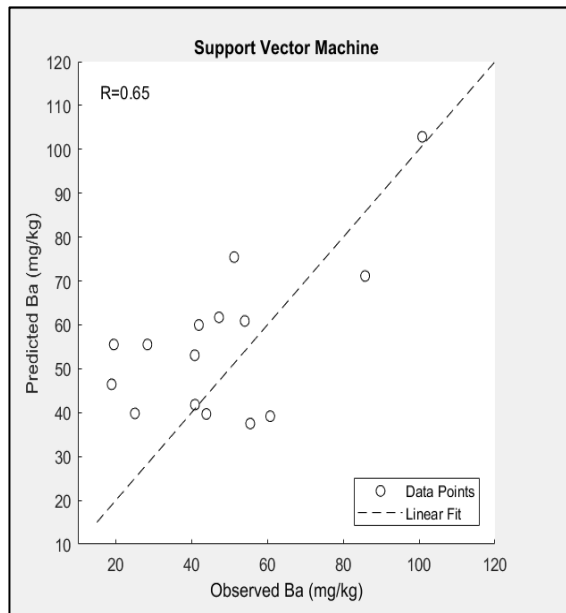
Figure C-36: Regression analysis for Ba in SVM (testing).
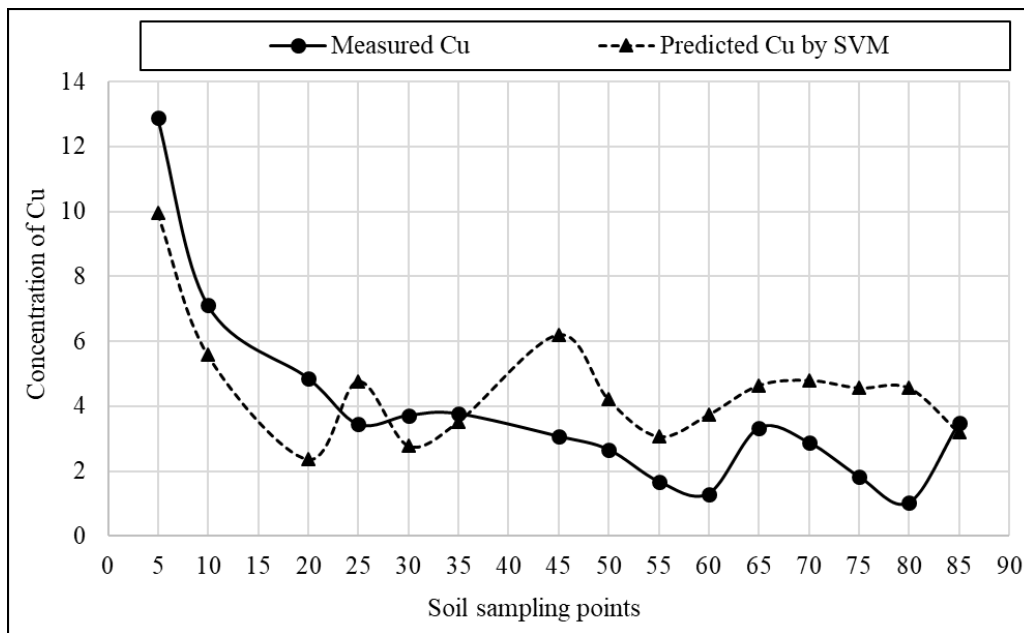
Figure C-37: Comparison of predicted and measured concentration of Cu from SVM (testing).
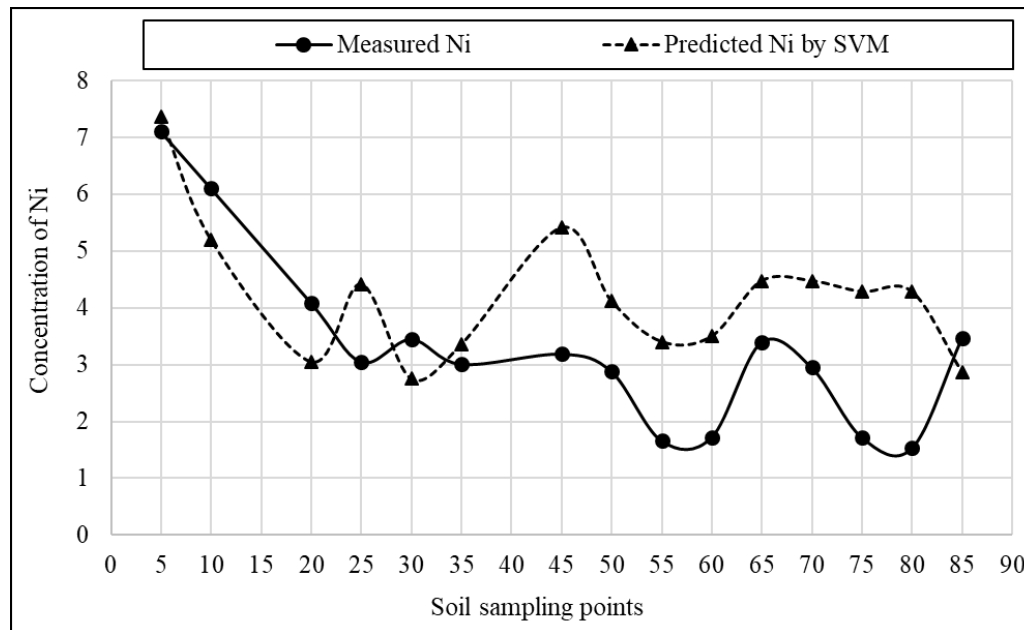
Figure C-38: Comparison of predicted and measured concentration of Ni from SVM (testing).
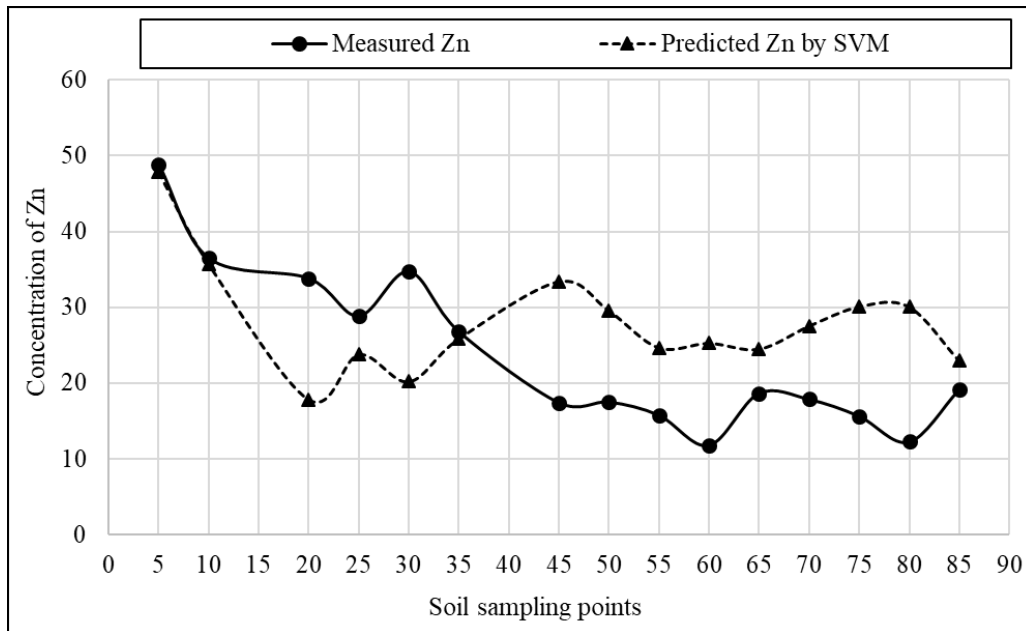
Figure A-39: Comparison of predicted and measured concentration of Zn from
SVM (testing).
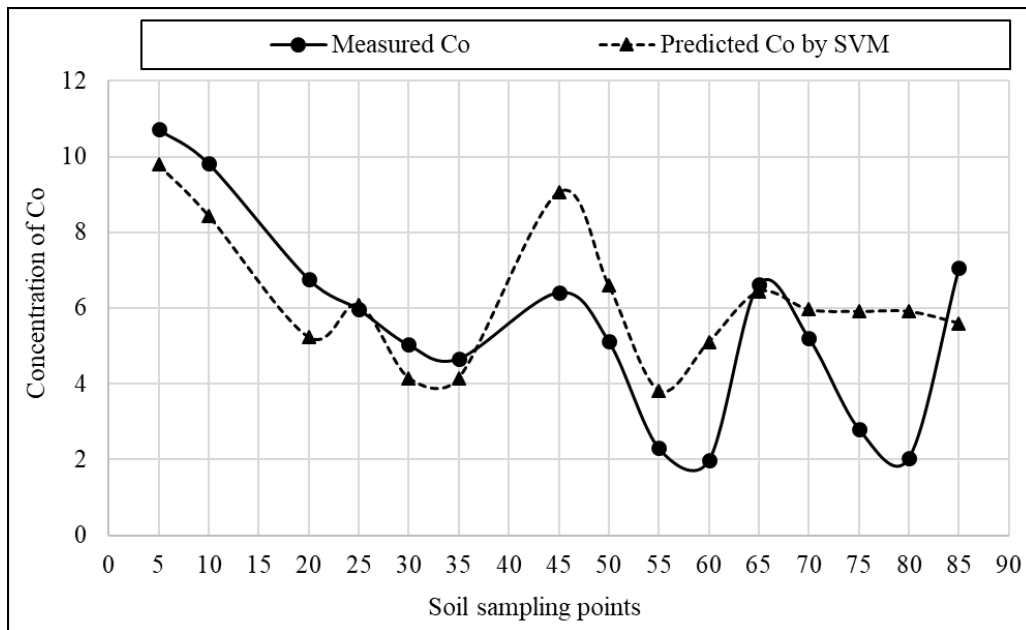
Figure C-40: Comparison of predicted and measured concentration of Co from
SVM (testing).
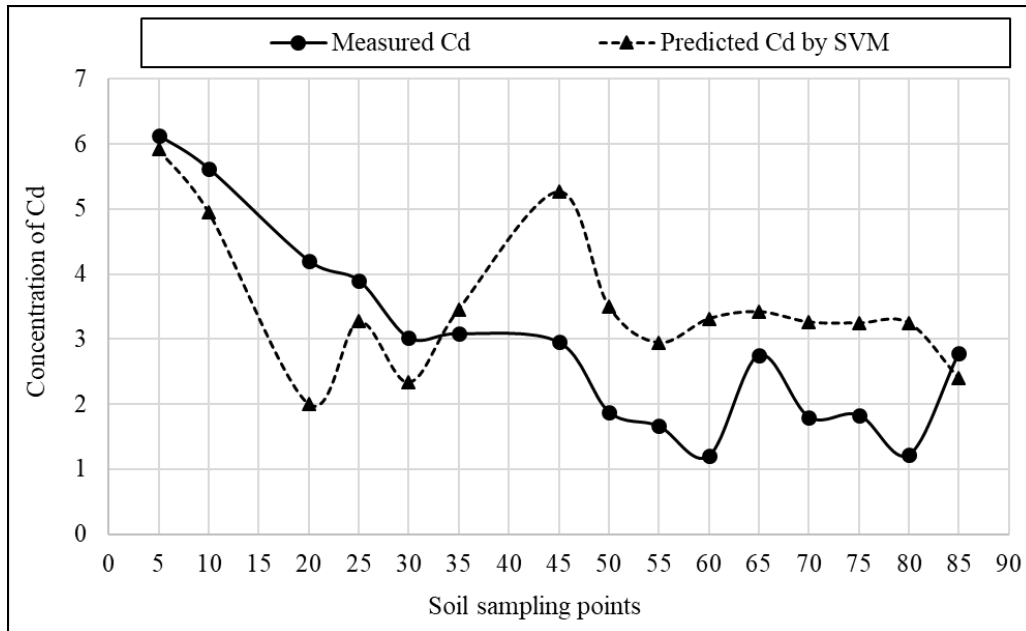
Figure C-41: Comparison of predicted and measured concentration of Cd from SVM (testing).
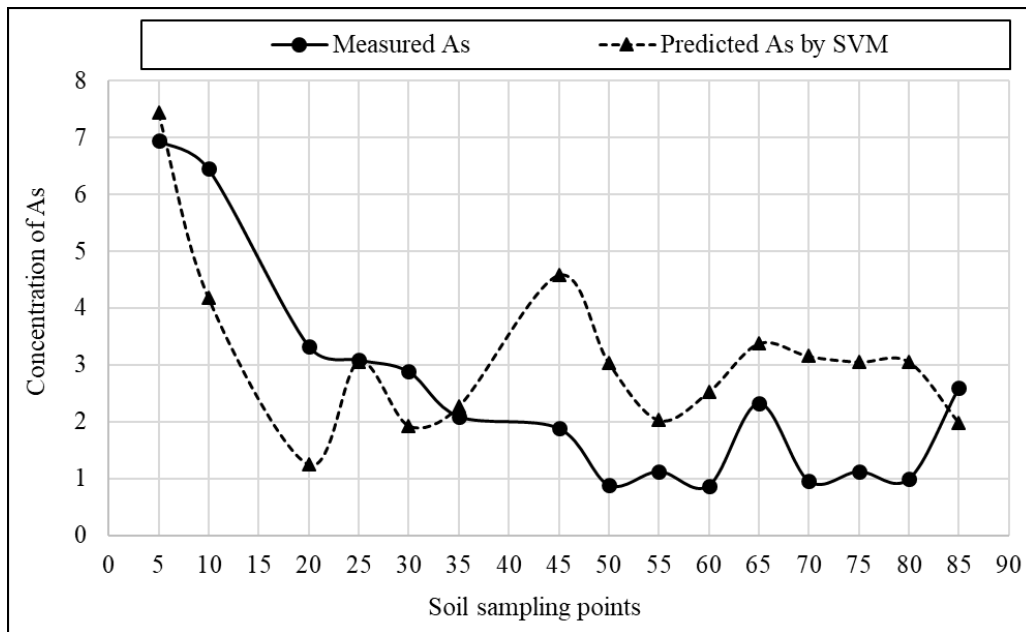
Figure C-42: Comparison of predicted and measured concentration of As from SVM (testing).
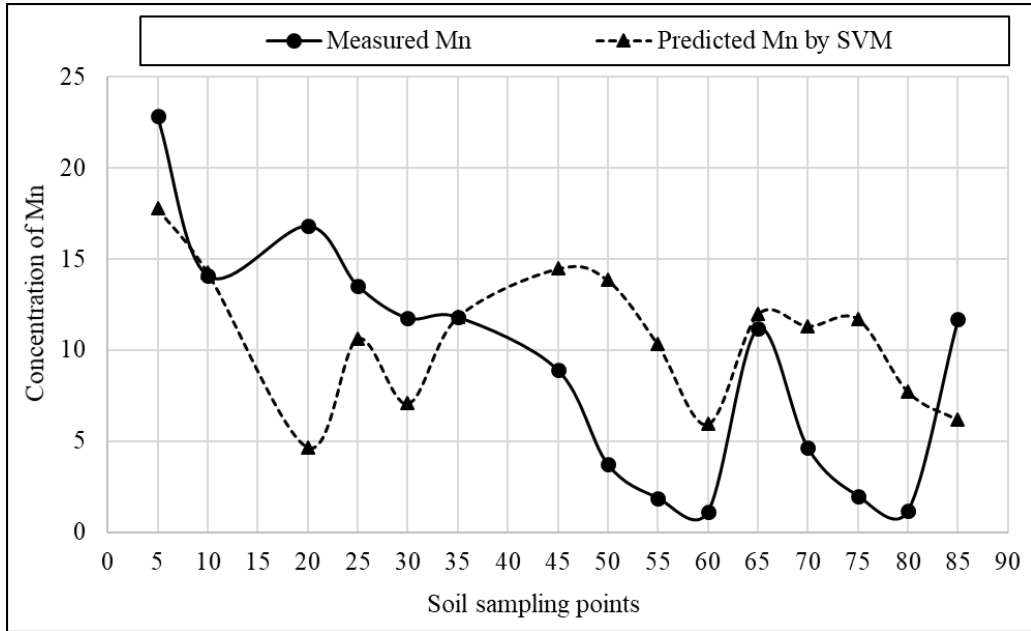
Figure C-43: Comparison of predicted and measured concentration of Mn from SVM (testing).
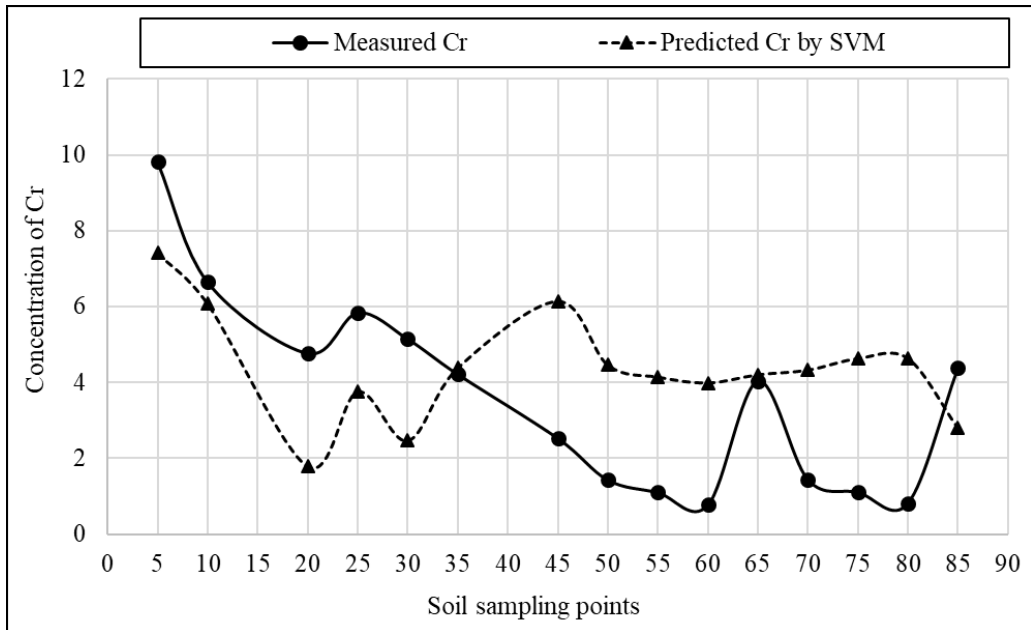
Figure C-44: Comparison of predicted and measured concentration of Cr from SVM (testing).
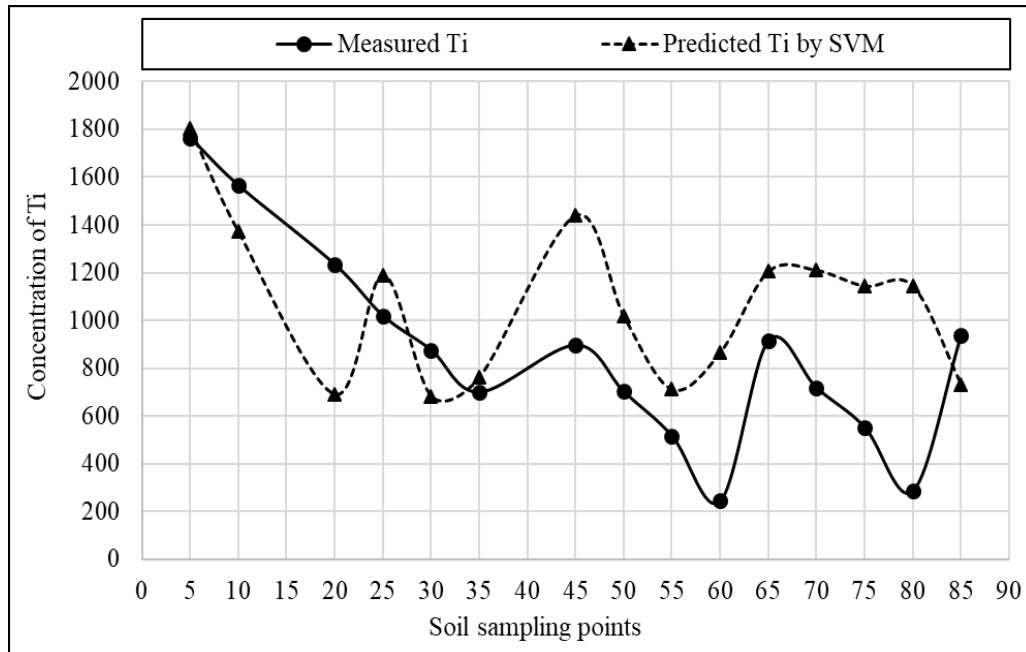
Figure C-45: Comparison of predicted and measured concentration of Ti from SVM (testing).
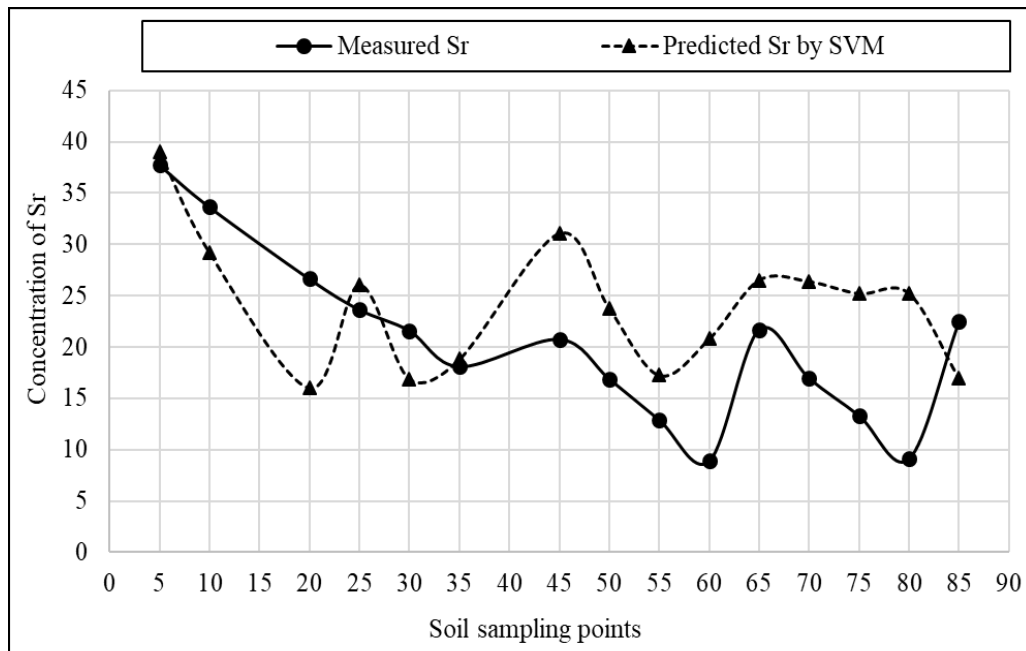
Figure C-46: Comparison of predicted and measured concentration of Sr from SVM (testing).
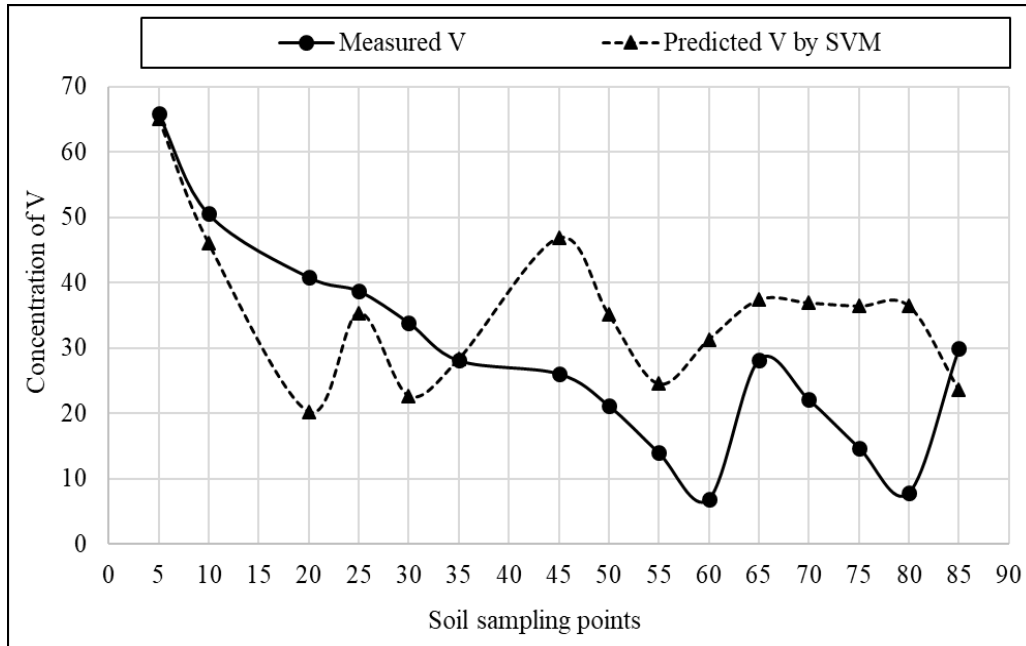
Figure C-47: Comparison of predicted and measured concentration of V from SVM (testing).
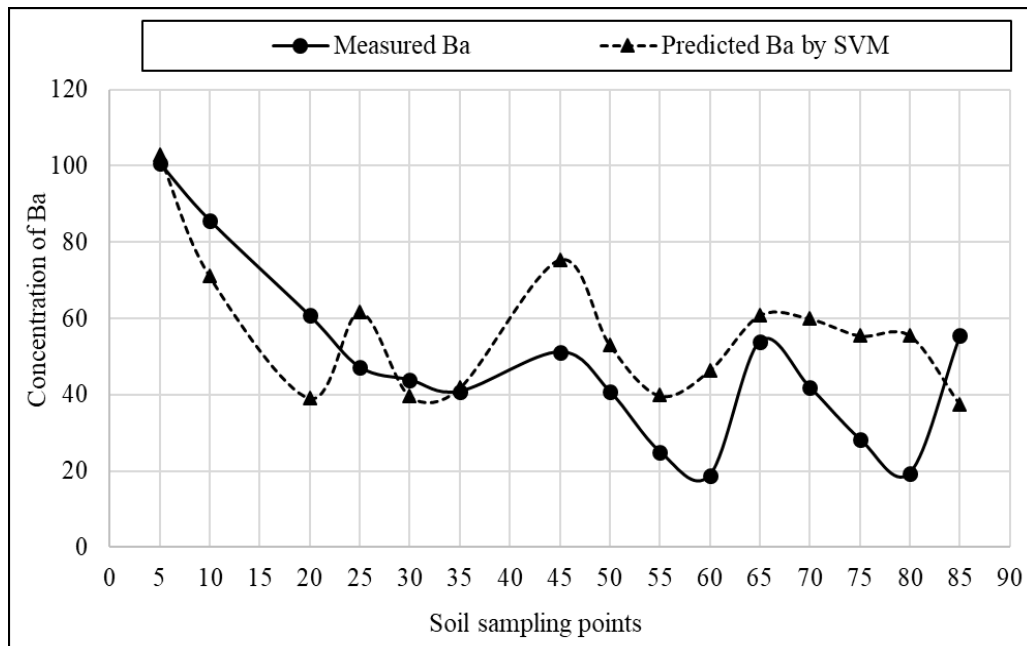
Figure C-48: Comparison of predicted and measured concentration of Ba from SVM (testing).
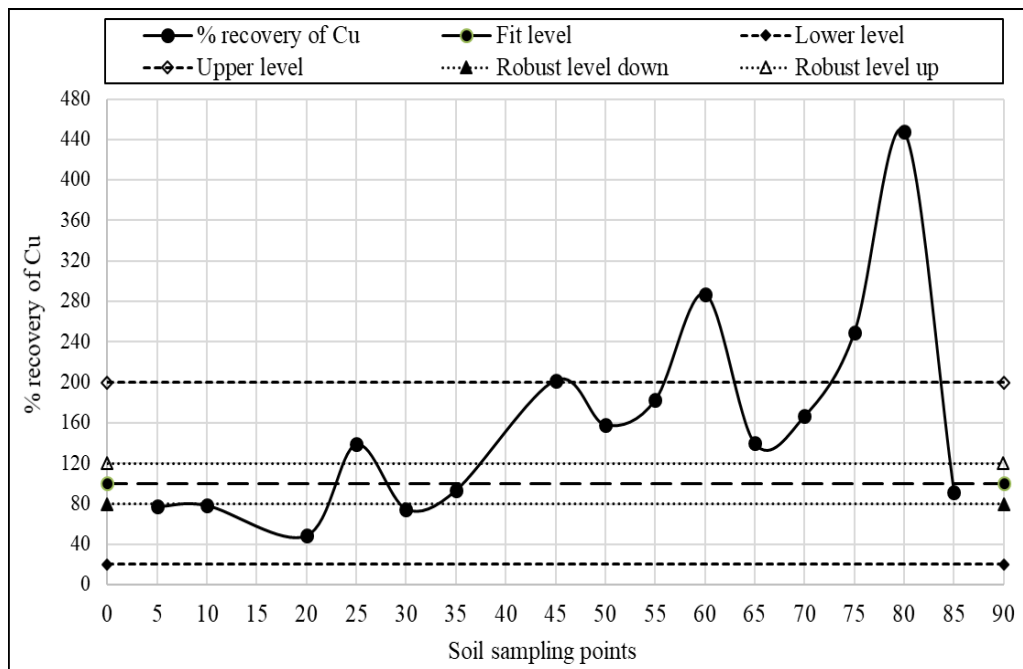
Figure C-49: Variation of recovery level for Cu in SVM.
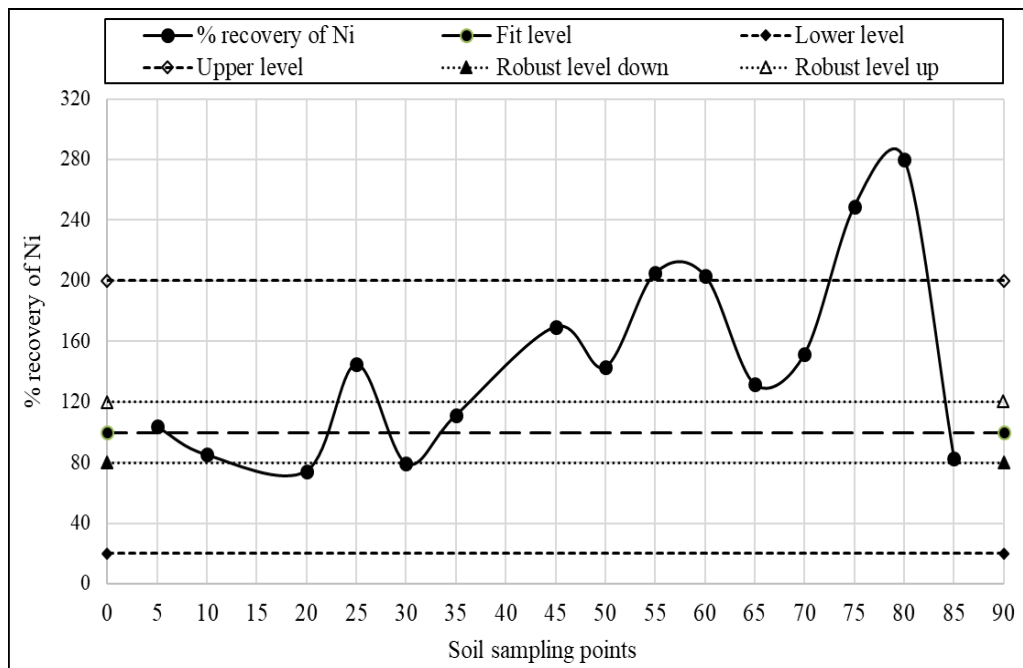
Figure C-50: Variation of recovery level for Ni in SVM.
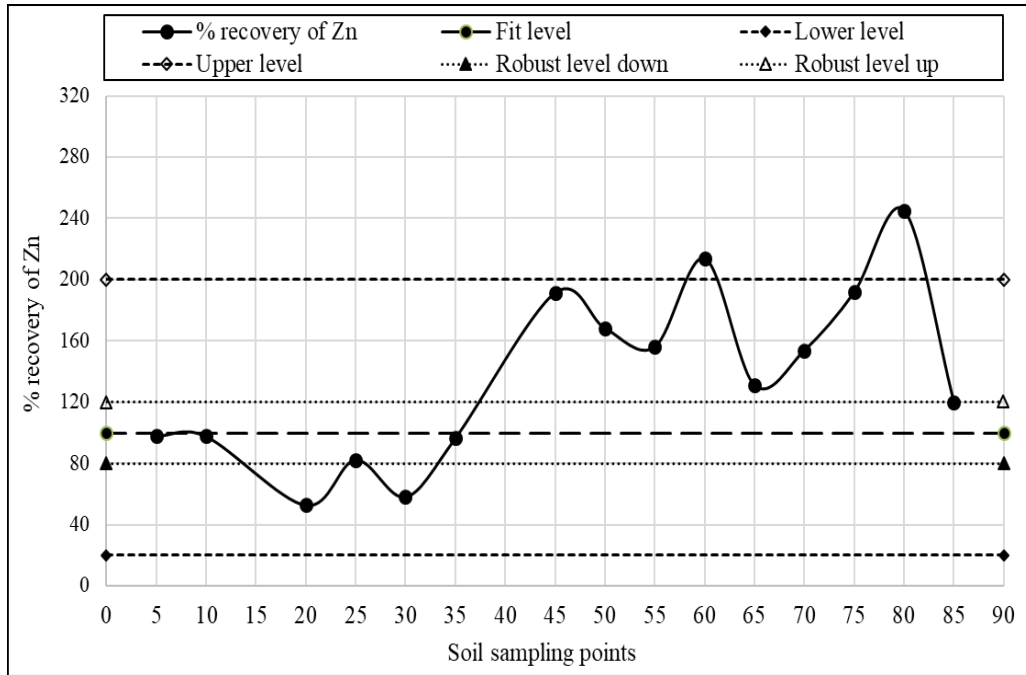
Figure C-51: Variation of recovery level for Zn in SVM.
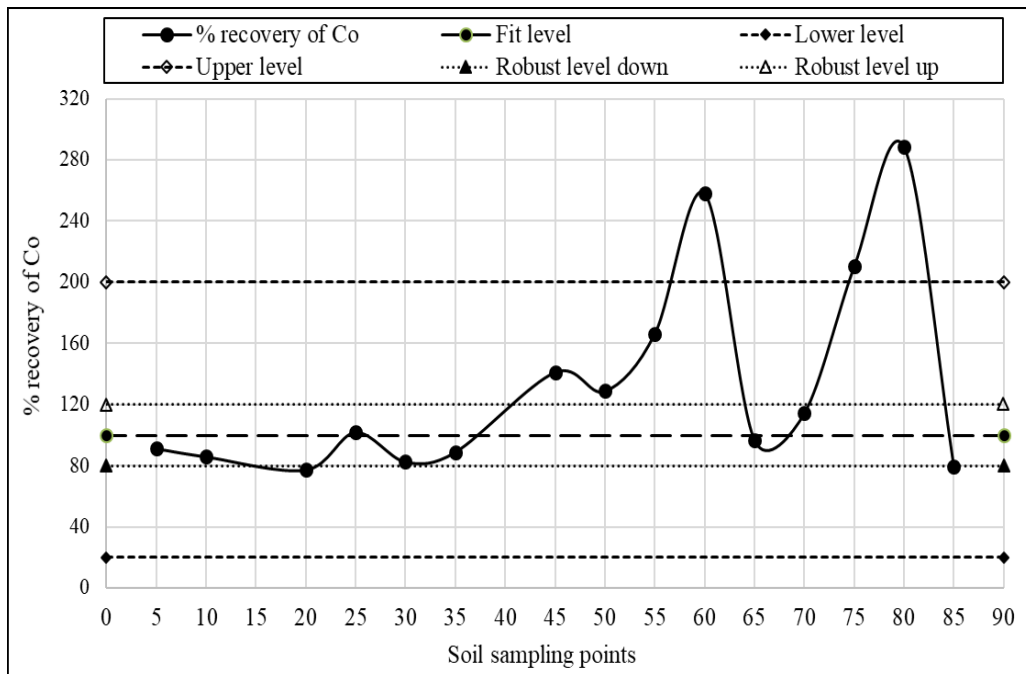
Figure C-52: Variation of recovery level for Co in SVM.
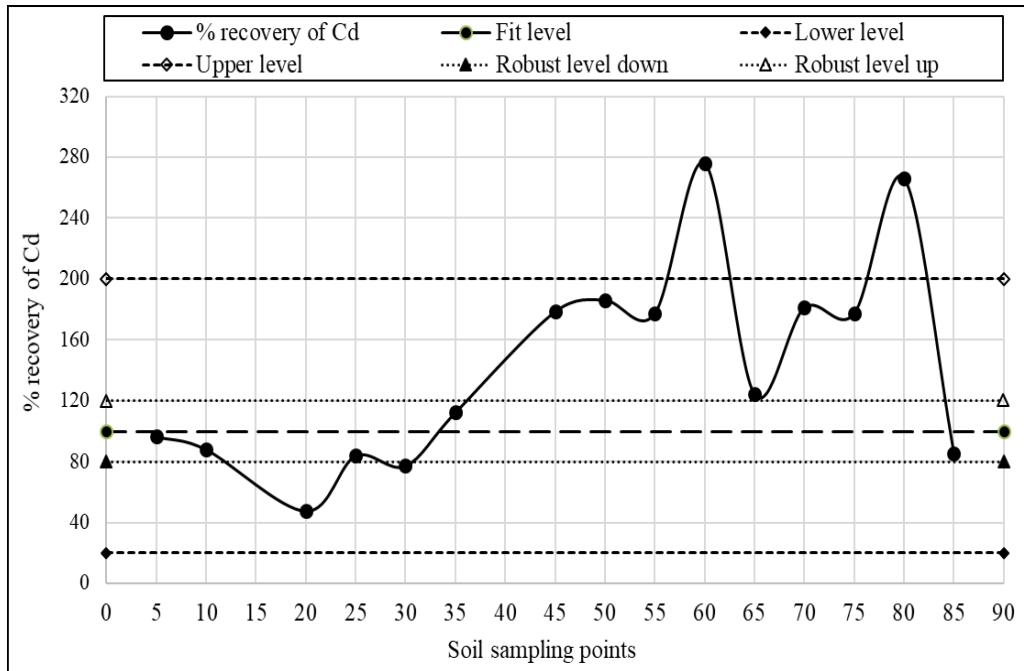
Figure C-53: Variation of recovery level for Cd in SVM.
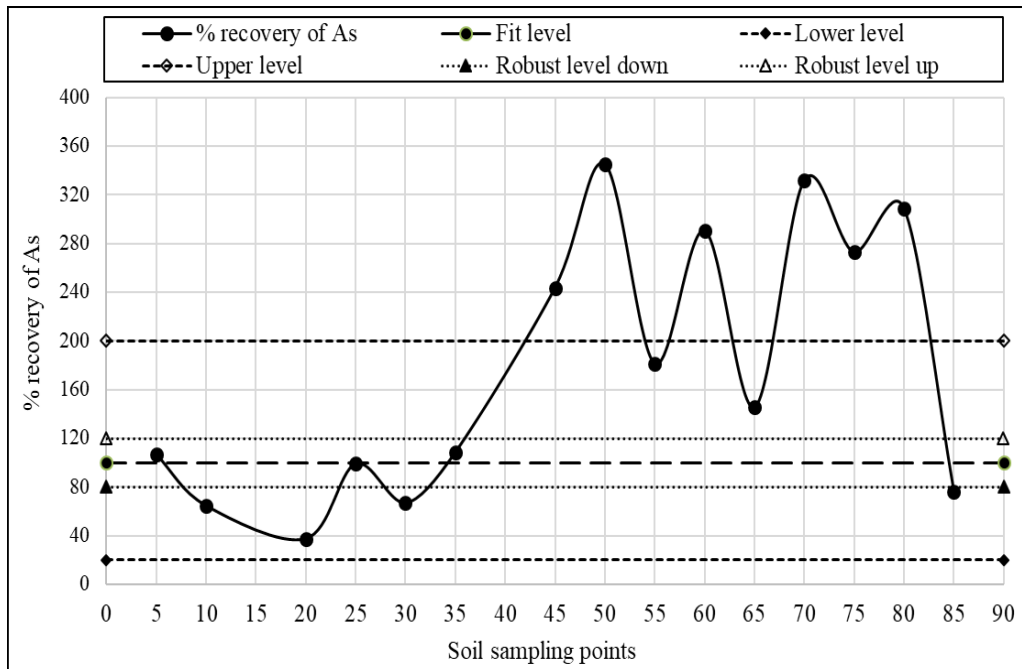
Figure C-54: Variation of recovery level for As in SVM.
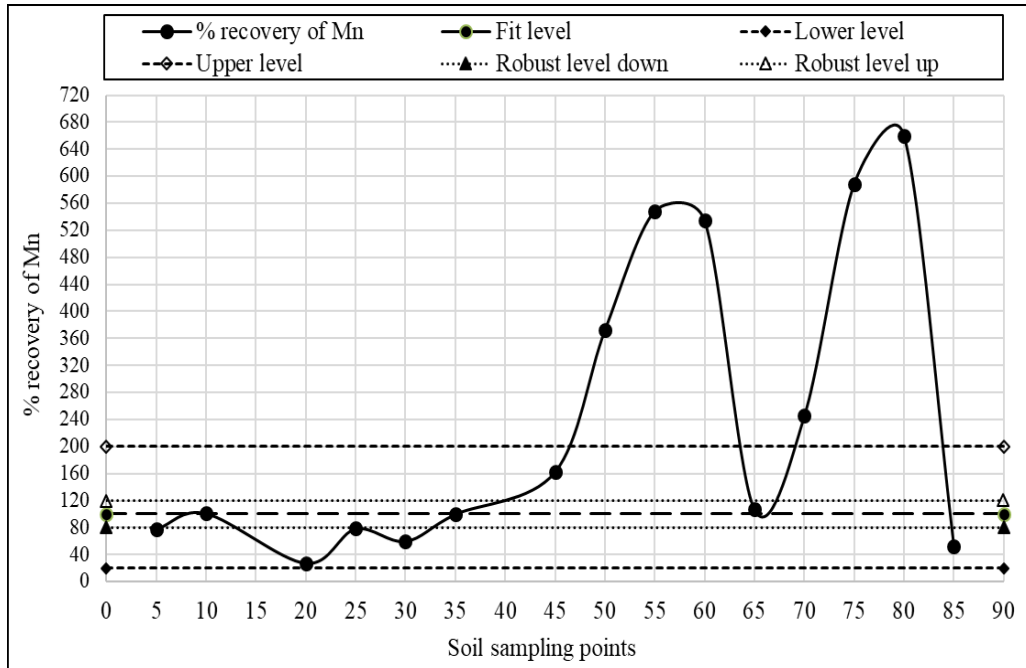
Figure C-55: Variation of recovery level for Mn in SVM.
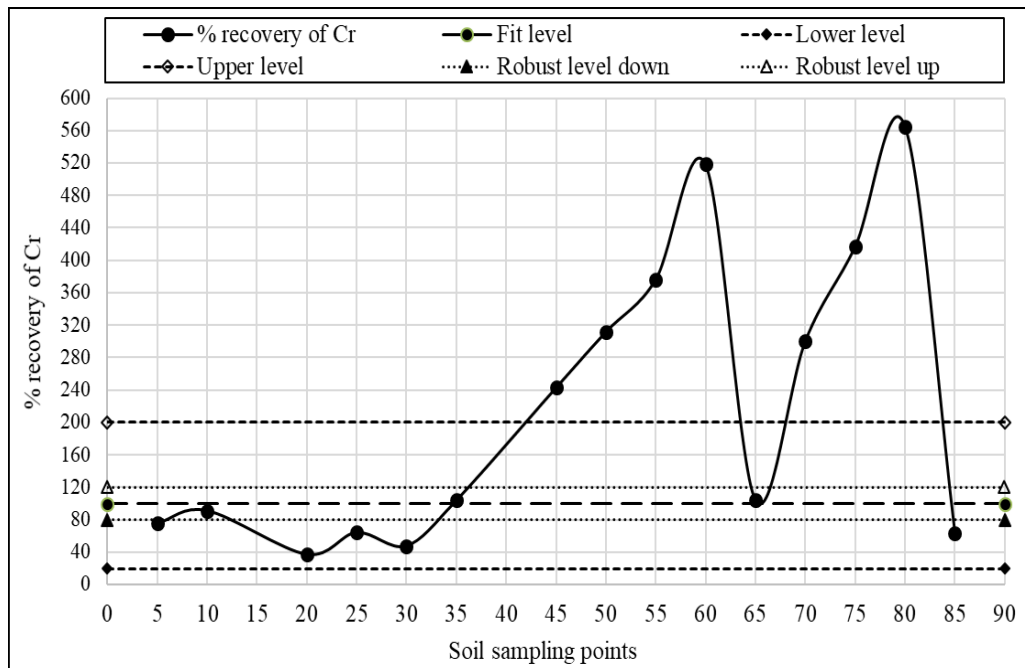
Figure C-56: Variation of recovery level for Cr in SVM.
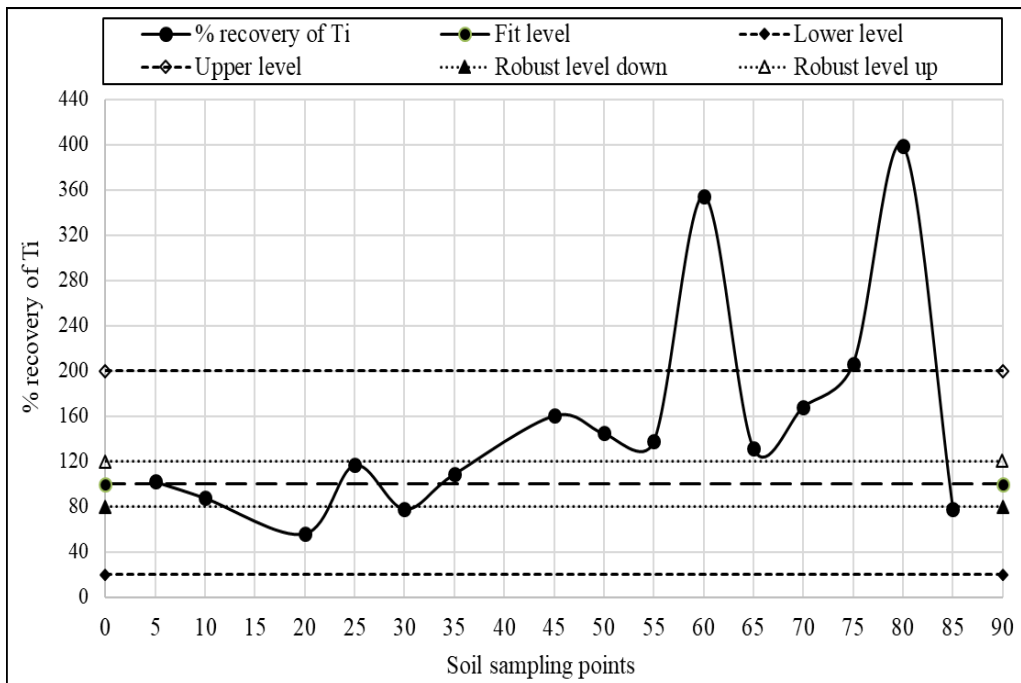
Figure C-57: Variation of recovery level for Ti in SVM.
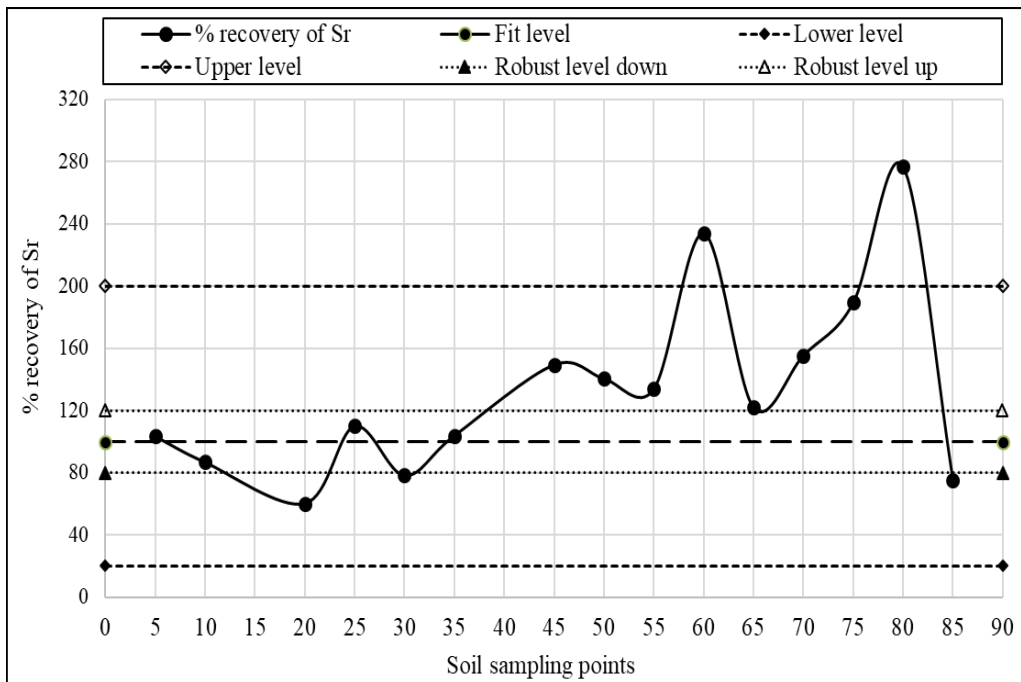
Figure C-58: Variation of recovery level for Sr in SVM.

231
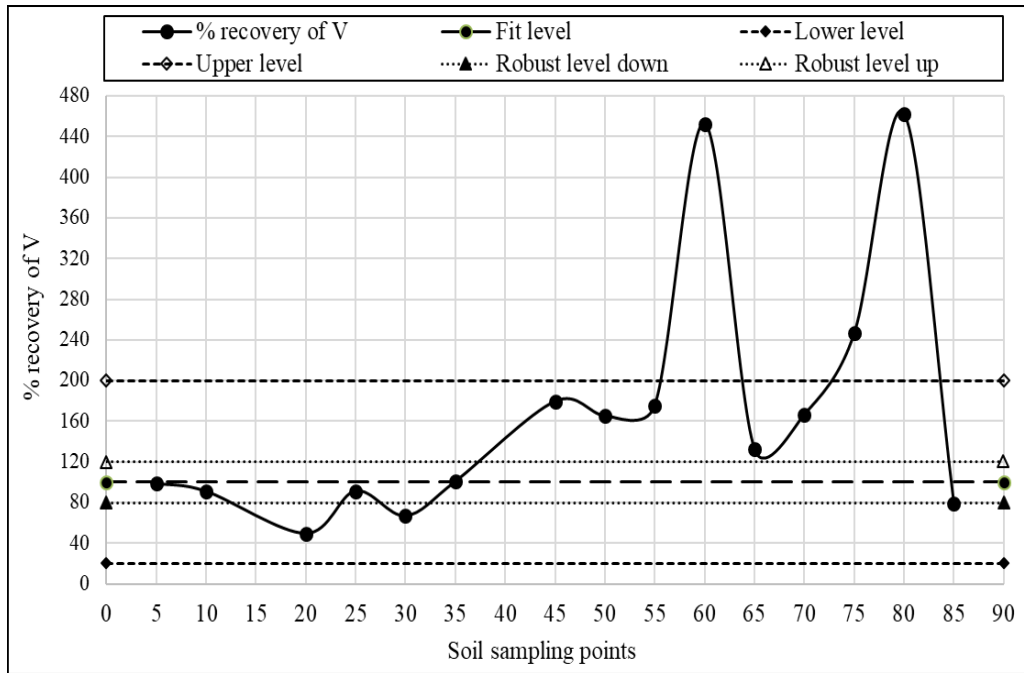
Figure C-59: Variation of recovery level for V in SVM.
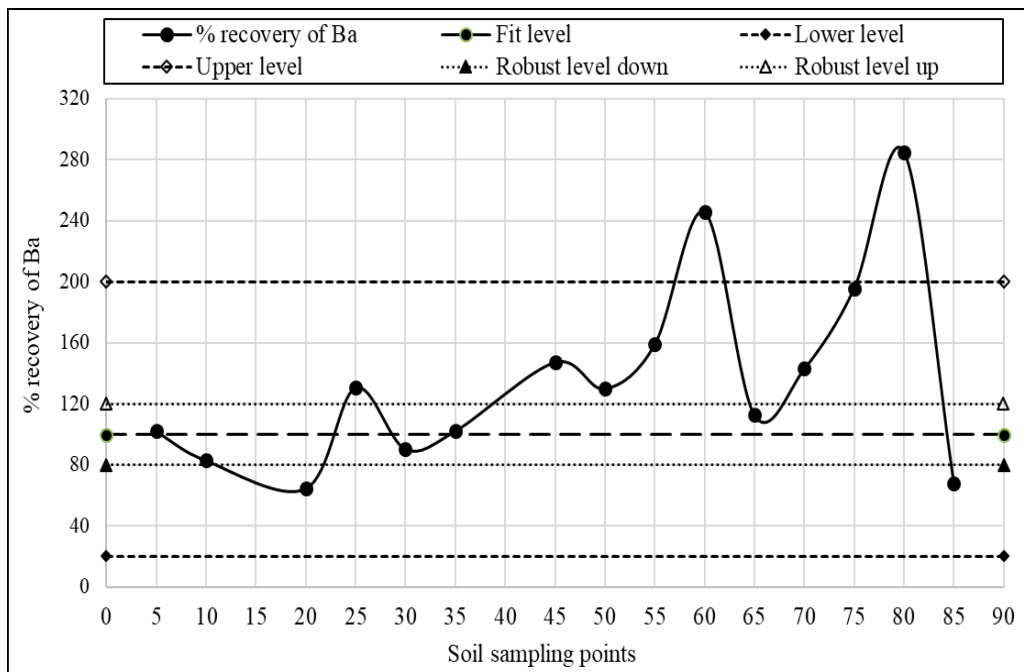
Figure C-60: Variation of recovery level for Ba in SVM.

Appendix-D

Recovery Level for Various Heavy Metals of SVM

Table D-1: Recovery level of Cu in SVM (testing)

Soils sampling points	Cu		
	Measured	Predicted	% recovery
5	12.88	9.95	77.27
10	7.11	5.58	78.42
20	4.86	2.36	48.49
25	3.44	4.77	138.61
30	3.71	2.78	74.94
35	3.765	3.51	93.19
45	3.07	6.21	202.12
50	2.66	4.20	158.07
55	1.68	3.06	182.31
60	1.3	3.73	287.13
65	3.32	4.64	139.69
70	2.88	4.80	166.66
75	1.83	4.56	249.34
80	1.02	4.56	447.34
85	3.47	3.18	91.72
% mean recovery			162.35
RMSE			2.09
MAPE			80.48
GRI			1.86

Table D-2: Recovery level of Ni in SVM (testing)

Soils sampling points	Ni		
	Measured	Predicted	% recovery
5	7.11	7.37	103.71
10	6.11	5.20	85.16
20	4.09	3.05	74.51
25	3.05	4.41	144.70
30	3.45	2.75	79.64
35	3.01	3.35	111.28
45	3.19	5.41	169.64
50	2.88	4.12	143.01
55	1.66	3.40	204.96
60	1.72	3.50	203.29
65	3.4	4.47	131.58
70	2.95	4.48	151.72
75	1.72	4.28	249.12
80	1.53	4.28	280.06
85	3.47	2.87	82.58
% mean recovery			147.66
RMSE			1.53
MAPE			58.08
GRI			1.64

Table D-3: Recovery level for Zn in SVM (testing)

Soils sampling points	Zn		
	Measured	Predicted	% recovery
5	48.87	47.81	97.82
10	36.5	35.68	97.77
20	33.83	17.84	52.74
25	28.87	23.78	82.38
30	34.72	20.21	58.21
35	26.815	25.85	96.41
45	17.45	33.39	191.37
50	17.55	29.56	168.41
55	15.8	24.67	156.14
60	11.82	25.28	213.89
65	18.67	24.50	131.23
70	17.89	27.52	153.85
75	15.68	30.08	191.85
80	12.27	30.08	245.17
85	19.13	22.99	120.15
% mean recovery			137.16
RMSE			11.01
MAPE			52.45
GRI			1.63

Table D-4: Recovery level of Co in SVM (testing)

Soils sampling points	Co		
	Measured	Predicted	% recovery
5	10.72	9.78	91.28
10	9.82	8.44	85.94
20	6.77	5.23	77.33
25	5.97	6.08	101.91
30	5.03	4.15	82.60
35	4.67	4.15	88.91
45	6.42	9.06	141.20
50	5.12	6.60	128.83
55	2.3	3.83	166.33
60	1.98	5.11	258.07
65	6.64	6.43	96.90
70	5.21	5.96	114.46
75	2.81	5.91	210.43
80	2.05	5.91	288.44
85	7.07	5.60	79.24
% mean recovery			134.12
RMSE			1.91
MAPE			47.16
GRI			1.56

Table D-5: Recovery level of Cd in SVM (testing)

Soils sampling points	Cd		
	Measured	Predicted	% recovery
5	6.13	5.92	96.52
10	5.62	4.95	88.09
20	4.20	2.00	47.57
25	3.90	3.27	83.83
30	3.02	2.33	77.27
35	3.08	3.46	112.30
45	2.95	5.27	178.63
50	1.88	3.50	186.29
55	1.66	2.95	177.44
60	1.20	3.32	276.32
65	2.75	3.42	124.48
70	1.80	3.26	181.30
75	1.83	3.25	177.55
80	1.22	3.25	266.32
85	2.79	2.39	85.75
% mean recovery			143.98
RMSE			1.40
MAPE			60.11
GRI			1.70

Table D-6: Recovery level of As in SVM (testing)

Soils sampling points	As		
	Measured	Predicted	% recovery
5	6.95	7.43	106.98
10	6.45	4.18	64.78
20	3.33	1.25	37.56
25	3.09	3.06	99.00
30	2.88	1.93	67.07
35	2.09	2.27	108.50
45	1.88	4.58	243.68
50	0.88	3.03	344.84
55	1.12	2.03	181.40
60	0.87	2.53	290.39
65	2.32	3.38	145.56
70	0.95	3.15	331.92
75	1.12	3.05	272.67
80	0.99	3.05	308.48
85	2.6	1.98	76.23
% mean recovery			178.60
RMSE			1.64
MAPE			99.32
GRI			2.11

Table D-7: Recovery level of Mn in SVM (testing)

Soils sampling points	Mn		
	Measured	Predicted	% recovery
5	22.84	17.79	77.88
10	14.09	14.27	101.25
20	16.83	4.66	27.71
25	13.54	10.63	78.54
30	11.76	7.07	60.09
35	11.815	11.81	99.99
45	8.91	14.48	162.50
50	3.72	13.84	372.01
55	1.88	10.31	548.25
60	1.11	5.93	533.96
65	11.17	11.97	107.15
70	4.62	11.33	245.16
75	1.99	11.71	588.61
80	1.17	7.71	659.26
85	11.68	6.16	52.75
% mean recovery			247.67
RMSE			7.24
MAPE			174.75
GRI			2.65

Table D-8: Recovery level of Cr in SVM (testing)

Soils sampling points	Cr		
	Measured	Predicted	% recovery
5	9.82	7.42	75.52
10	6.65	6.08	91.37
20	4.77	1.78	37.33
25	5.84	3.76	64.32
30	5.15	2.46	47.80
35	4.22	4.38	103.85
45	2.53	6.15	242.93
50	1.43	4.46	311.93
55	1.1	4.14	376.70
60	0.77	3.99	518.30
65	4.04	4.21	104.13
70	1.44	4.33	300.61
75	1.11	4.63	417.52
80	0.82	4.63	565.18
85	4.39	2.79	63.58
% mean recovery			221.40
RMSE			2.66
MAPE			150.75
GRI			2.52

Table D-9: Recovery level of Ti in SVM (testing)

Soils sampling points	Ti		
	Measured	Predicted	% recovery
5	1765.91	1802.17	102.05
10	1566.17	1371.30	87.56
20	1234.98	690.94	55.95
25	1020.28	1191.20	116.75
30	876.93	683.18	77.91
35	700.83	762.06	108.74
45	898.11	1439.77	160.31
50	702.30	1017.62	144.90
55	516.89	711.30	137.61
60	243.88	865.29	354.80
65	916.17	1204.97	131.52
70	718.54	1210.05	168.40
75	553.55	1143.68	206.61
80	286.55	1143.68	399.12
85	938.02	732.69	78.11
% mean recovery			155.36
RMSE			422.11
MAPE			68.75
GRI			1.75

Table D-10: Recovery level of Sr in SVM (testing)

Soils sampling points	Sr		
	Measured	Predicted	% recovery
5	37.75	39.04	103.42
10	33.66	29.21	86.79
20	26.65	16.04	60.17
25	23.65	26.03	110.07
30	21.61	16.91	78.26
35	18.10	18.80	103.87
45	20.77	31.06	149.53
50	16.90	23.78	140.68
55	12.90	17.27	133.91
60	8.88	20.80	234.26
65	21.71	26.52	122.17
70	16.97	26.39	155.50
75	13.31	25.22	189.48
80	9.10	25.22	277.14
85	22.51	16.96	75.35
% mean recovery			134.71
RMSE			8.25
MAPE			47.96
GRI			1.57

Table D-11: Recovery level of V in SVM (testing)

Soils sampling points	V		
	Measured	Predicted	% recovery
5	65.88	65.08	98.79
10	50.55	46.03	91.06
20	40.88	20.30	49.66
25	38.77	35.41	91.34
30	33.92	22.71	66.95
35	28.13	28.33	100.74
45	26.10	46.84	179.48
50	21.21	35.10	165.51
55	13.98	24.53	175.49
60	6.92	31.32	452.54
65	28.23	37.40	132.48
70	22.15	36.91	166.62
75	14.78	36.41	246.33
80	7.88	36.41	462.03
85	29.97	23.59	78.71
% mean recovery			170.52
RMSE			15.35
MAPE			86.98
GRI			1.91

Table D-12: Recovery level of Ba in SVM (testing)

Soils sampling points	Ba		
	Measured	Predicted	% recovery
5	100.82	102.87	102.03
10	85.76	71.13	82.94
20	60.73	39.17	64.50
25	47.22	61.71	130.69
30	43.88	39.69	90.45
35	40.88	41.81	102.27
45	51.22	75.45	147.31
50	40.82	53.07	130.00
55	25.04	39.83	159.05
60	18.90	46.44	245.71
65	54.01	60.88	112.72
70	41.91	59.96	143.08
75	28.34	55.52	195.91
80	19.48	55.52	285.01
85	55.46	37.51	67.64
% mean recovery			137.29
RMSE			18.90
MAPE			49.88
GRI			1.58

Appendix-E

## Results of Heavy Metal Analysis and Assessment of ANN
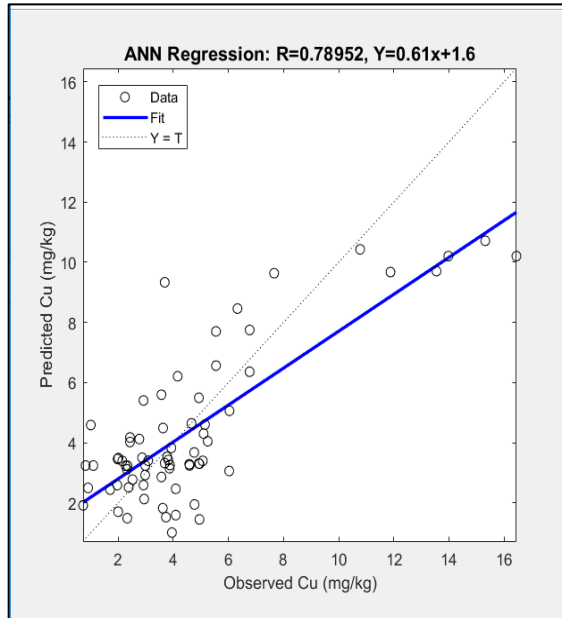
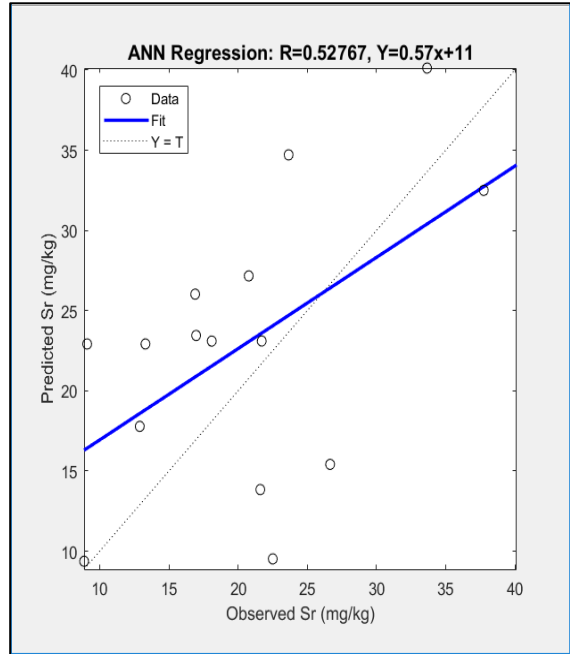Figure E-1: Regression analysis for Cu in ANN (training).
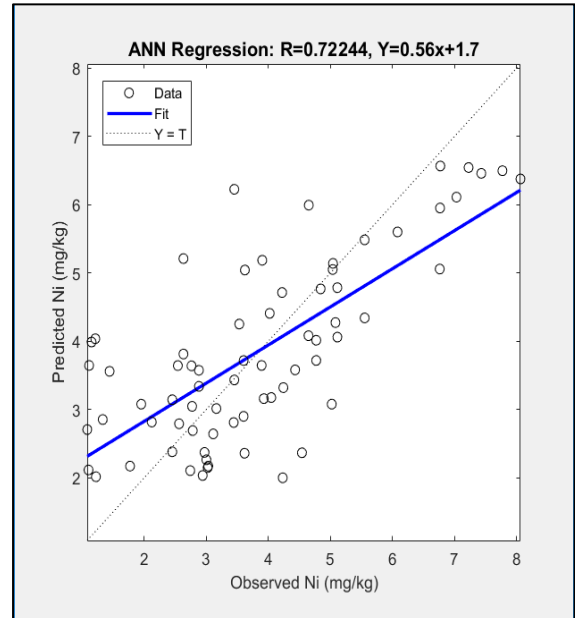
Figure E-2: Regression analysis for Ni in ANN (training).
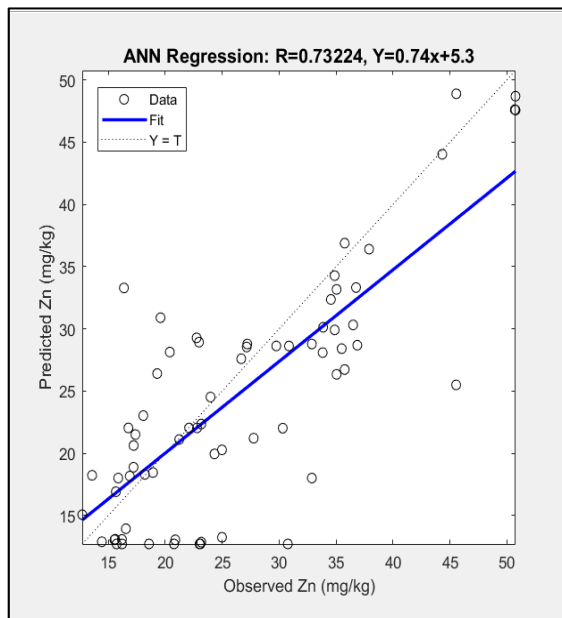
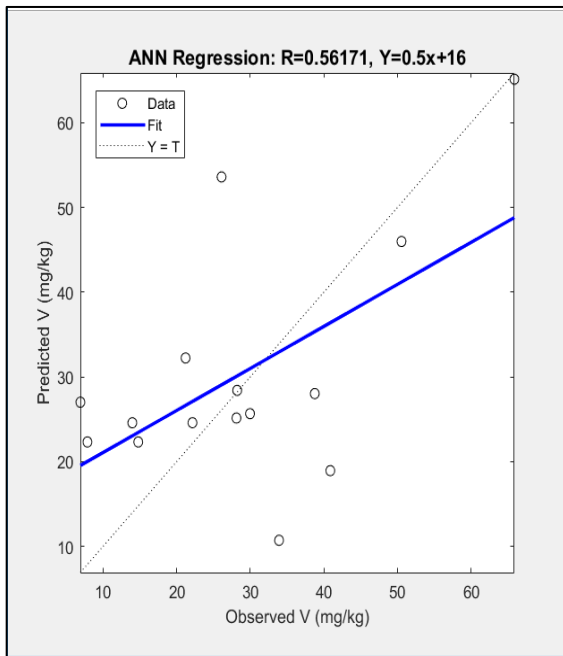Figure E-3: Regression analysis for Zn in ANN (training).
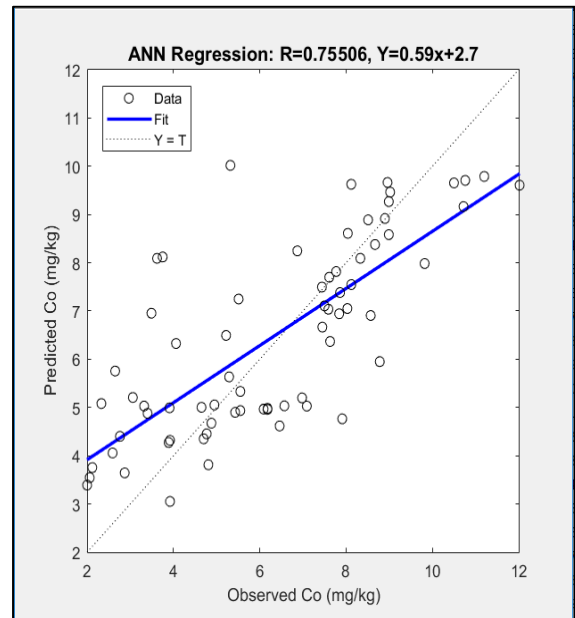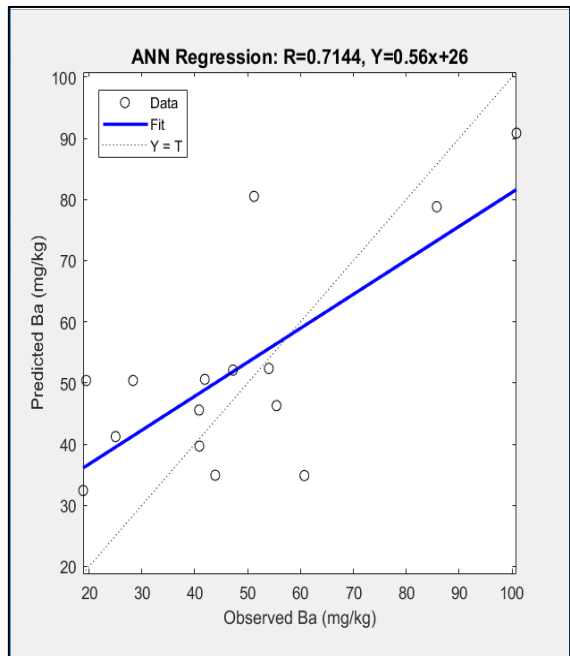

Figure E-4: Regression analysis for Co in ANN (training).
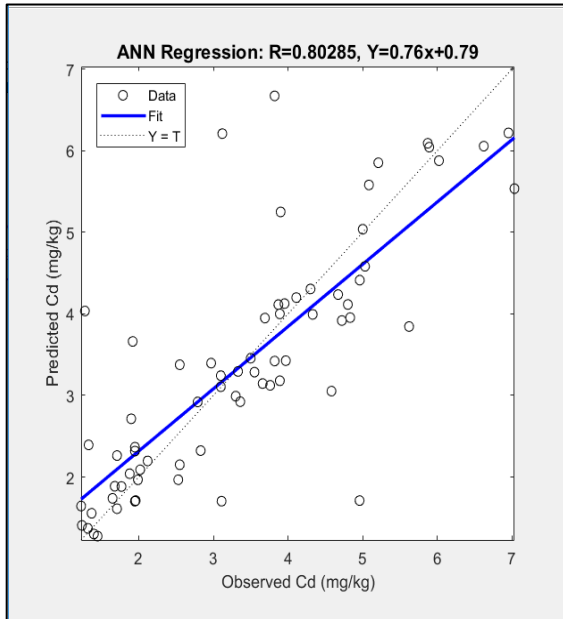
Figure E-5: Regression analysis for Cd in ANN (training).
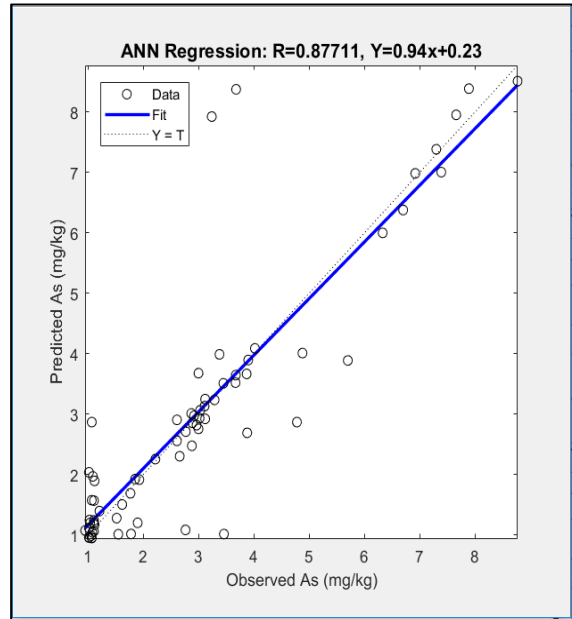
Figure E-6: Regression analysis for As in ANN (training).
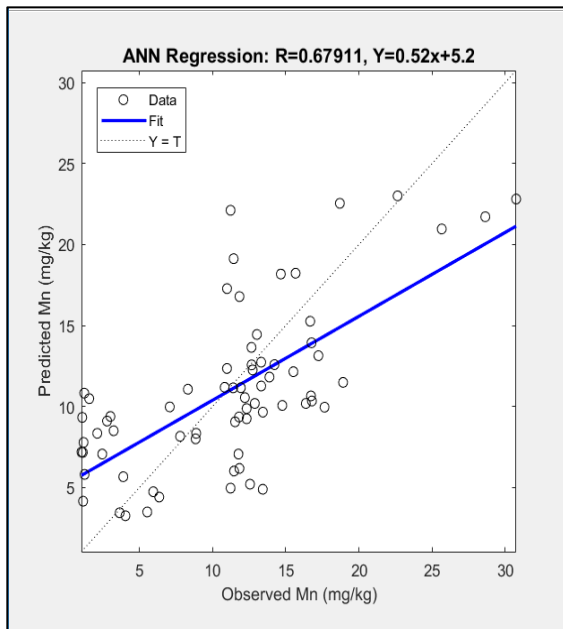
Figure E-7: Regression analysis for Mn in ANN (training).
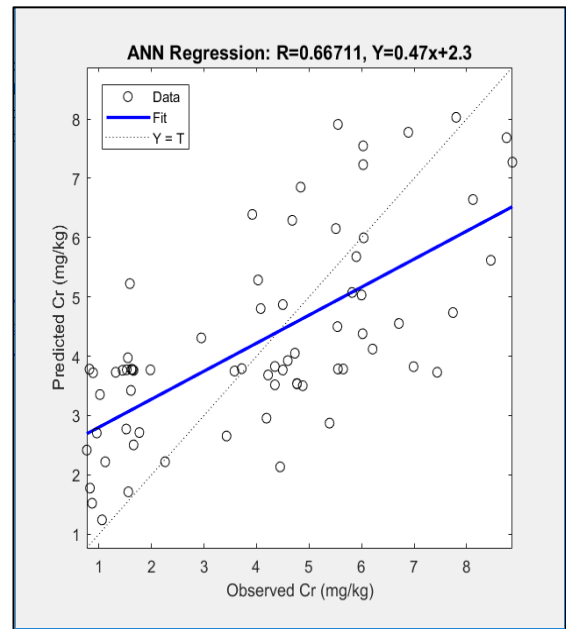
Figure E-8: Regression analysis for Cr in ANN (training).
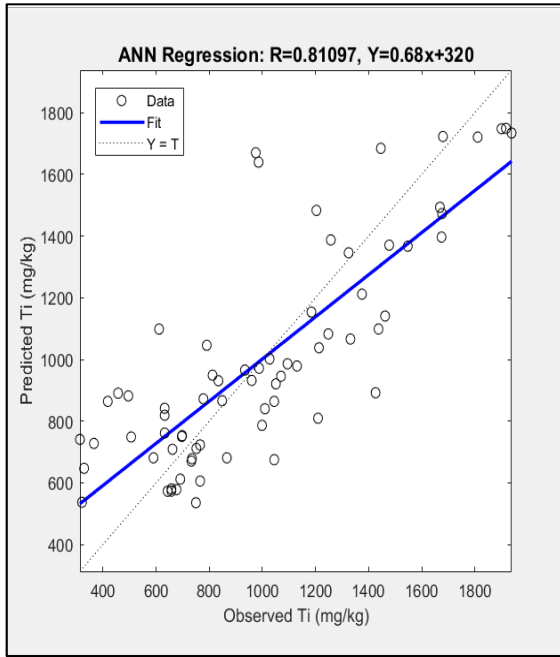
Figure E-9: Regression analysis for Ti in ANN (training).
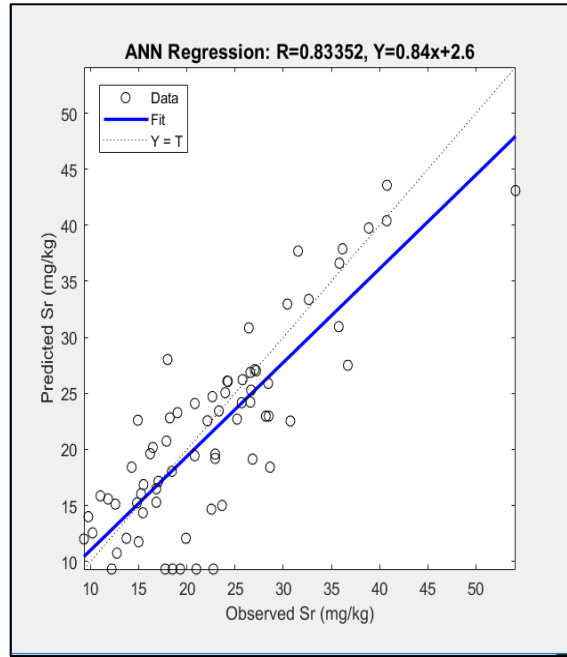
Figure E-10: Regression analysis for Sr in ANN (training).
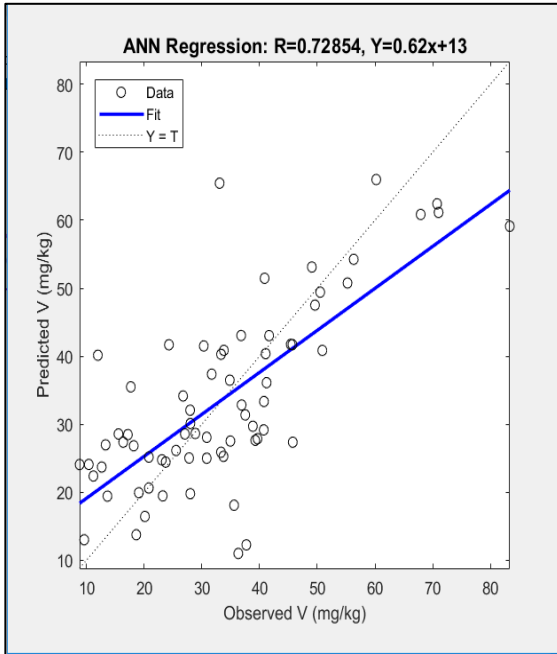
Figure E-11: Regression analysis for V in ANN (training).
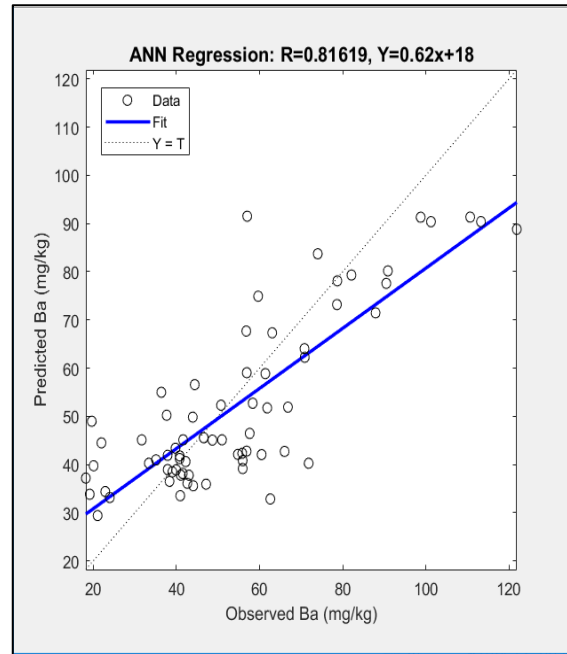
Figure E-12: Regression analysis for Ba in ANN (training).

241
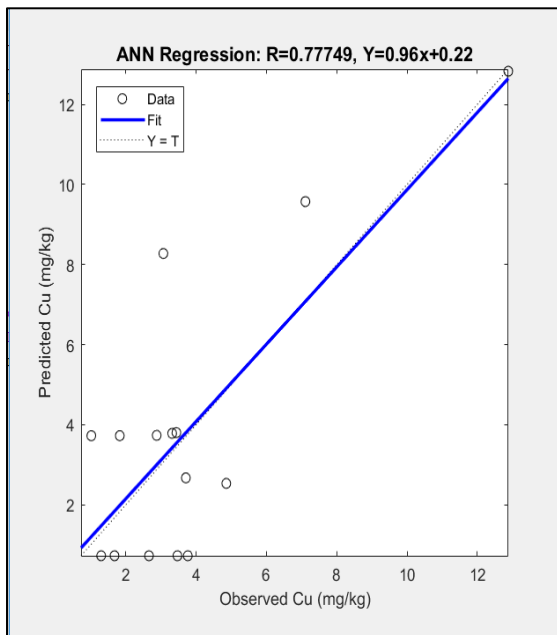
Figure E-13: Regression analysis for Cu in ANN (testing).
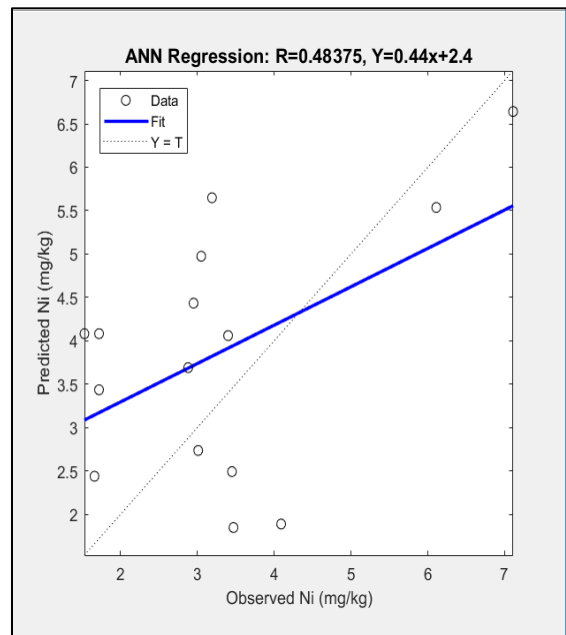
Figure E-14: Regression analysis for Ni in ANN (testing).
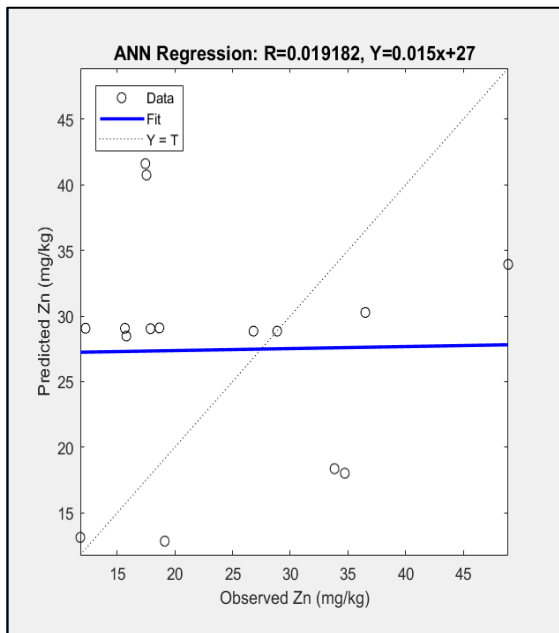
Figure E-15: Regression analysis for Zn in ANN (testing).
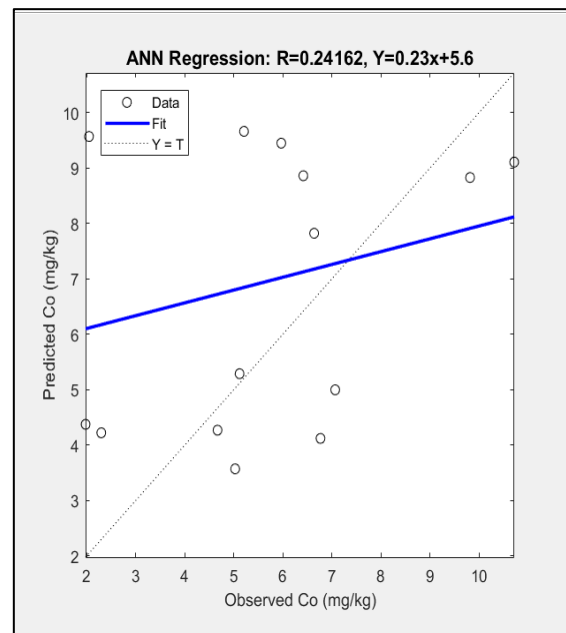
Figure E-16: Regression analysis for Co in ANN (testing).
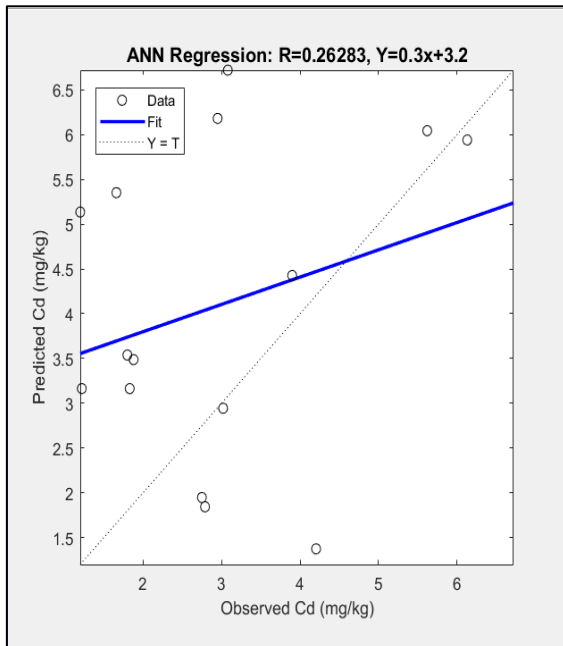
Figure E-17: Regression analysis for Cd
in ANN (testing).
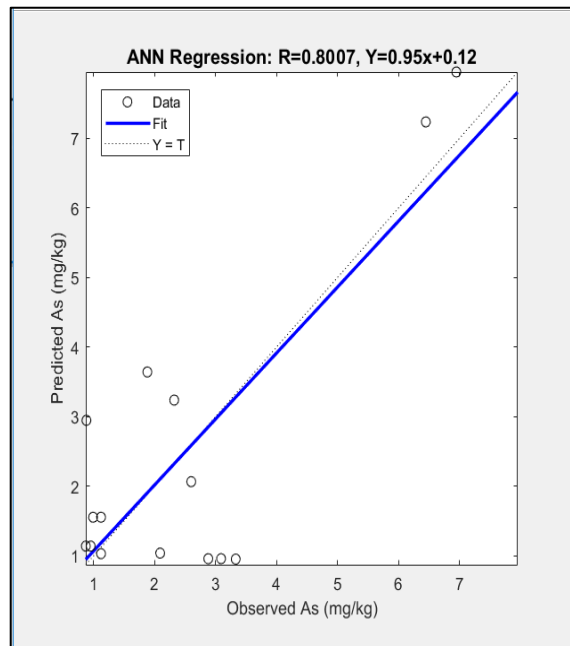
Figure E-18: Regression analysis for As
in ANN (testing).
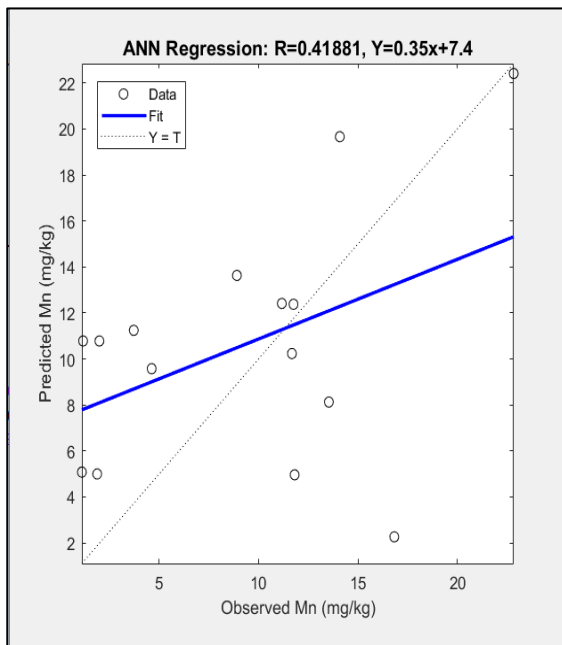
Figure E-19: Regression analysis for Mn
in ANN (testing).
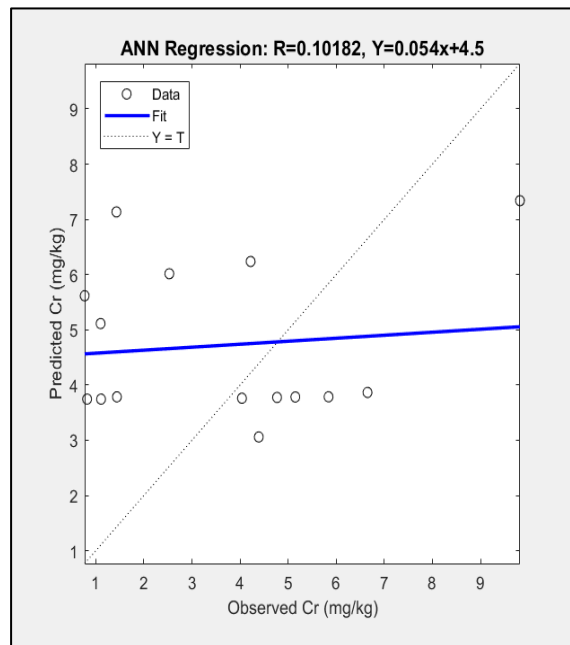
Figure E-20: Regression analysis for Cr
in ANN (testing).
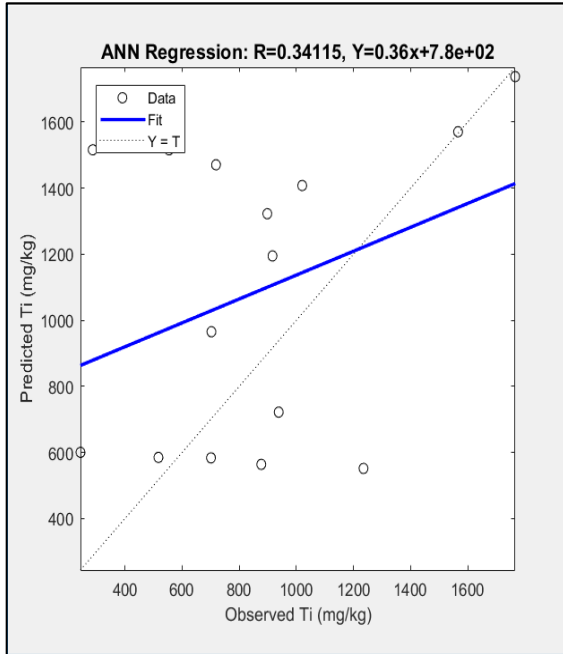
Figure E-21: Regression analysis for Ti in ANN (testing).

Figure E-22: Regression analysis for Sr in ANN (testing).

Figure E-23: Regression analysis for V in ANN (testing).

Figure E-24: Regression analysis for Ba in ANN (testing).
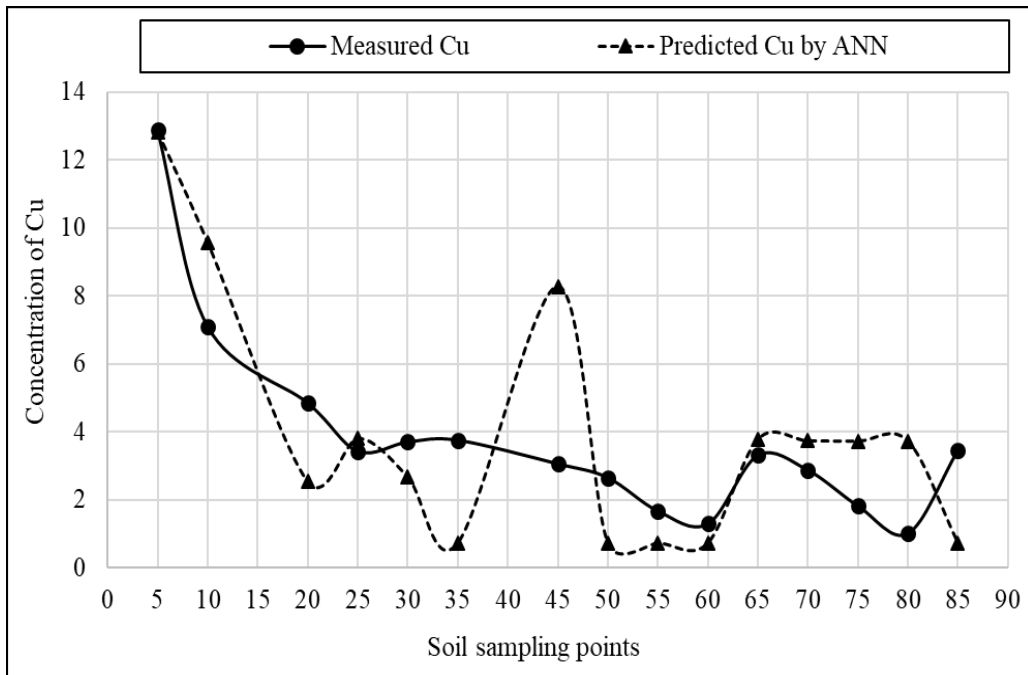
Figure E-25: Comparison of predicted and measured concentration of Cu from ANN (testing).
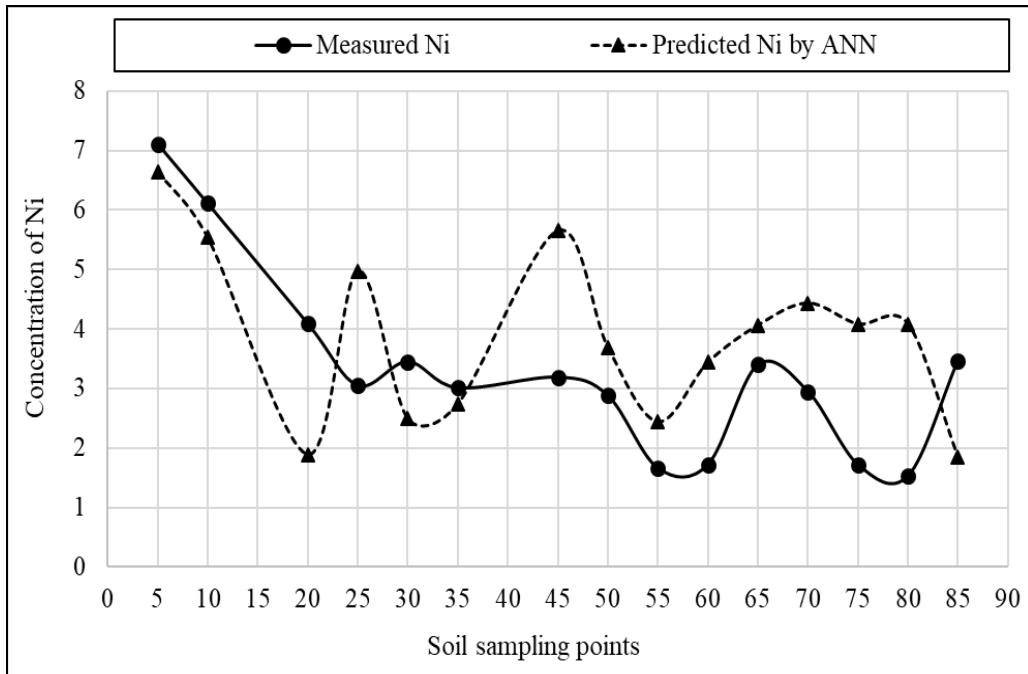
Figure E-26: Comparison of predicted and measured concentration of Ni from ANN (testing).
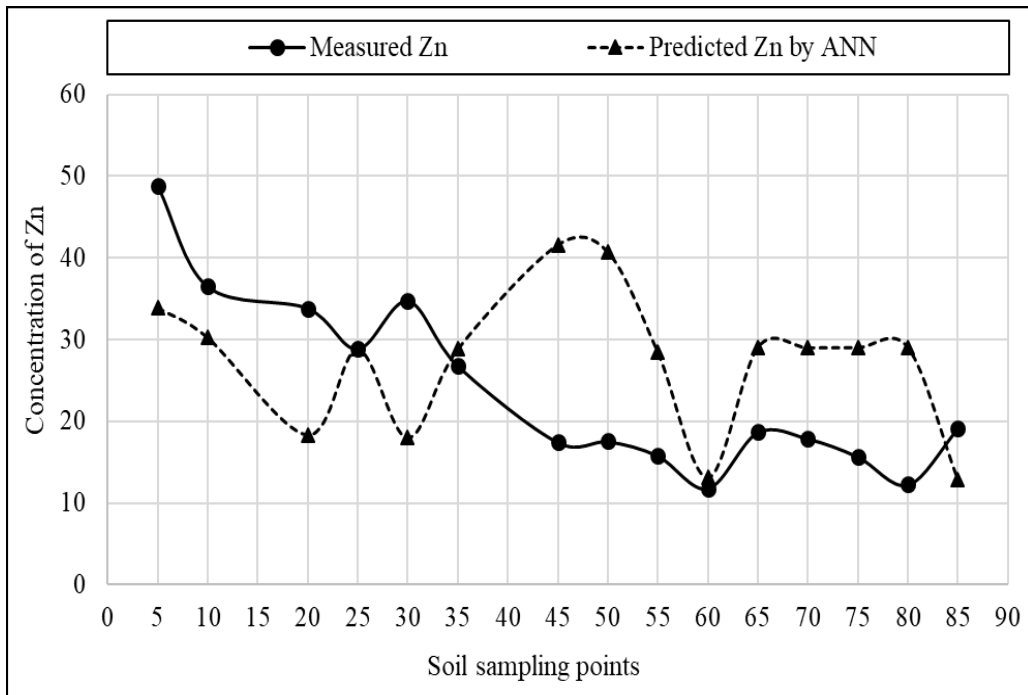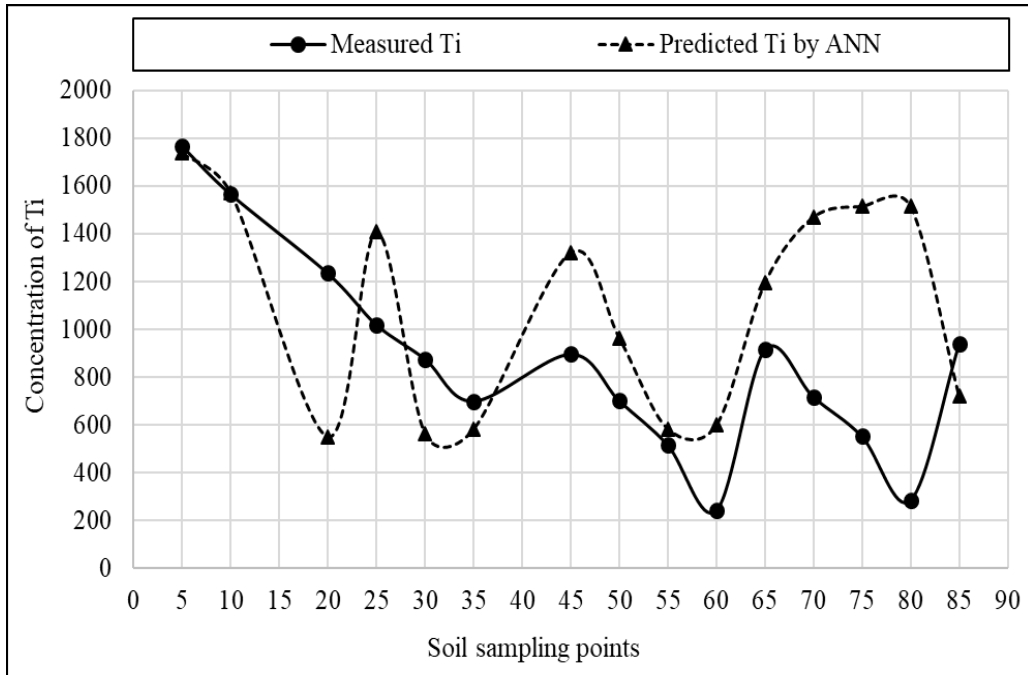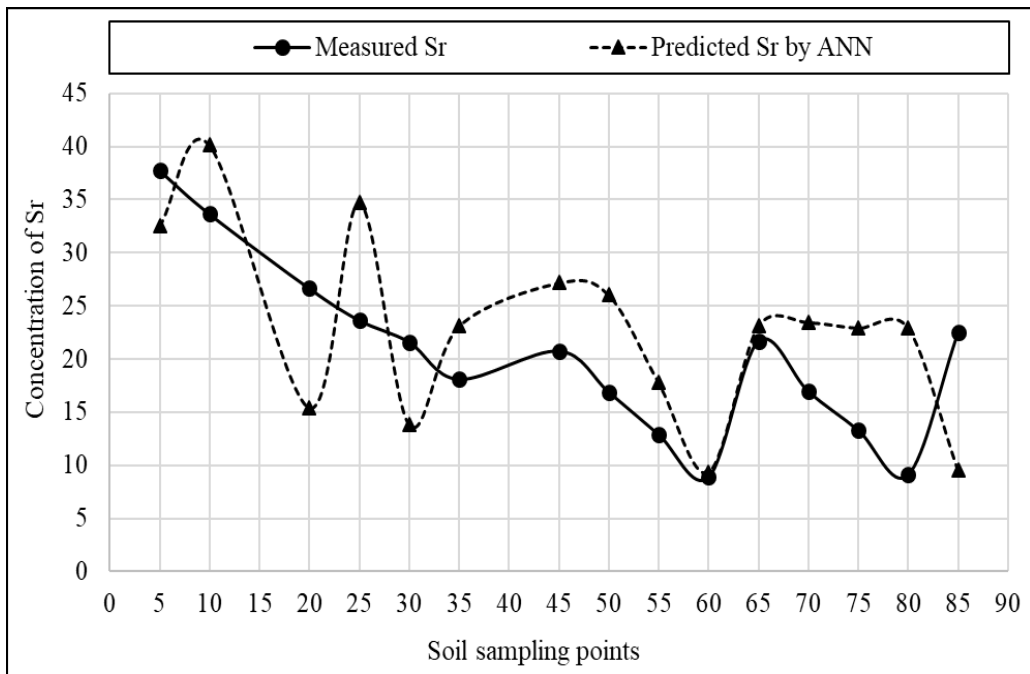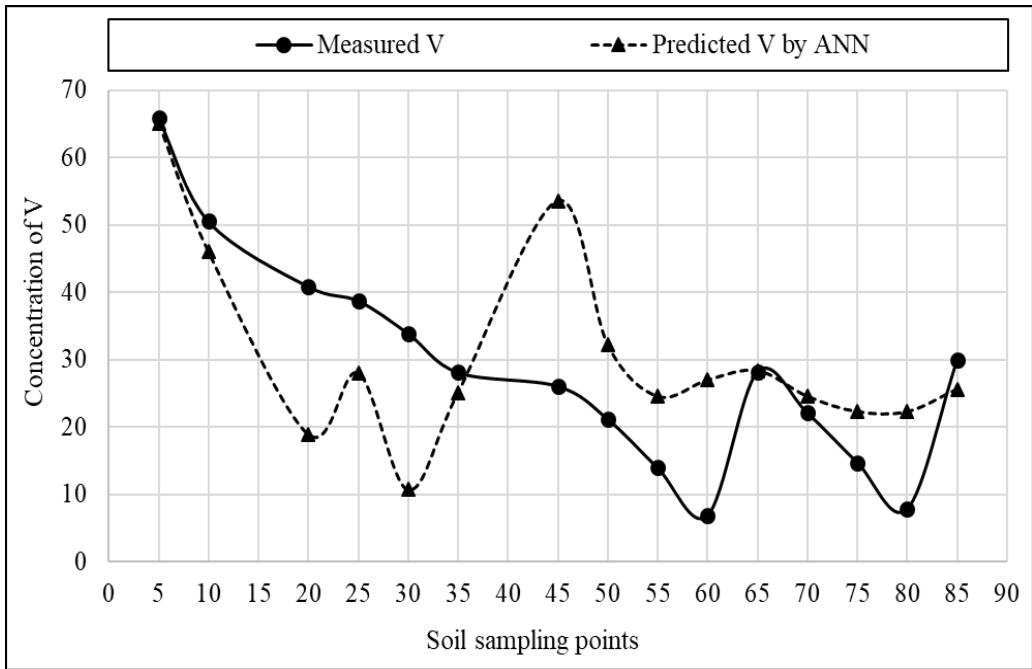
Figure E-27: Comparison of predicted and measured concentration of Zn from ANN (testing).
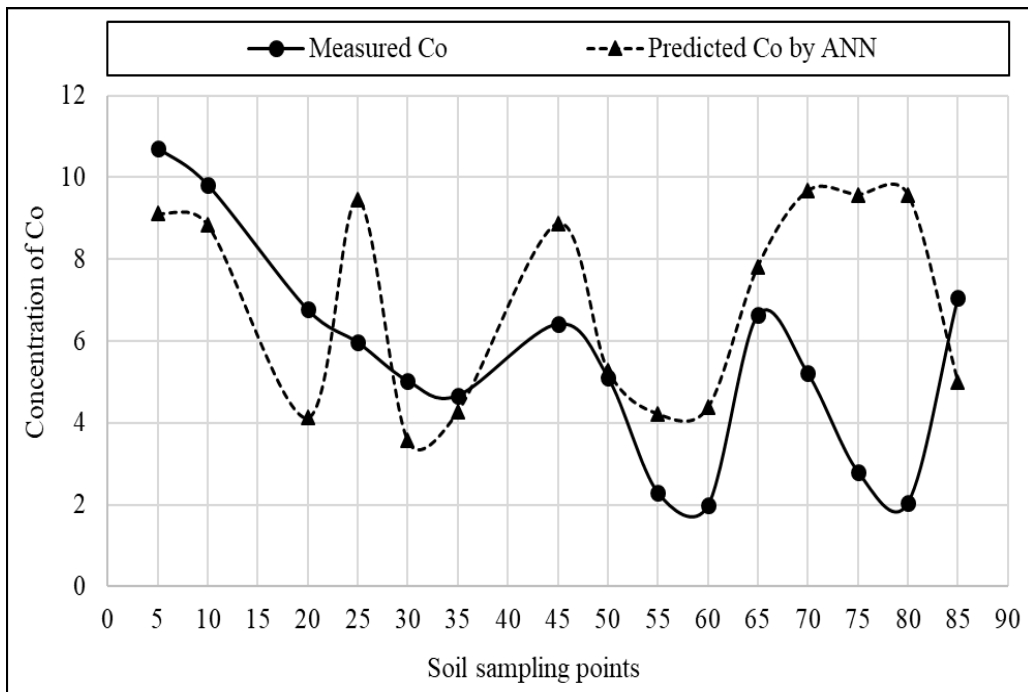
Figure E-28: Comparison of predicted and measured concentration of Co from ANN (testing).
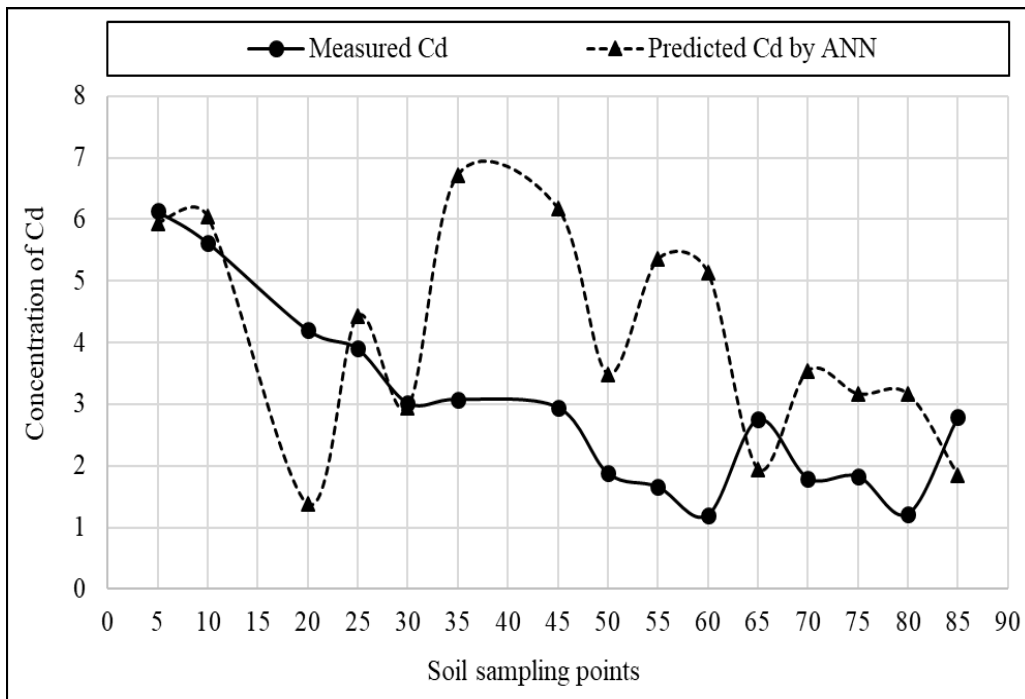
Figure E-29: Comparison of predicted and measured concentration of Cd from ANN (testing).
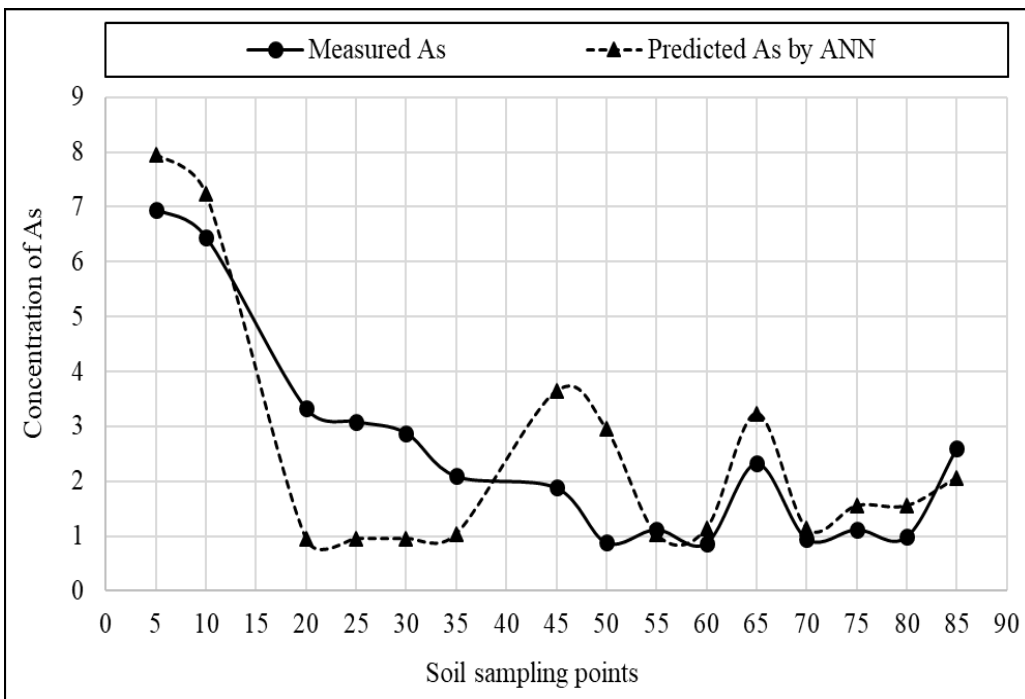
Figure E-30: Comparison of predicted and measured concentration of As from ANN (testing).
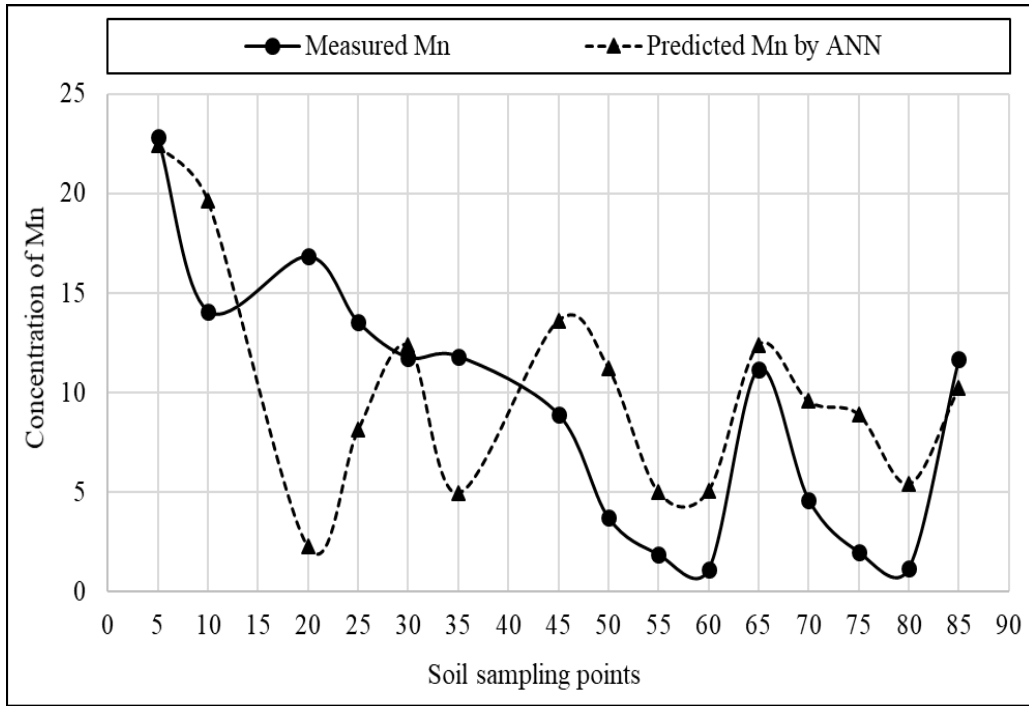
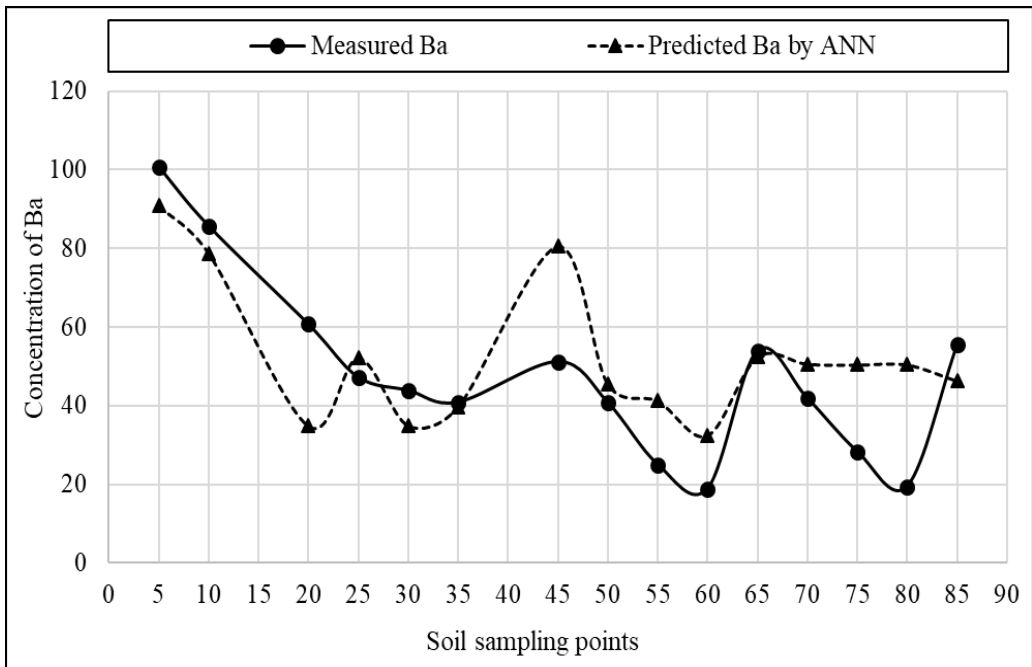Figure E-31: Comparison of predicted and measured concentration of Mn from ANN (testing).
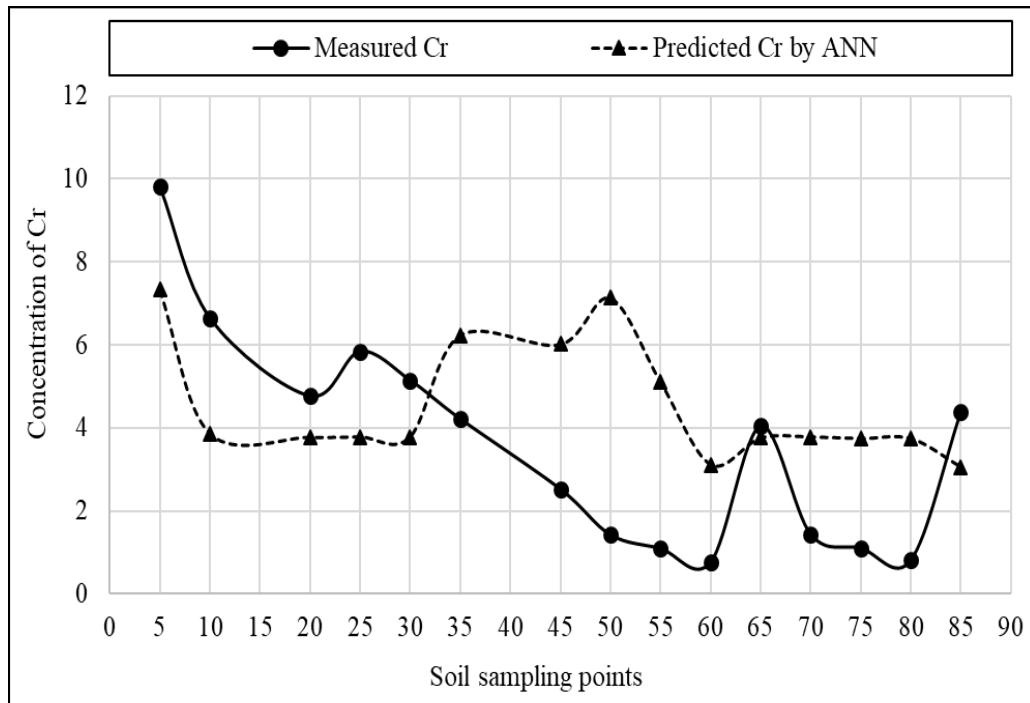
Figure E-32: Comparison of predicted and measured concentration of Cr from ANN (testing).
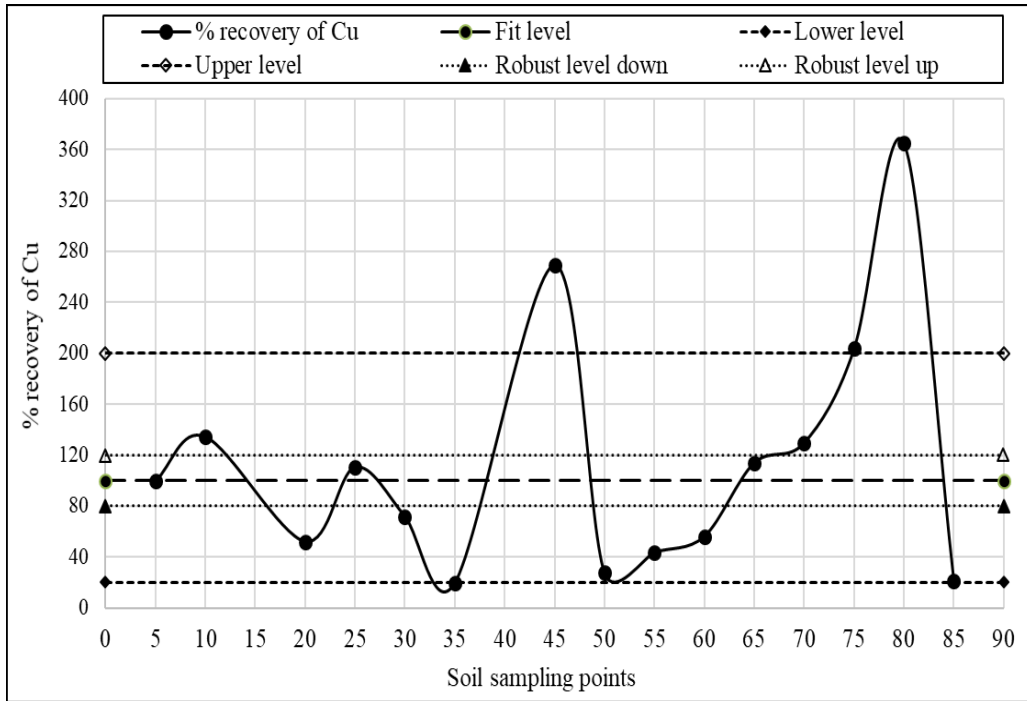
Figure E-33: Comparison of predicted and measured concentration of Ti from ANN (testing).

Figure E-34: Comparison of predicted and measured concentration of Sr from ANN (testing).

Figure E-35: Comparison of predicted and measured concentration of V from ANN (testing).

Figure E-36: Comparison of predicted and measured concentration of Ba from ANN (testing).

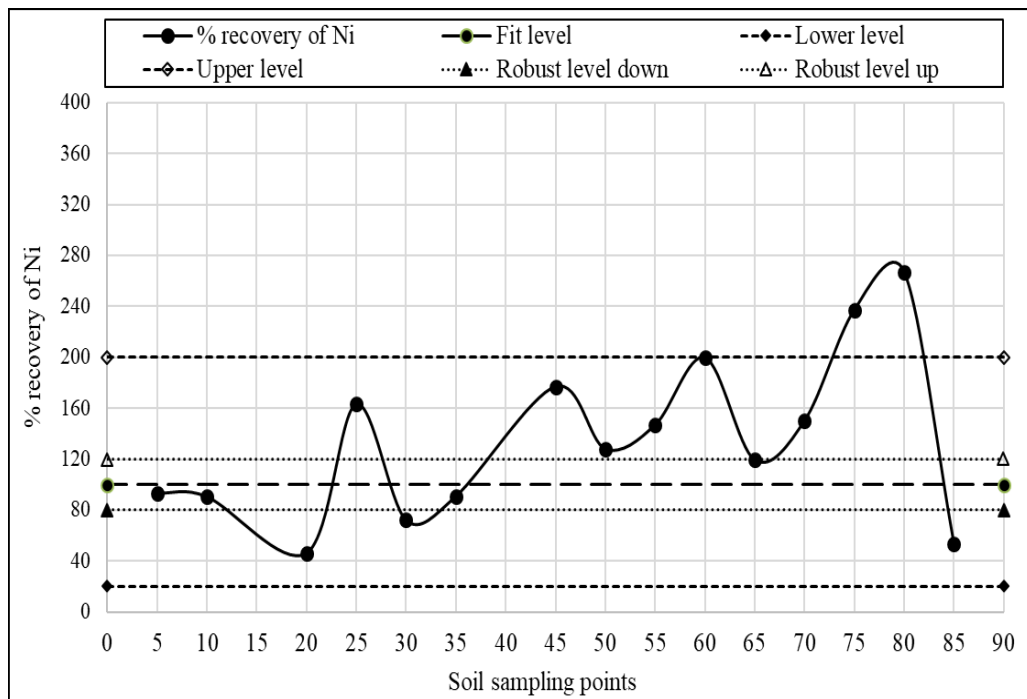Figure E-37: Variation of recovery level for Cu in ANN.

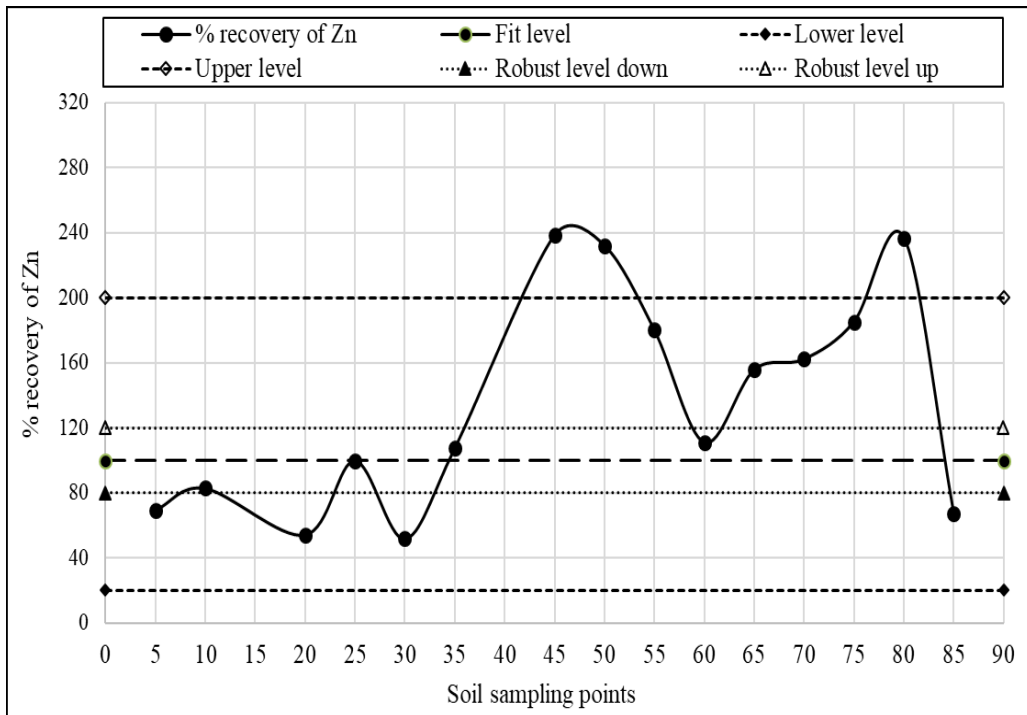Figure E-38: Variation of recovery level for Ni in ANN.

251

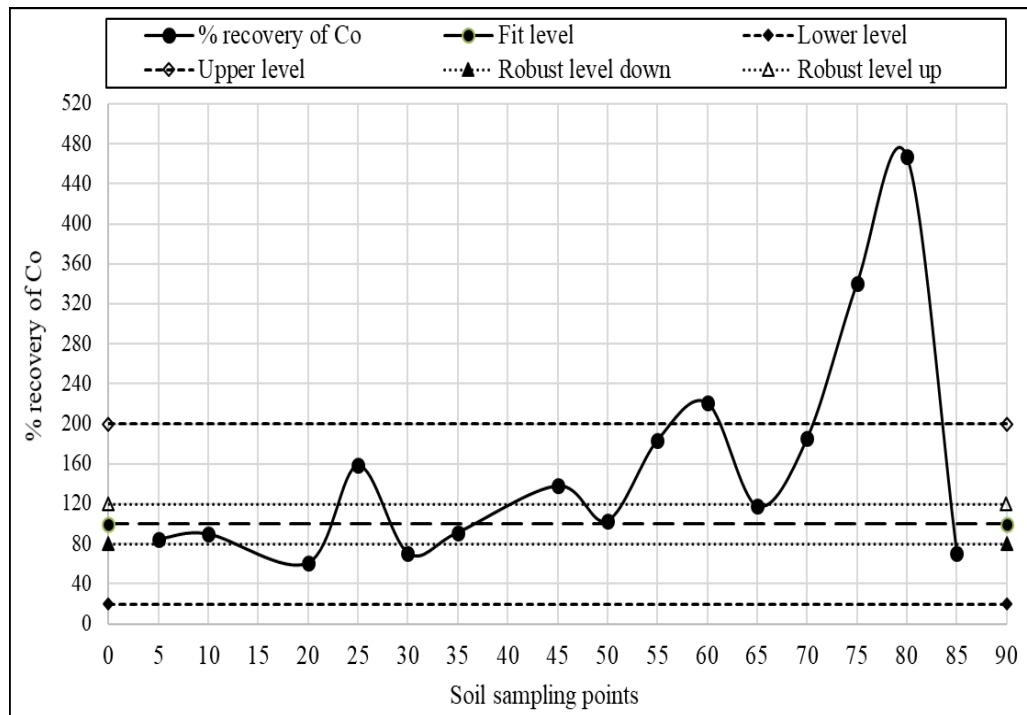Figure E-39: Variation of recovery level for Zn in ANN.

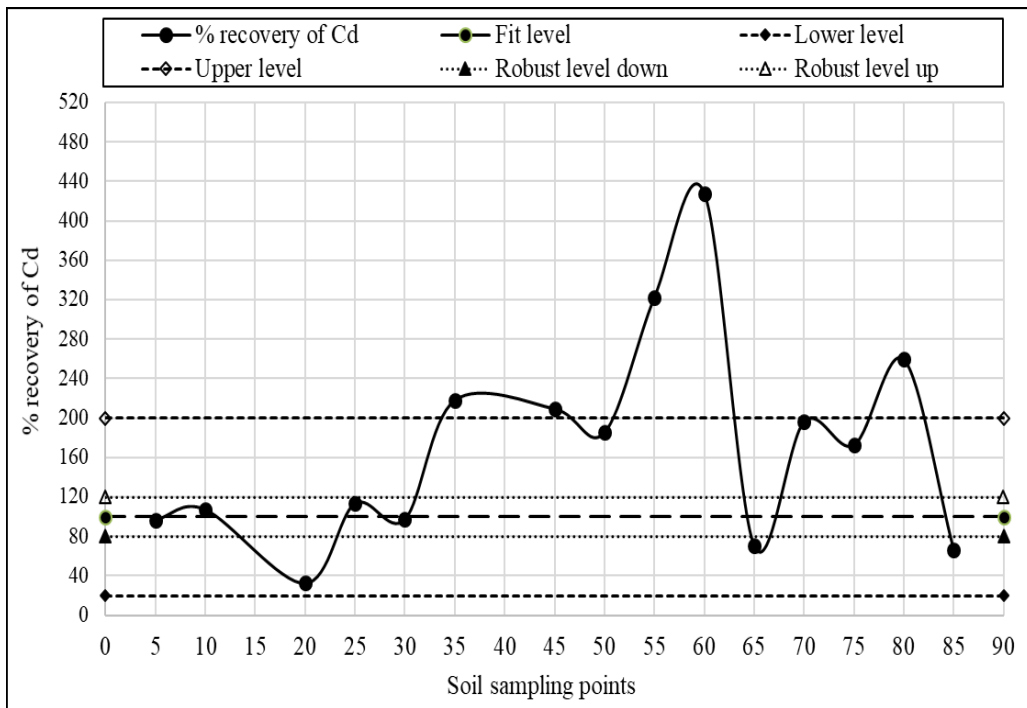Figure E-40: Variation of recovery level for Co in ANN.

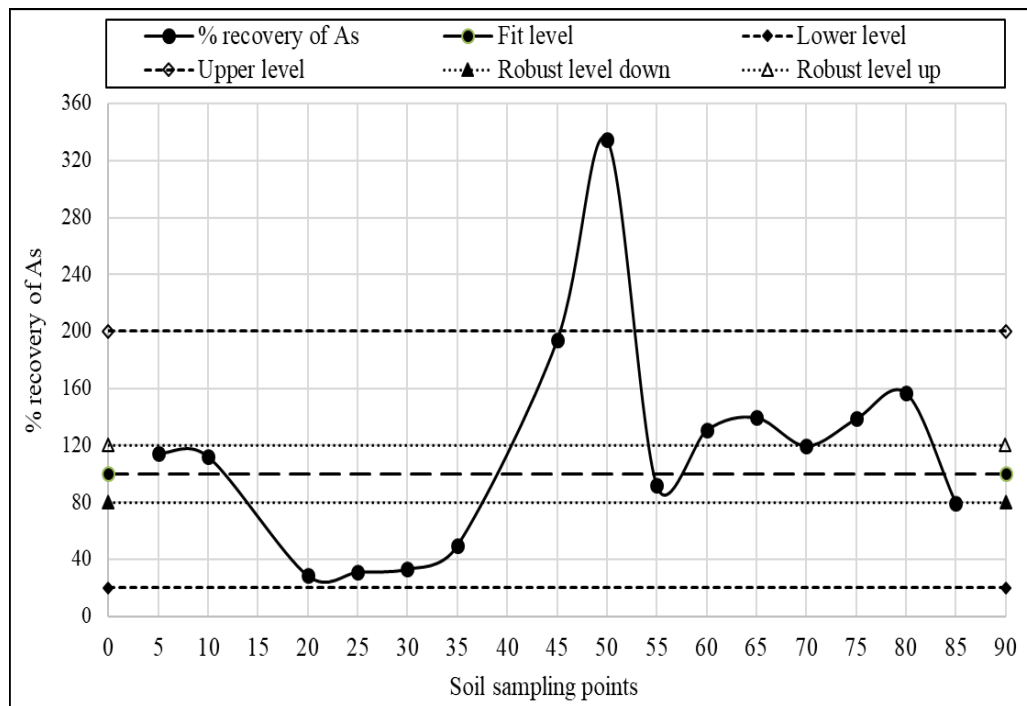Figure E-41: Variation of recovery level for Cd in ANN.

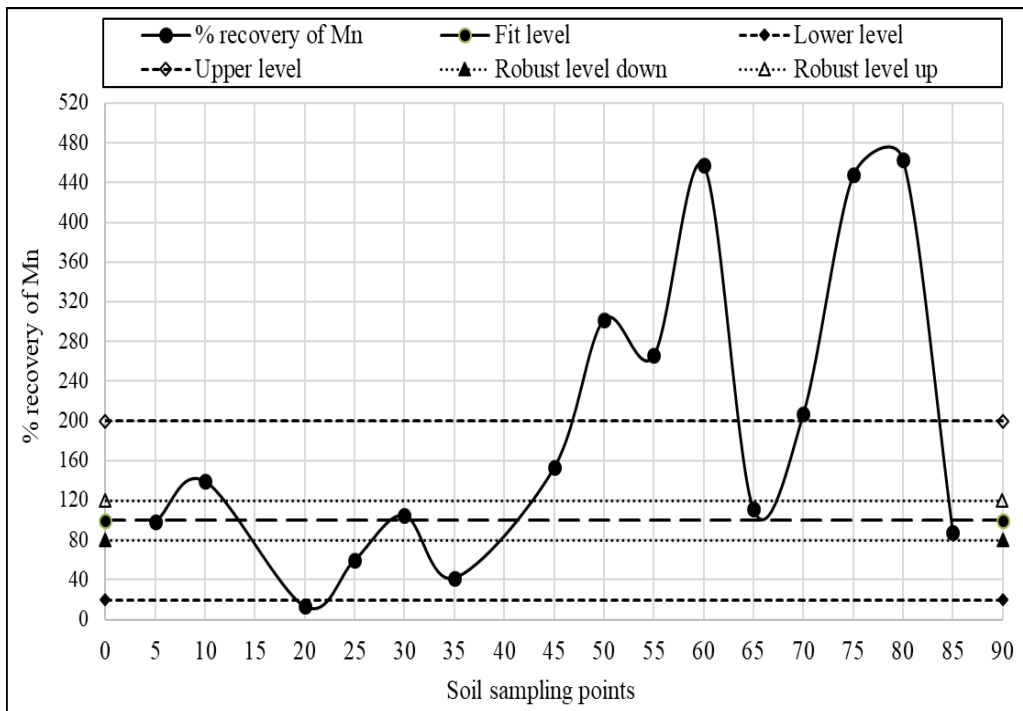Figure E-42: Variation of recovery level for As in ANN.

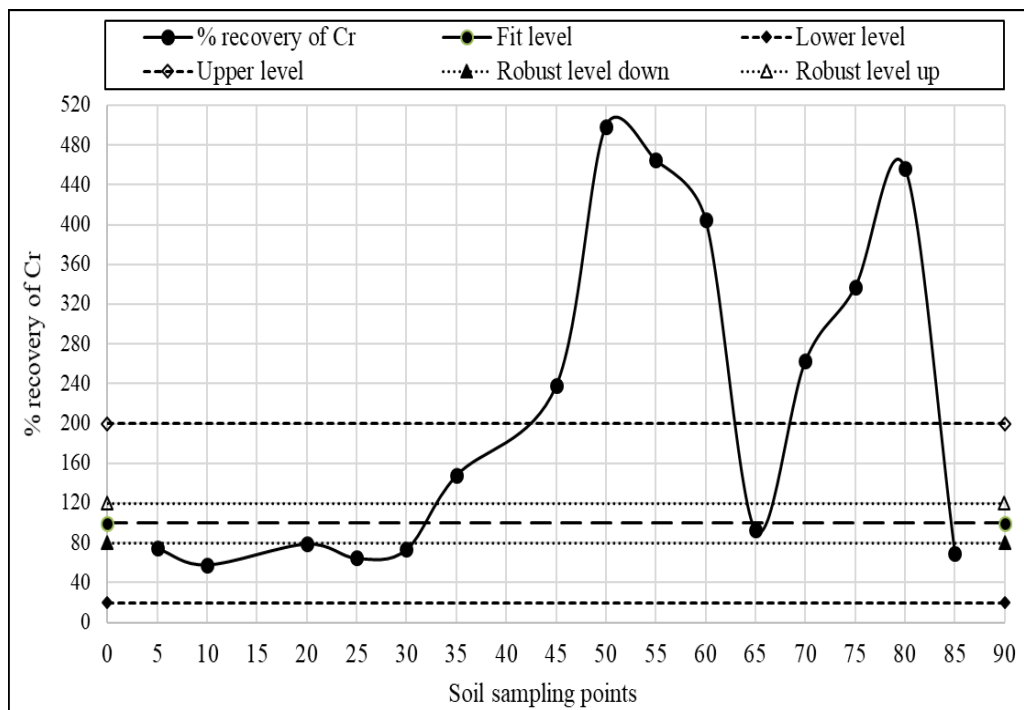Figure E-43: Variation of recovery level for Mn in ANN.

Figure E-44: Variation of recovery level for Cr in ANN.
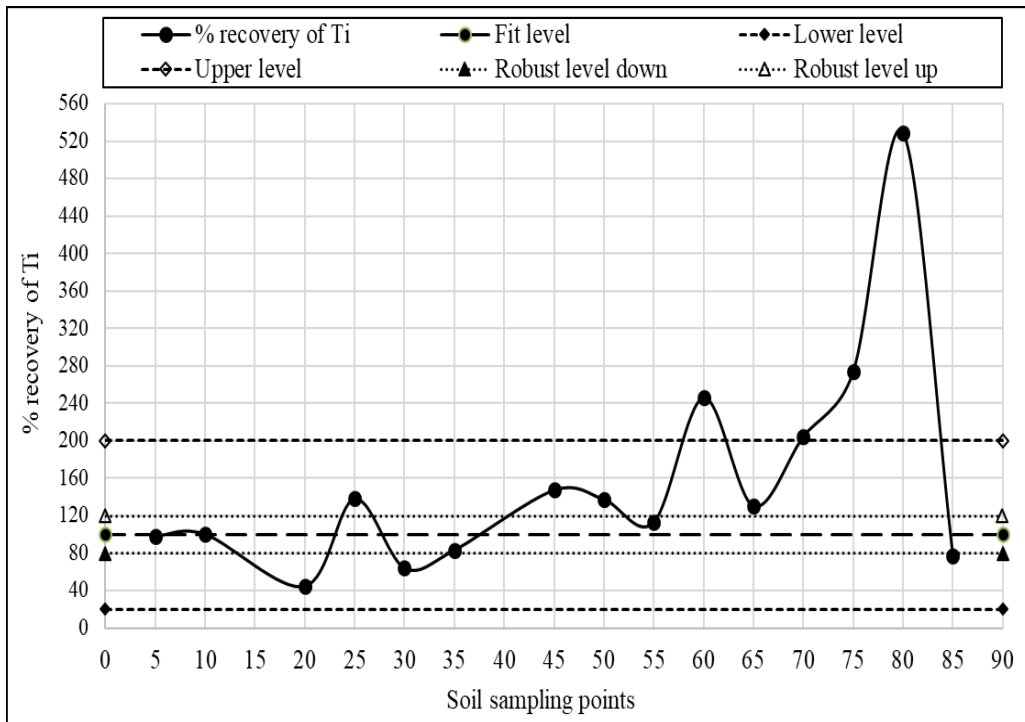
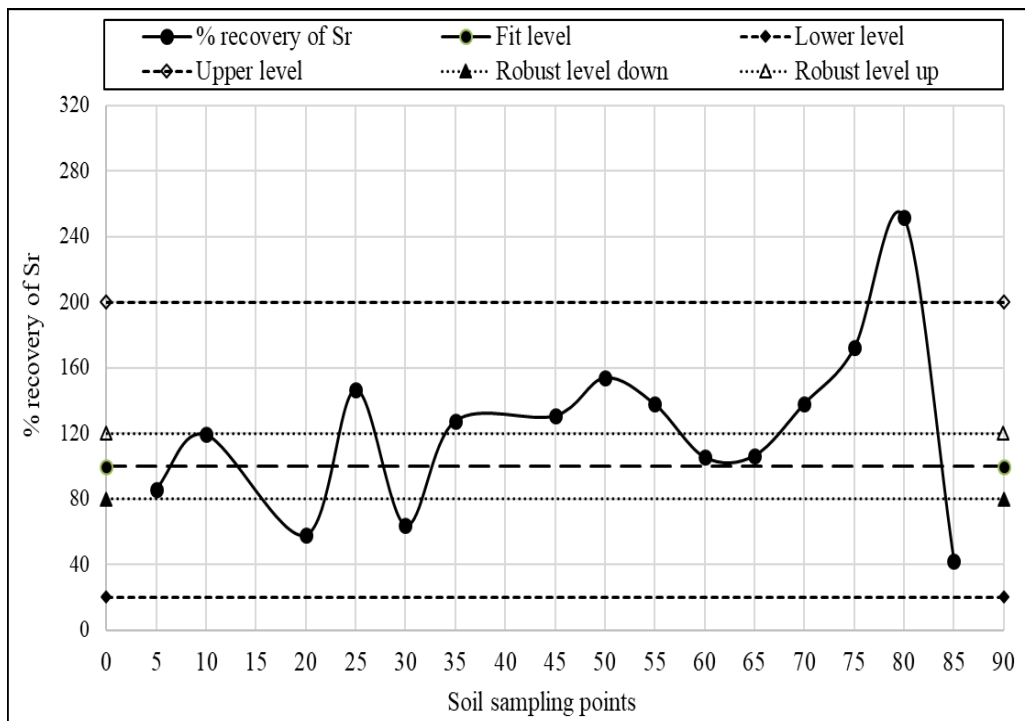Figure E-45: Variation of recovery level for Ti in ANN.

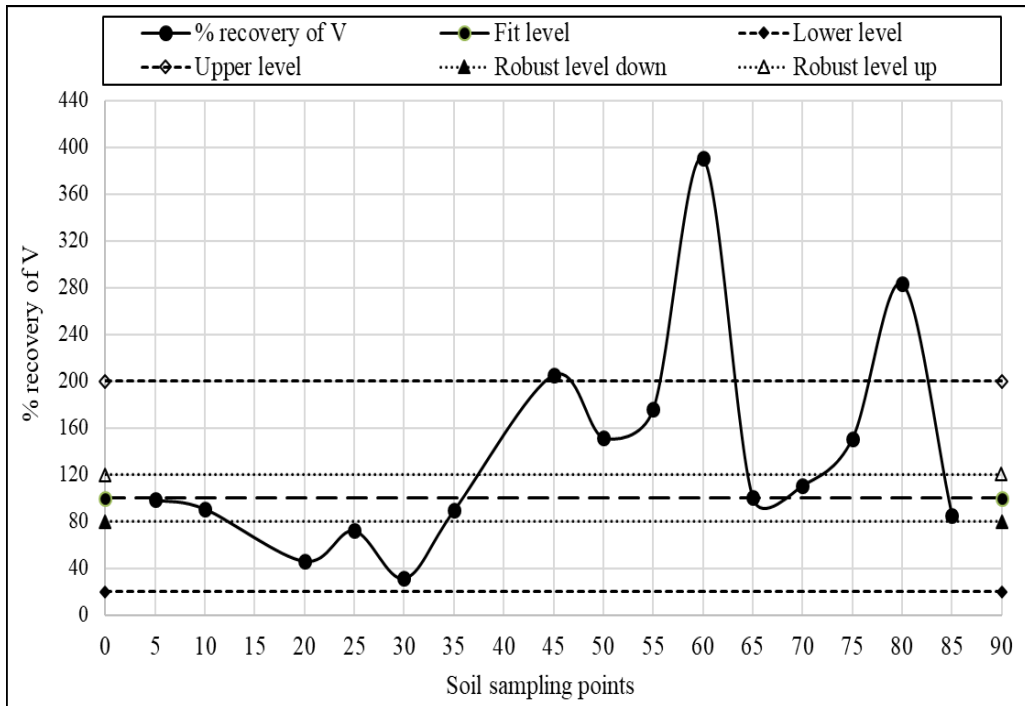Figure E-46: Variation of recovery level for Sr in ANN.

255

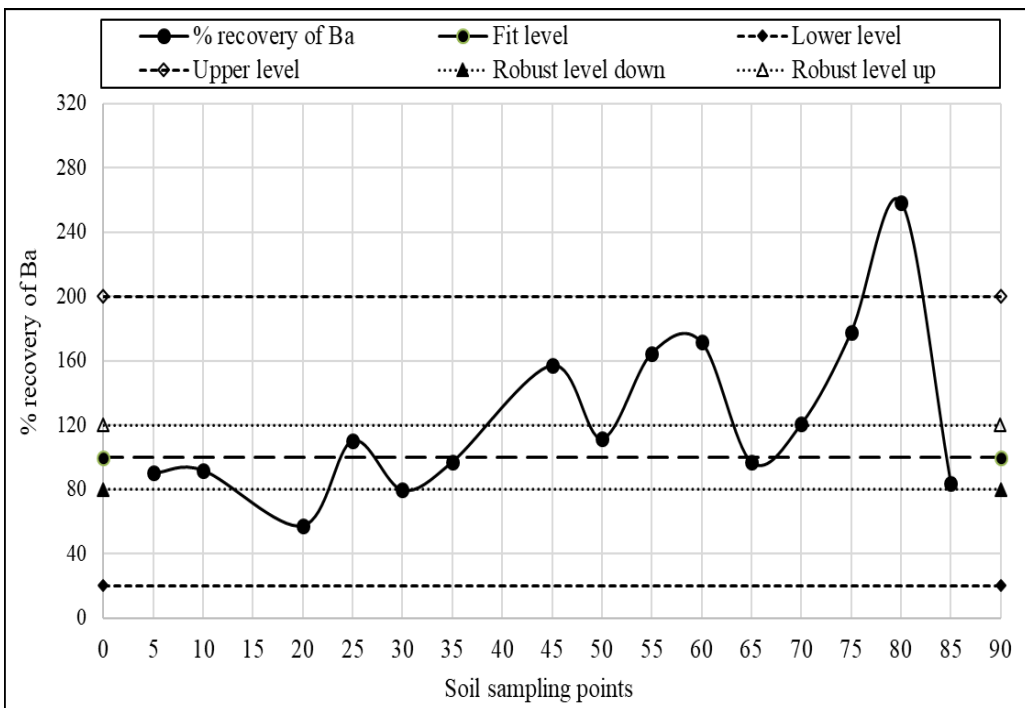Figure E-47: Variation of recovery level for V in ANN.

Figure E-48: Variation of recovery level for Ba in ANN.

256

Appendix-F

Recovery Level for Various Heavy Metals of ANN

Table F-1: Recovery level of Cu in ANN (testing)

Soils sampling points	Cu		
	Measured	Predicted	% recovery
5	12.88	12.83	99.60
10	7.11	9.57	134.65
20	4.86	2.54	52.18
25	3.44	3.80	110.58
30	3.71	2.67	72.07
35	3.77	0.73	19.39
45	3.07	8.28	269.64
50	2.66	0.73	27.44
55	1.68	0.73	43.45
60	1.30	0.73	56.15
65	3.32	3.79	114.01
70	2.88	3.74	129.73
75	1.83	3.73	203.68
80	1.02	3.73	365.43
85	3.47	0.73	21.04
% mean recovery			114.60
RMSE			2.21
MAPE			69.09
GRI			2.27

Table F-2: Recovery level of Ni in ANN (testing)

Soils sampling points	Ni		
	Measured	Predicted	% recovery
5	7.11	6.64	93.42
10	6.11	5.54	90.61
20	4.09	1.89	46.18
25	3.05	4.97	163.09
30	3.45	2.49	72.23
35	3.01	2.74	90.89
45	3.19	5.65	177.05
50	2.88	3.69	128.14
55	1.66	2.44	146.90
60	1.72	3.44	199.74
65	3.4	4.06	119.40
70	2.95	4.43	150.27
75	1.72	4.08	237.31
80	1.53	4.08	266.78
85	3.47	1.85	53.25
% mean recovery			135.68
RMSE			1.59
MAPE			56.14
GRI			1.69

	Table F-3: Recovery level of in ANN (testing)					Table F-4: Recovery level of Co in ANN (testing)		

Table F-3: Recovery level of in ANN (testing)

Soils sampling points	Zn		
	Measured	Predicted	% recovery
5	48.87	33.93	69.44
10	36.50	30.27	82.93
20	33.83	18.37	54.30
25	28.87	28.84	99.91
30	34.72	18.03	51.92
35	26.82	28.84	107.56
45	17.45	41.60	238.39
50	17.55	40.73	232.10
55	15.80	28.47	180.18
60	11.82	13.14	111.20
65	18.67	29.09	155.81
70	17.89	29.03	162.28
75	15.68	29.06	185.34
80	12.27	29.06	236.85
85	19.13	12.86	67.20
% mean recovery			135.69
RMSE			13.66
MAPE			58.93
GRI			1.71

Table F-4: Recovery level of Co in ANN (testing)

Soils sampling points	Co		
	Measured	Predicted	% recovery
5	10.72	9.11	84.98
10	9.82	8.83	89.93
20	6.77	4.12	60.90
25	5.97	9.45	158.33
30	5.03	3.57	71.02
35	4.67	4.27	91.47
45	6.42	8.86	138.08
50	5.12	5.29	103.32
55	2.3	4.22	183.65
60	1.98	4.38	221.03
65	6.64	7.83	117.85
70	5.21	9.66	185.51
75	2.81	9.57	340.67
80	2.05	9.57	466.96
85	7.07	5.00	70.72
% mean recovery			158.96
RMSE			3.35
MAPE			76.42
GRI			1.81

Table F-5: Recovery level of Cd in ANN (testing)

Soils sampling points	Cd		
	Measured	Predicted	% recovery
5	6.13	5.94	96.91
10	5.62	6.04	107.54
20	4.20	1.37	32.68
25	3.90	4.43	113.53
30	3.02	2.94	97.47
35	3.08	6.72	218.17
45	2.95	6.18	209.52
50	1.88	3.49	185.50
55	1.66	5.35	322.39
60	1.20	5.14	427.98
65	2.75	1.95	70.82
70	1.80	3.54	196.60
75	1.83	3.16	172.88
80	1.22	3.16	259.32
85	2.79	1.84	66.08
% mean recovery			171.83
RMSE			2.22
MAPE			89.96
GRI			2.03

Table F-6: Recovery level of As in ANN (testing)

Soils sampling points	As		
	Measured	Predicted	% recovery
5	6.95	7.96	114.49
10	6.45	7.24	112.25
20	3.33	0.95	28.53
25	3.09	0.96	30.97
30	2.88	0.96	33.20
35	2.09	1.04	49.56
45	1.88	3.64	193.71
50	0.88	2.95	334.79
55	1.12	1.03	91.81
60	0.87	1.14	130.75
65	2.32	3.24	139.53
70	0.95	1.14	119.72
75	1.12	1.55	138.63
80	0.99	1.55	156.84
85	2.6	2.06	79.39
% mean recovery			116.94
RMSE			1.31
MAPE			55.15
GRI			1.92

Table F-7: Recovery level of Mn in ANN (testing)

Soils sampling points	Mn		
	Measured	Predicted	% recovery
5	22.84	22.41	98.13
10	14.09	19.66	139.55
20	16.83	2.27	13.48
25	13.54	8.13	60.03
30	11.76	12.38	105.26
35	11.815	4.96	42.01
45	8.91	13.63	153.00
50	3.72	11.25	302.29
55	1.88	5.00	265.99
60	1.11	5.08	457.72
65	11.17	12.41	111.13
70	4.62	9.58	207.30
75	1.99	8.91	447.74
80	1.17	5.42	463.25
85	11.68	10.24	87.63
% mean recovery			196.97
RMSE			4.79
MAPE			123.46
GRI			2.47

Table F-8: Recovery level of Cr in ANN (testing)

Soils sampling points	Cr		
	Measured	Predicted	% recovery
5	9.82	7.34	74.73
10	6.65	3.87	58.13
20	4.77	3.78	79.16
25	5.84	3.79	64.85
30	5.15	3.78	73.48
35	4.22	6.24	147.82
45	2.53	6.02	237.80
50	1.43	7.14	498.97
55	1.1	5.11	464.91
60	0.77	3.12	404.92
65	4.04	3.76	93.10
70	1.44	3.79	262.92
75	1.11	3.75	337.44
80	0.82	3.75	456.77
85	4.39	3.06	69.71
% mean recovery			221.65
RMSE			2.97
MAPE			146.56
GRI			2.40

Table F-9: Recovery level of Ti in ANN (testing)

Soils sampling points	Ti		
	Measured	Predicted	% recovery
5	1765.91	1737.60	98.40
10	1566.17	1570.64	100.29
20	1234.98	551.40	44.65
25	1020.28	1407.71	137.97
30	876.93	563.68	64.28
35	700.83	583.35	83.24
45	898.11	1322.45	147.25
50	702.30	965.14	137.43
55	516.89	584.66	113.11
60	243.88	600.12	246.07
65	916.17	1194.61	130.39
70	718.54	1470.54	204.66
75	553.55	1515.86	273.84
80	286.55	1515.86	529.00
85	938.02	721.83	76.95
% mean recovery			159.17
RMSE			531.52
MAPE			76.83
GRI			1.84

Table F-10: Recovery level of Sr in ANN (testing)

Soils sampling points	Sr		
	Measured	Predicted	% recovery
5	37.75	32.50	86.09
10	33.66	40.12	119.21
20	26.65	15.42	57.86
25	23.65	34.71	146.76
30	21.61	13.85	64.08
35	18.10	23.10	127.64
45	20.77	27.16	130.78
50	16.90	26.03	154.01
55	12.90	17.79	137.87
60	8.88	9.38	105.58
65	21.71	23.11	106.43
70	16.97	23.45	138.19
75	13.31	22.92	172.22
80	9.10	22.92	251.89
85	22.51	9.53	42.32
% mean recovery			122.73
RMSE			8.36
MAPE			42.68
GRI			1.56

Table F-11: Recovery level of V in ANN (testing)

Soils sampling points	V		
	Measured	Predicted	% recovery
5	65.88	65.14	98.88
10	50.55	45.99	90.98
20	40.88	18.94	46.32
25	38.77	28.03	72.30
30	33.92	10.72	31.61
35	28.13	25.15	89.41
45	26.10	53.61	205.39
50	21.21	32.24	151.99
55	13.98	24.59	175.88
60	6.92	27.02	390.46
65	28.23	28.41	100.64
70	22.15	24.60	111.04
75	14.78	22.33	151.06
80	7.88	22.33	283.33
85	29.97	25.67	85.64
% mean recovery			139.00
RMSE			13.78
MAPE			63.64
GRI			1.84

Table F-12: Recovery level of Ba in ANN (testing)

Soils sampling points	Ba		
	Measured	Predicted	% recovery
5	100.82	90.85	90.11
10	85.76	78.81	91.89
20	60.73	34.87	57.41
25	47.22	52.12	110.37
30	43.88	34.92	79.59
35	40.88	39.69	97.08
45	51.22	80.51	157.19
50	40.82	45.58	111.66
55	25.04	41.25	164.72
60	18.90	32.43	171.56
65	54.01	52.38	96.97
70	41.91	50.60	120.74
75	28.34	50.41	177.88
80	19.48	50.41	258.78
85	55.46	46.31	83.49
% mean recovery			124.63
RMSE			16.03
MAPE			38.42
GRI			1.48

Appendix-G

Variation of Predicted Results of Heavy Metals with Different AI Techniques for Training
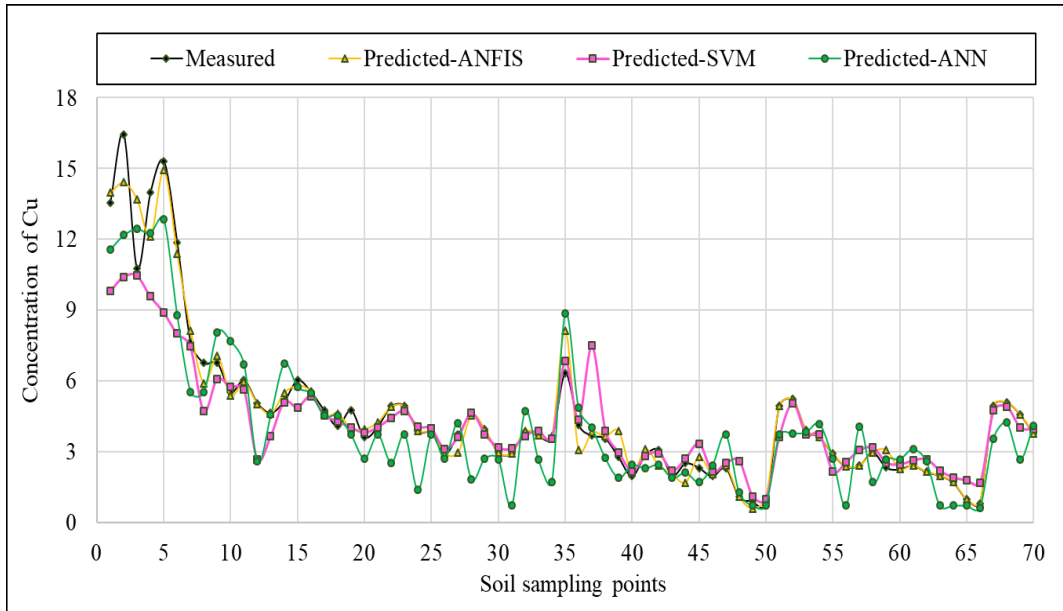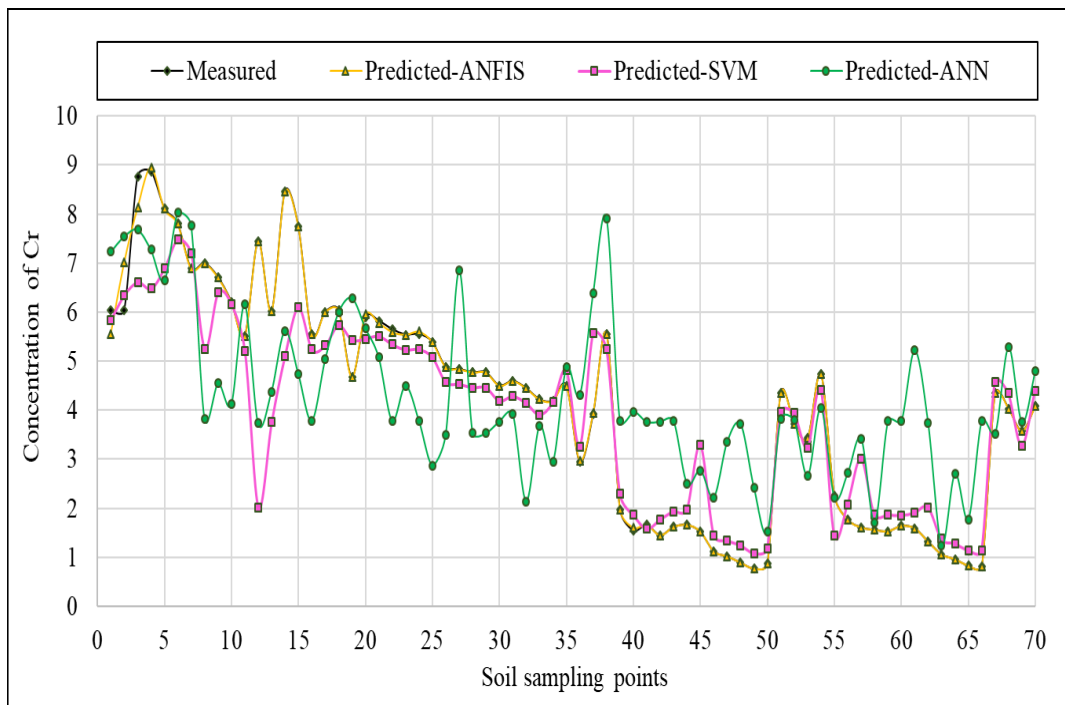

Figure G-1: Variation of predicted results of Cu in soil from various AI techniques in training.
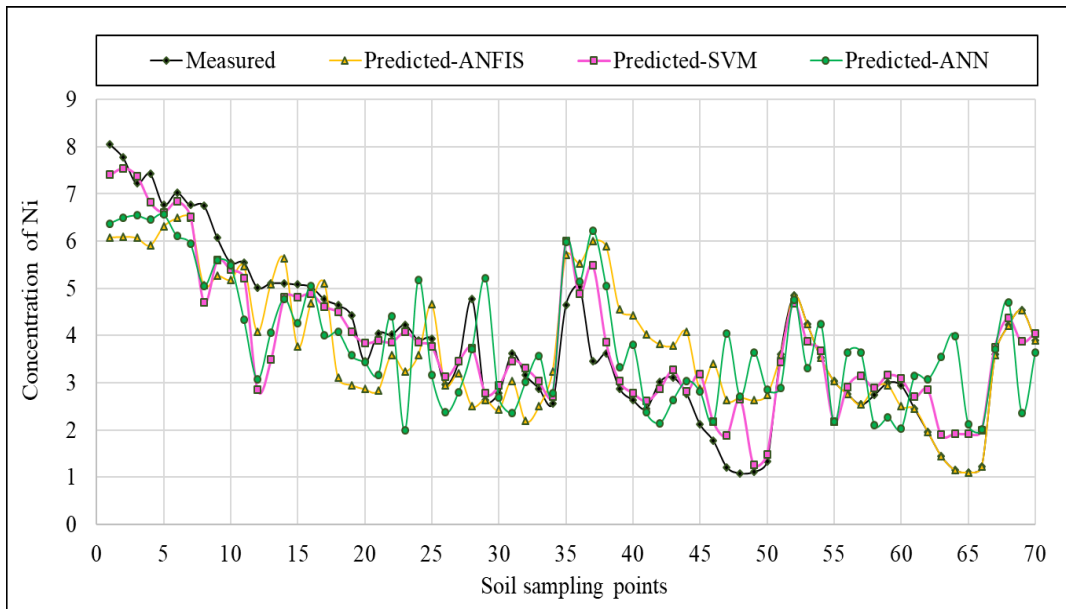
Figure G-2: Variation of predicted results of Ni in soil from various AI techniques in training.
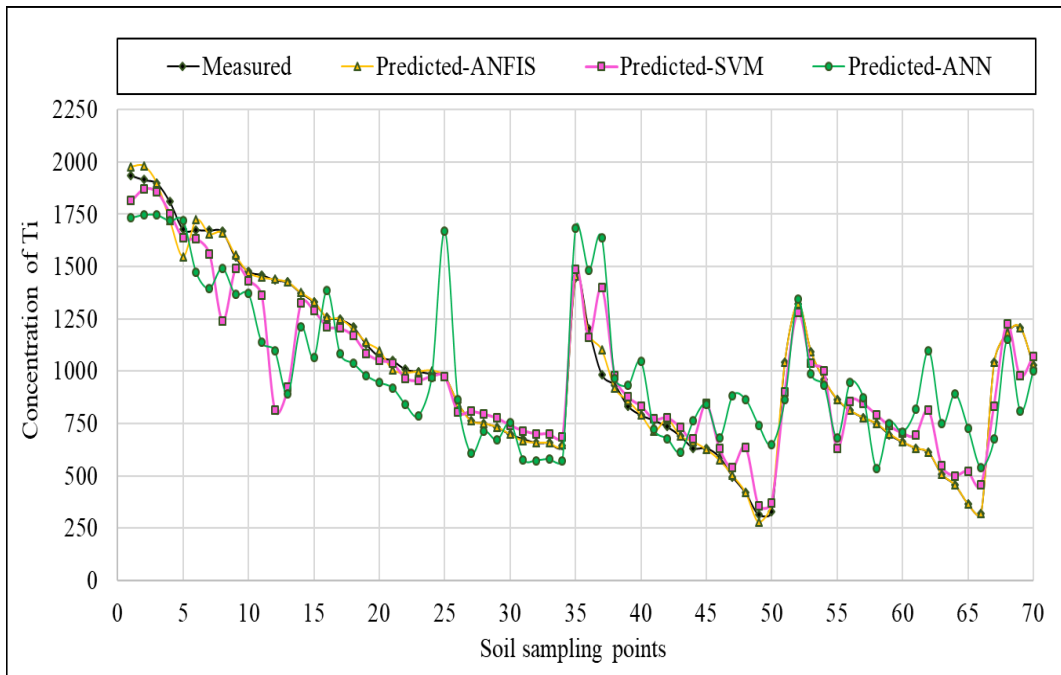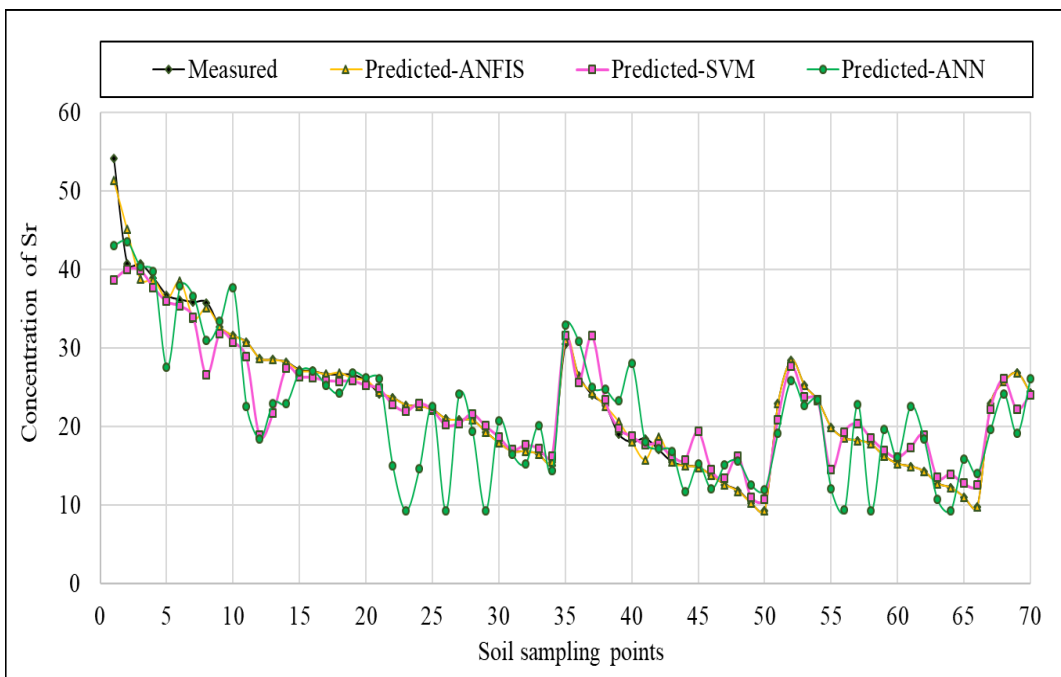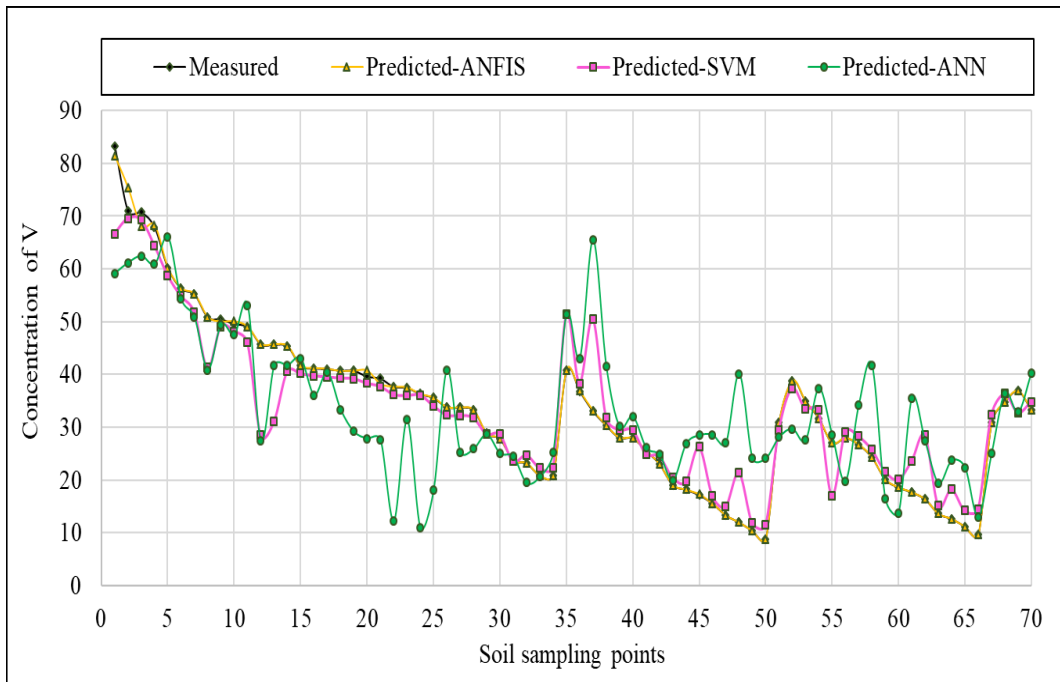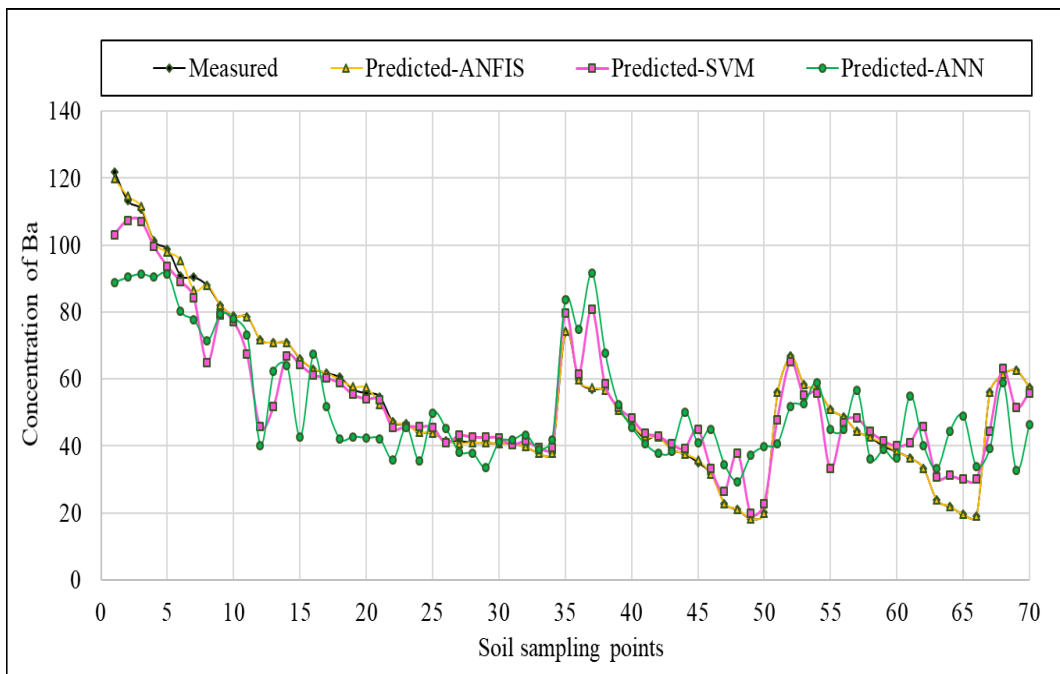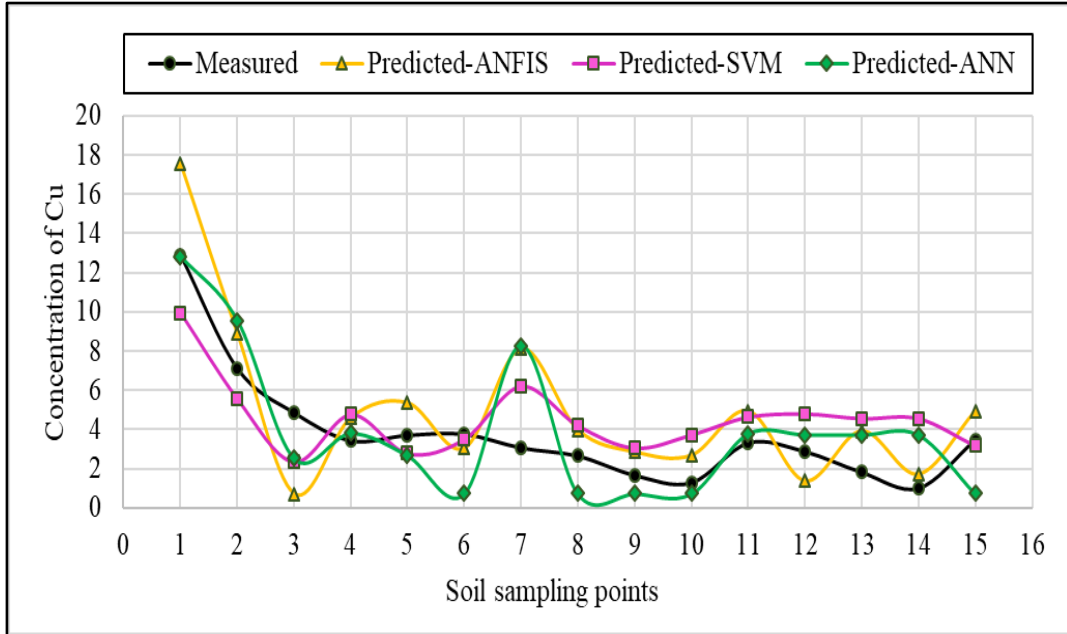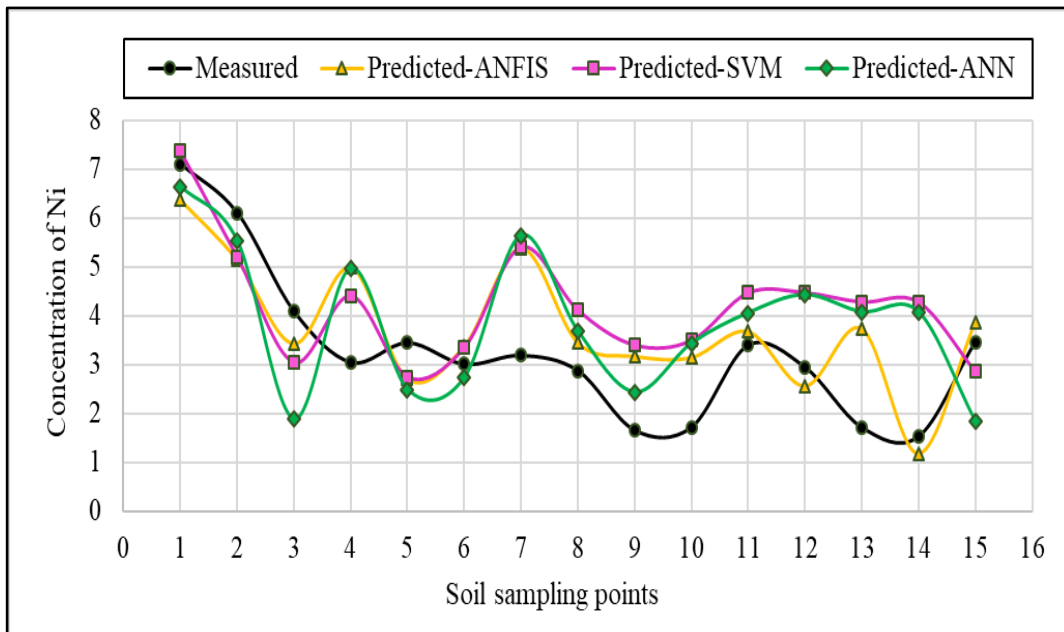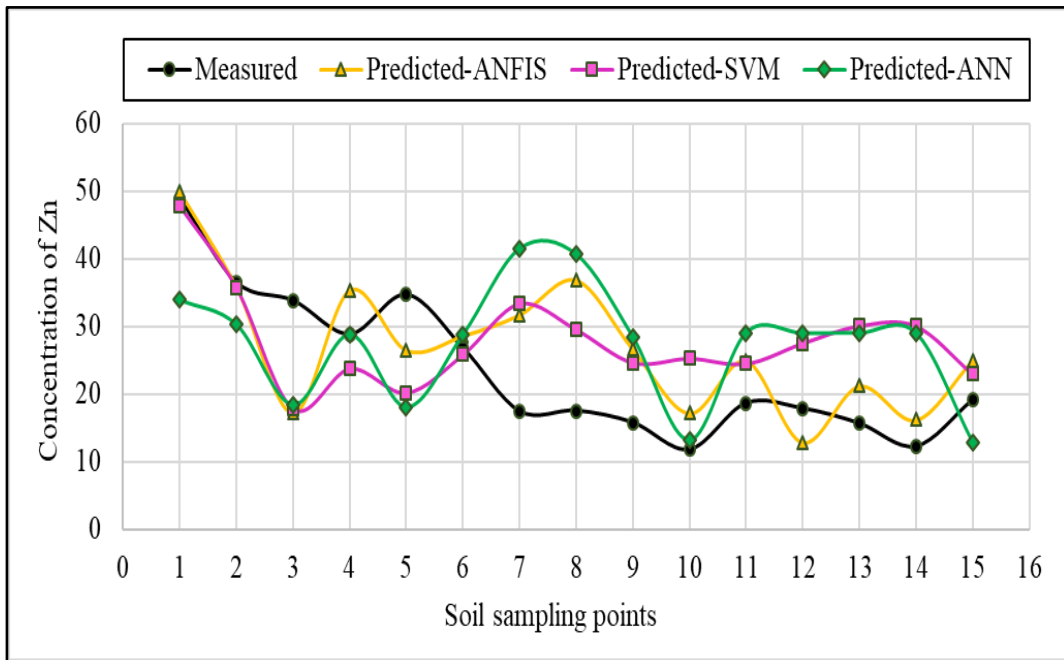
263
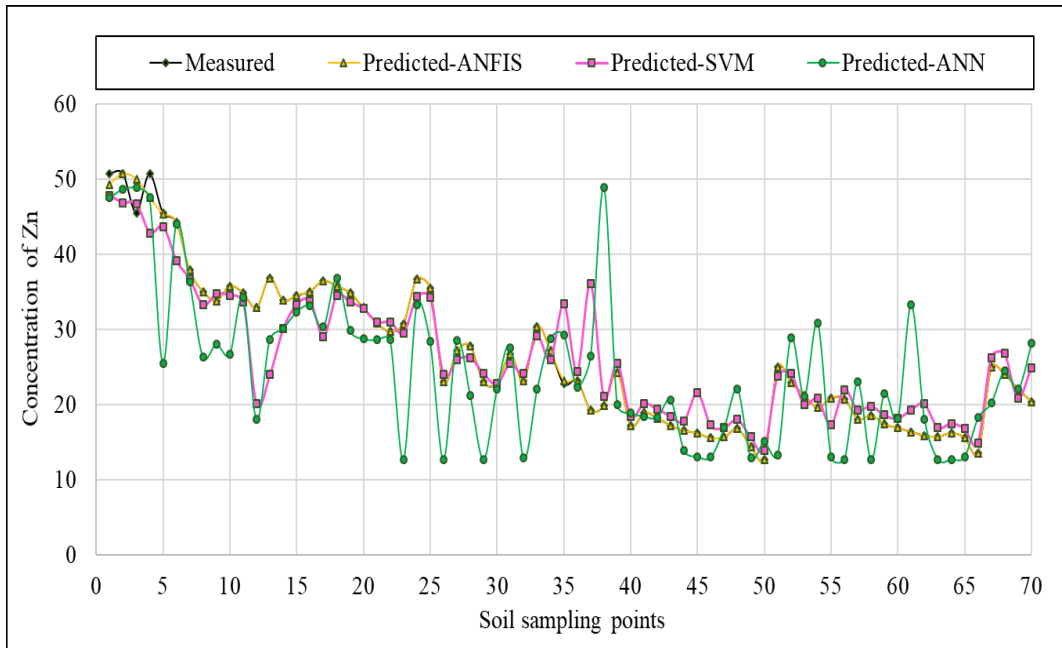
Figure G-3: Variation of predicted results of Zn in soil from various AI techniques in training.
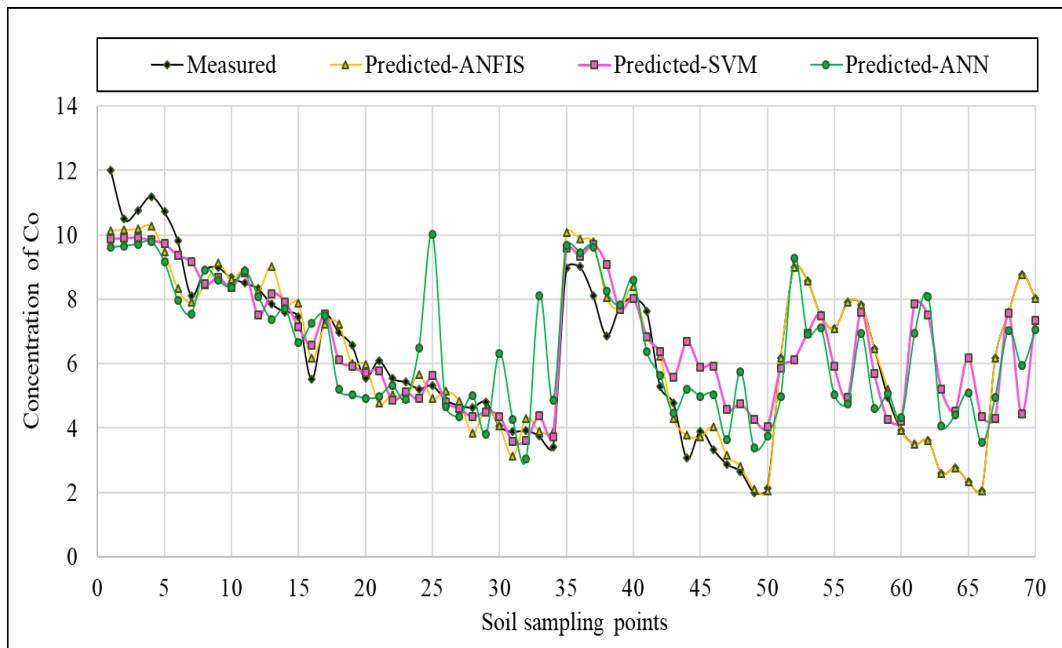
Figure G-4: Variation of predicted results of Co in soil from various AI techniques in training.

264
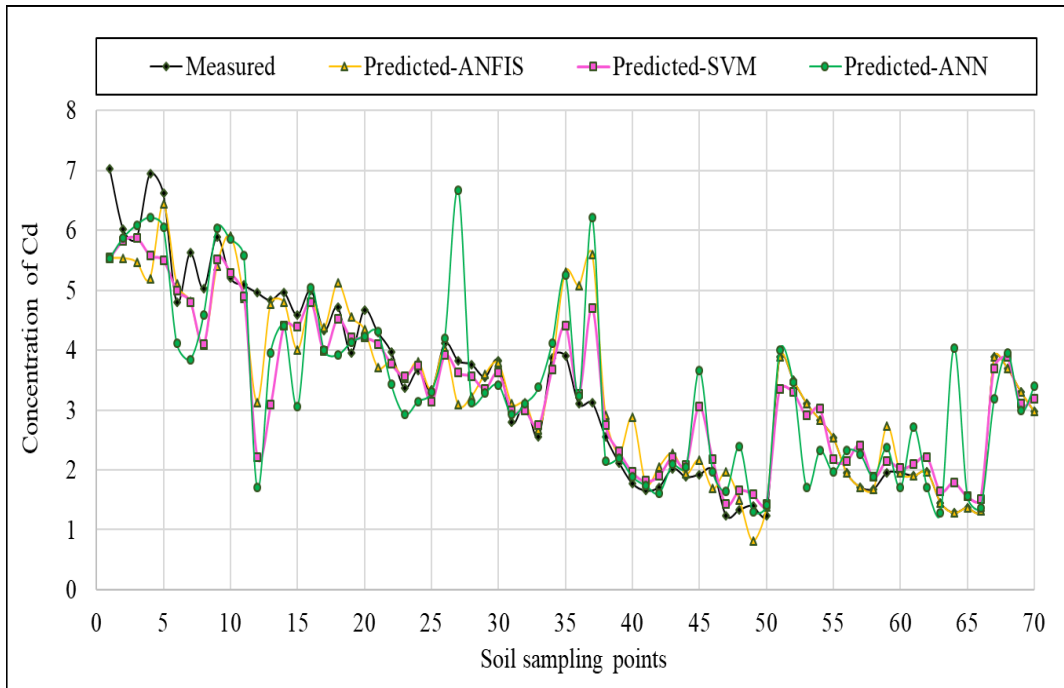
Figure G-5: Variation of predicted results of Cd in soil from various AI techniques in training.
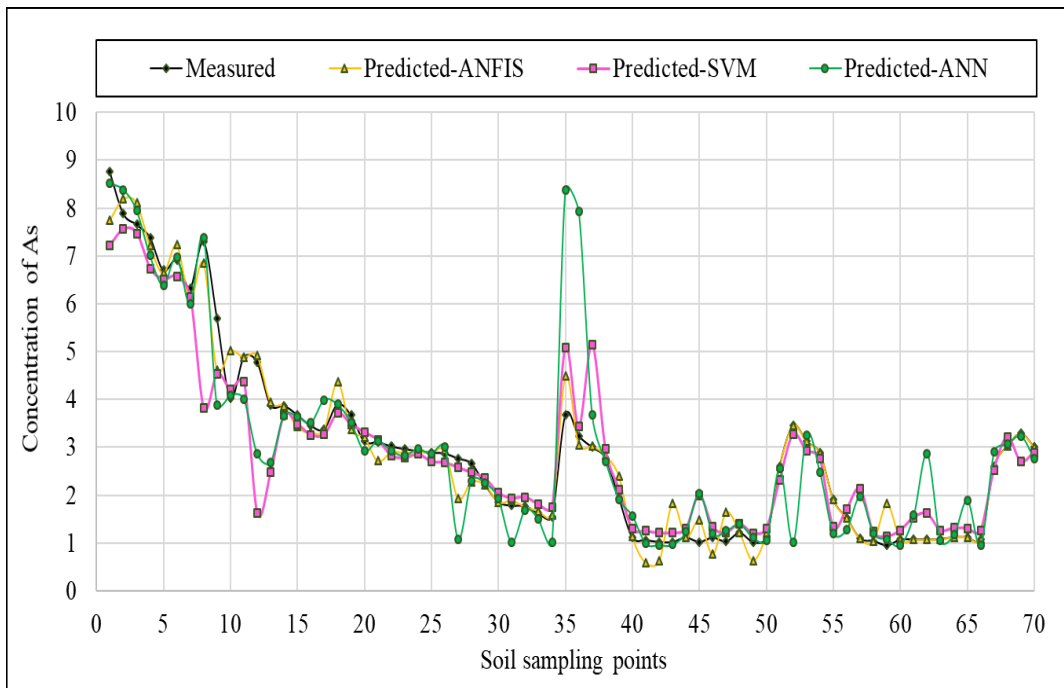
Figure G-6: Variation of predicted results of As in soil from various AI techniques in training.
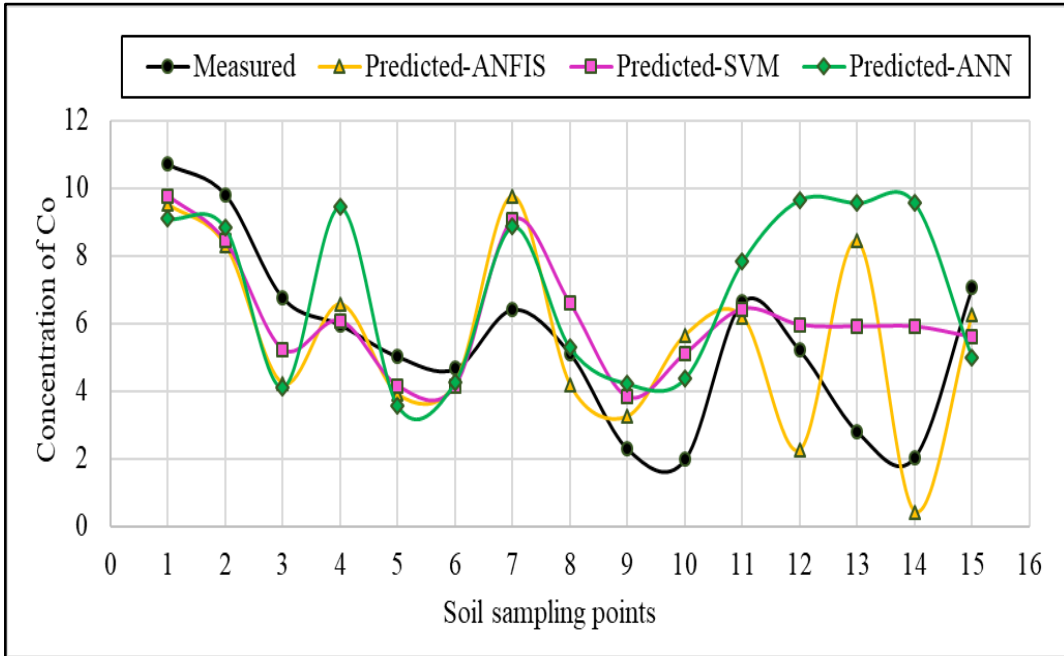
265
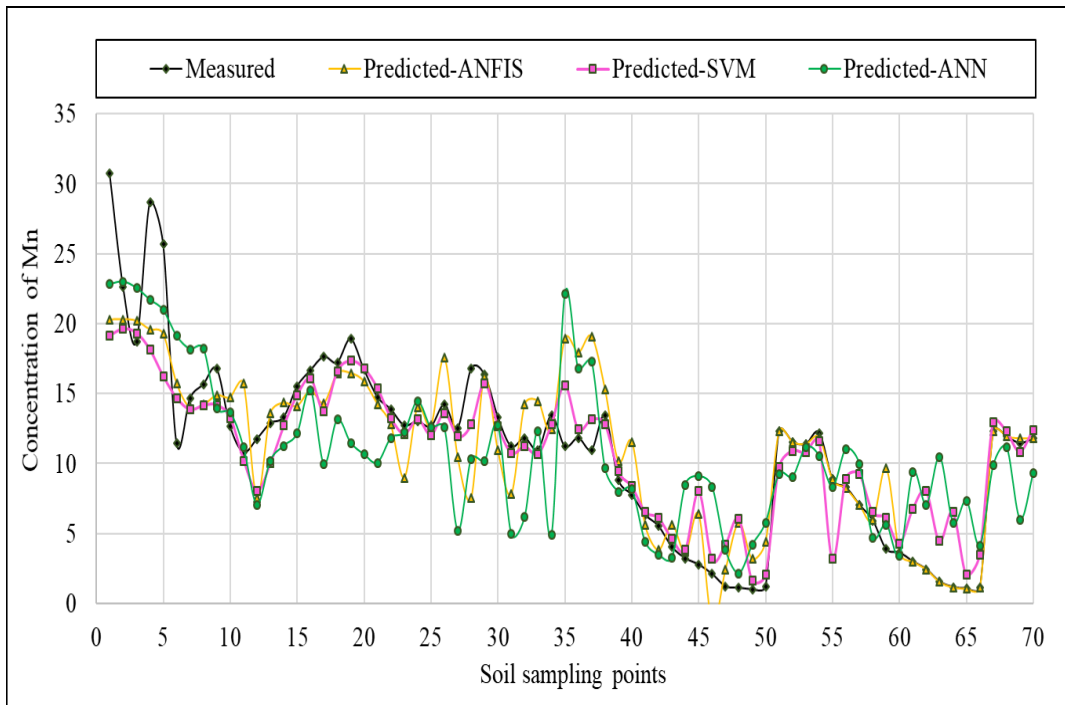
Figure G-7: Variation of predicted results of Mn in soil from various AI techniques in training.

Figure G-8: Variation of predicted results of Cr in soil from various AI techniques in training.
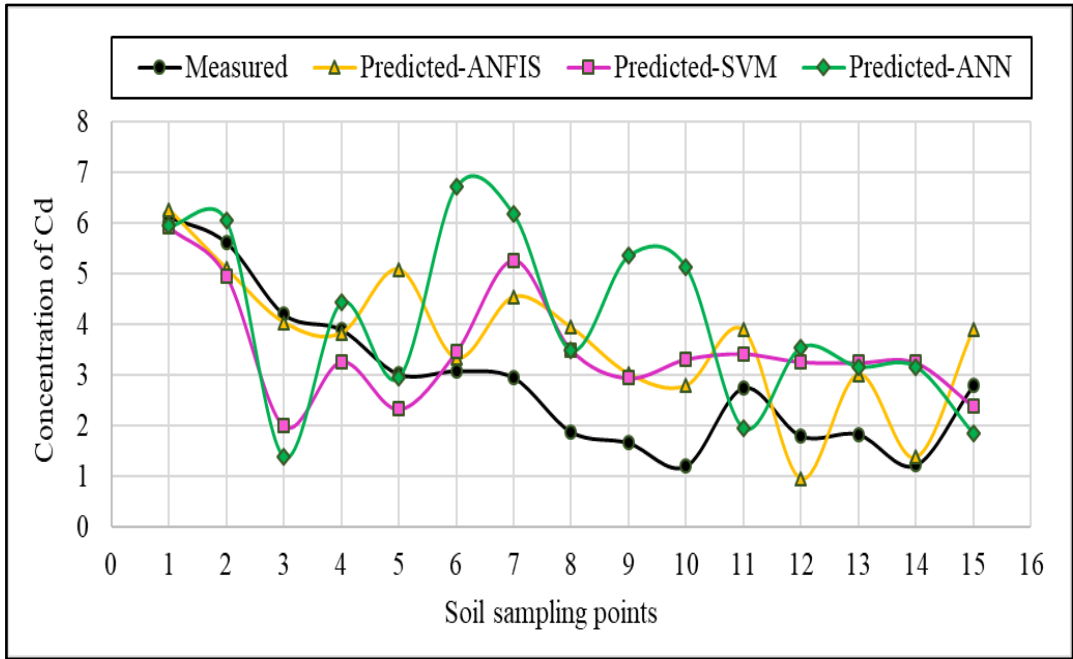
266

Figure G-9: Variation of predicted results of Ti in soil from various AI techniques in training.

Figure G-10: Variation of predicted results of Sr in soil from various AI techniques in training.

267

Figure G-11: Variation of predicted results of V in soil from various AI
techniques in training.

Figure G-12: Variation of predicted results of Ba in soil from various AI
techniques in training.

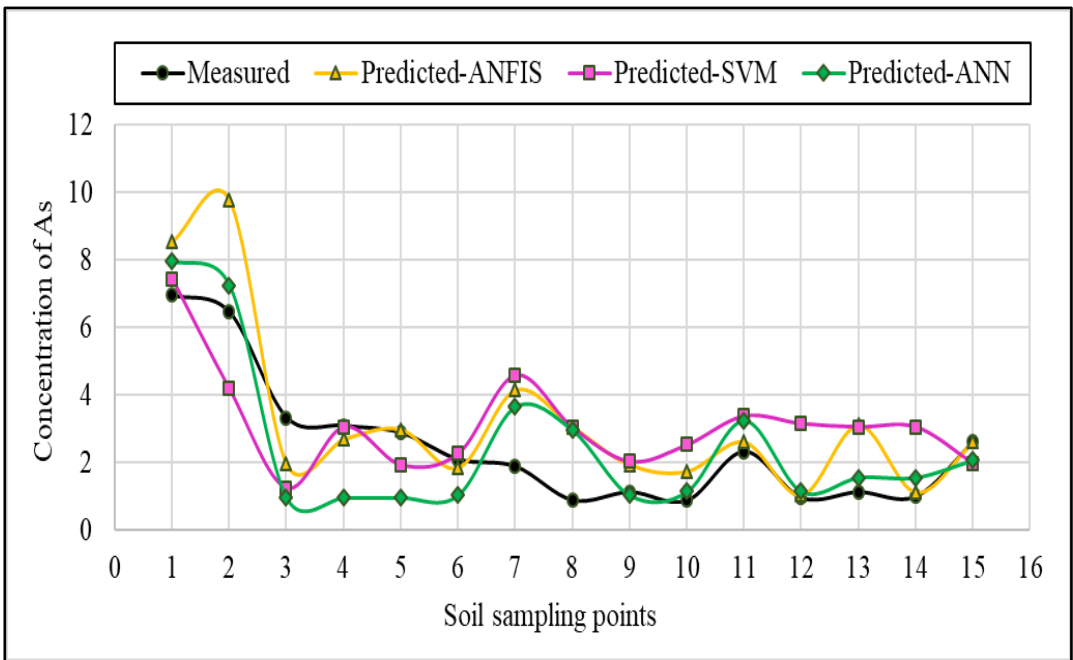**Variation of predicted results of heavy metals with different AI techniques for testing**

Figure H-1: Variation of predicted results of Cu in soil from various AI techniques in testing.

Figure H-2: Variation of predicted results of Ni in soil from various AI techniques in testing.
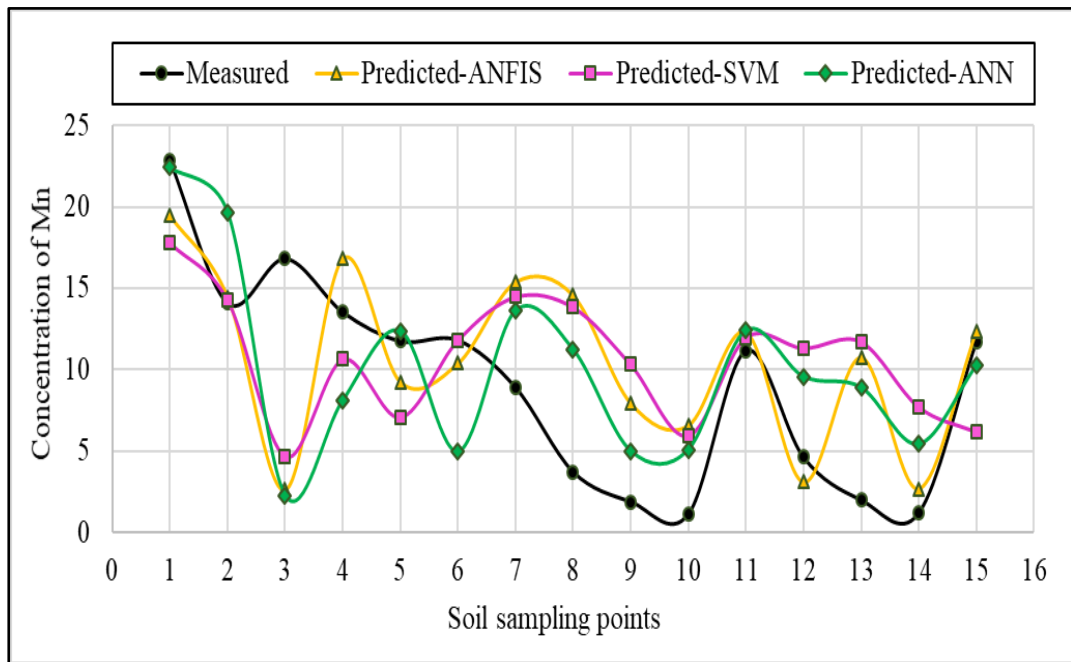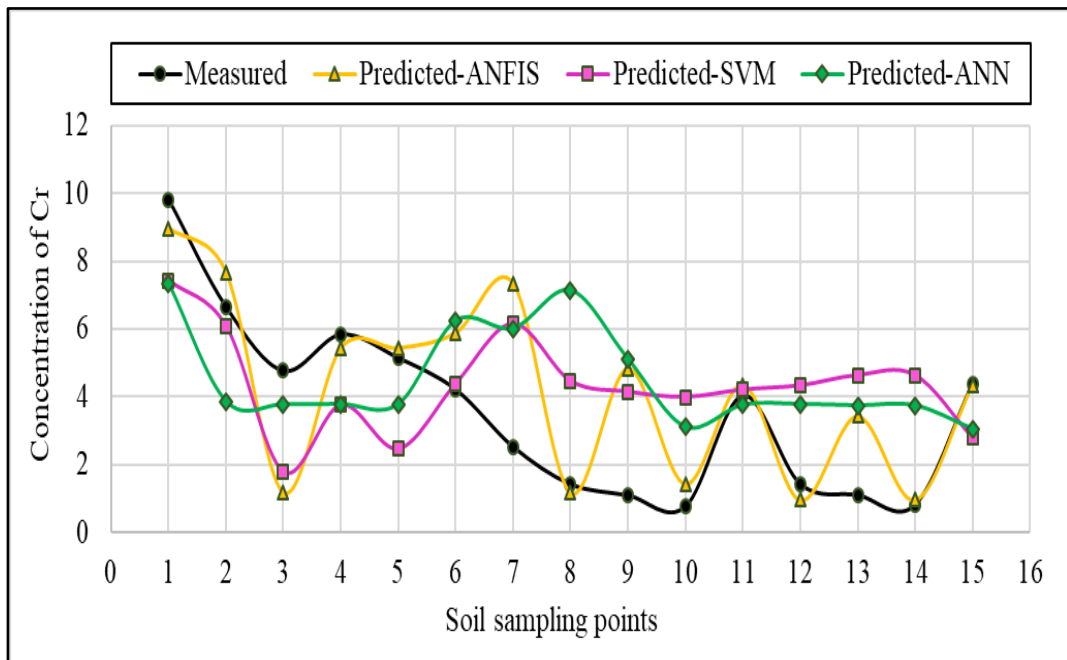
Figure H-3: Variation of predicted results of Zn in soil from various AI techniques in testing.

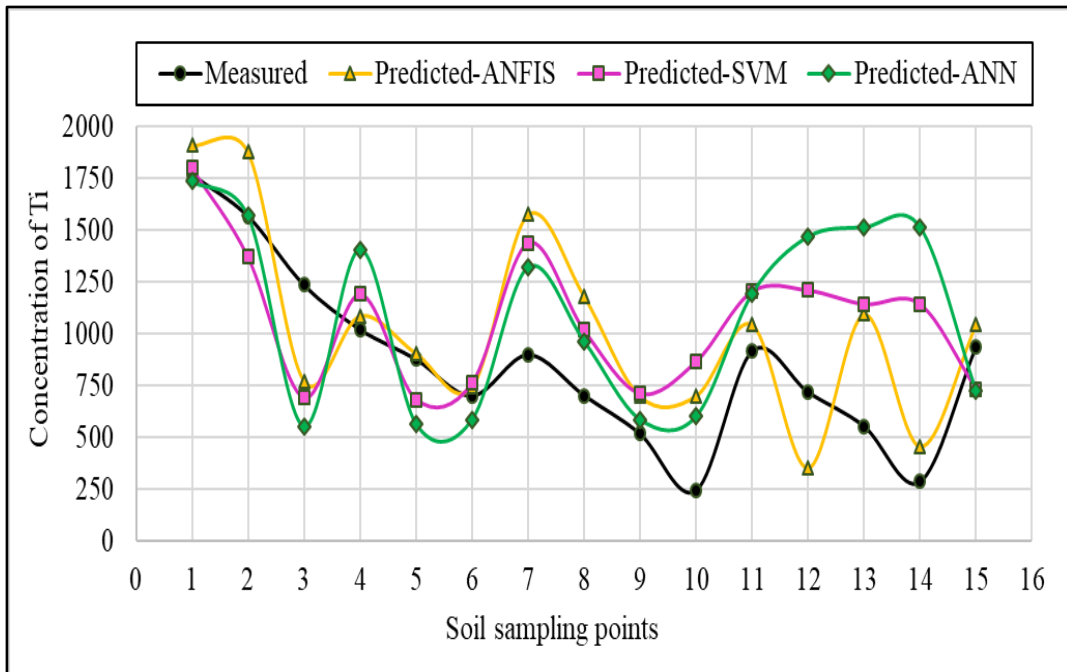Figure H-4: Variation of predicted results of Co in soil from various AI techniques in testing.

270

Figure H-5: Variation of predicted results of Cd in soil from various AI techniques in testing.

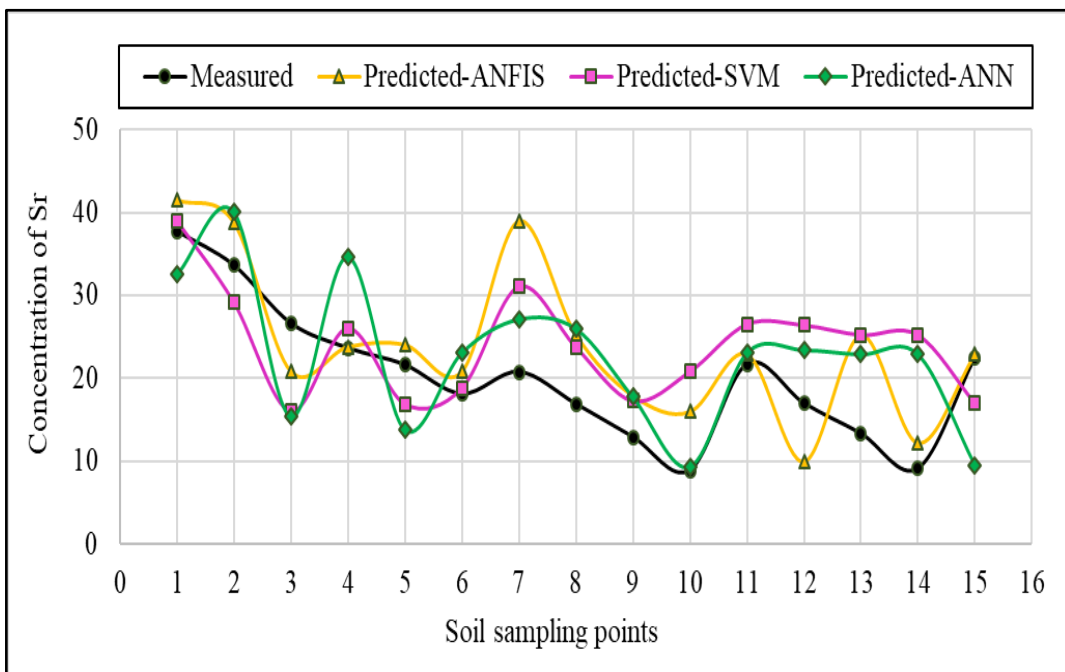Figure H-6: Variation of predicted results of As in soil from various AI techniques in testing.

271

Figure H-7: Variation of predicted results of Mn in soil from various AI techniques in testing.

Figure H-8: Variation of predicted results of Cr in soil from various AI techniques in testing.

272

Figure H-9: Variation of predicted results of Ti in soil from various AI techniques in testing.

Figure H-10: Variation of predicted results of Sr in soil from various AI techniques in testing.
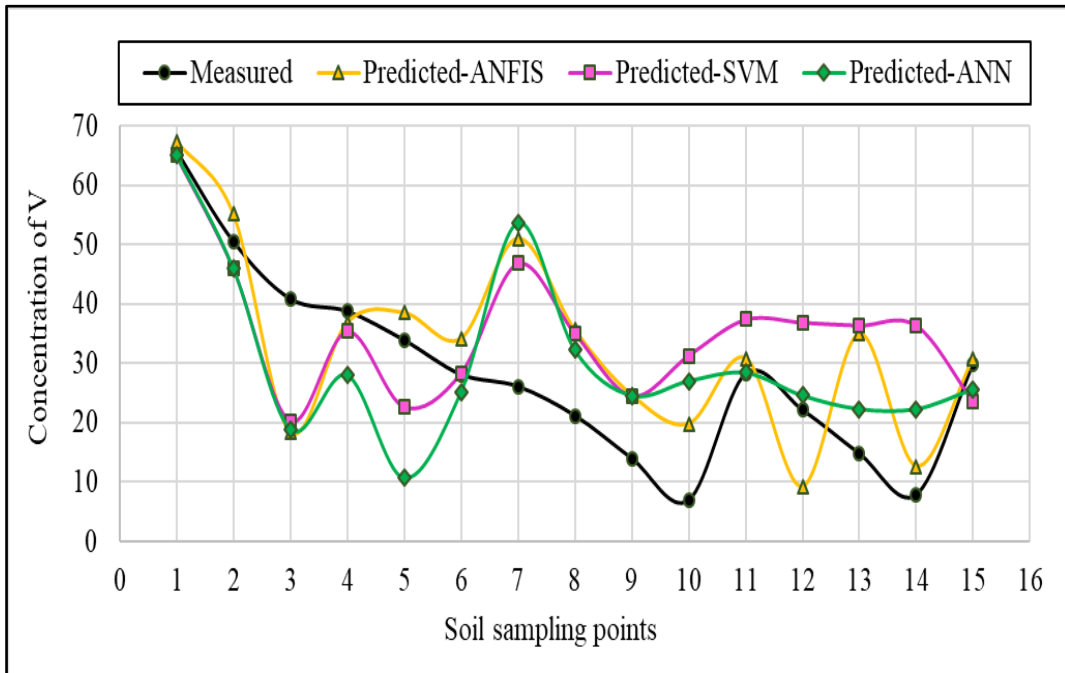
273

Figure H-11: Variation of predicted results of V in soil from various AI techniques in testing.
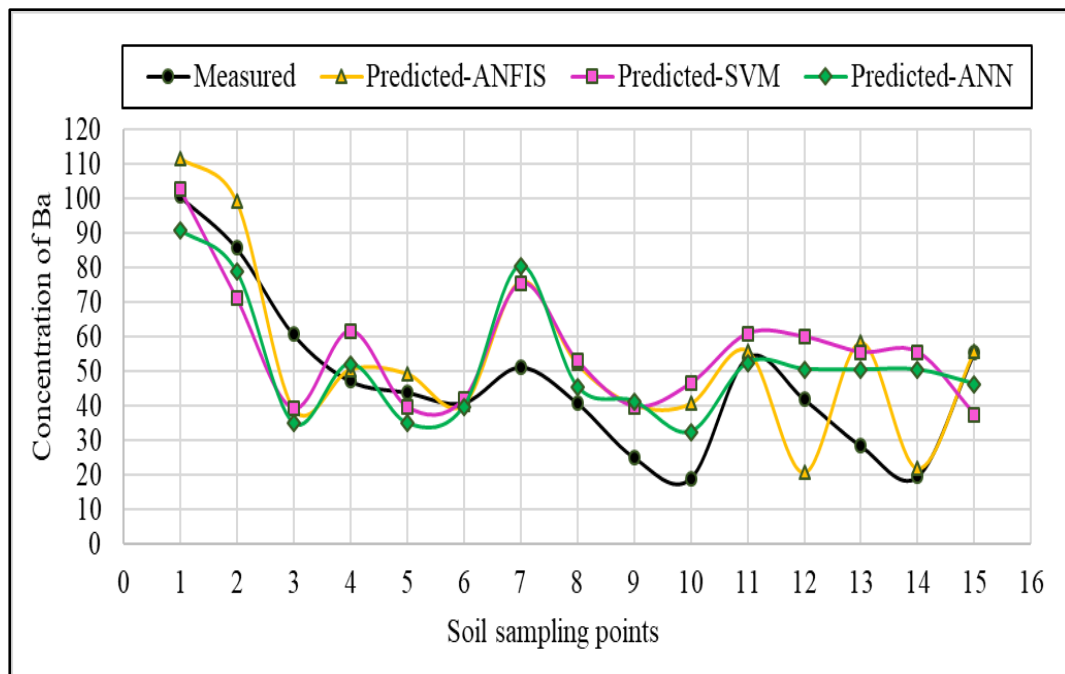

Figure H-12: Variation of predicted results of Ba in soil from various AI techniques in testing.

274