T.C.

SELÇUK UNIVERSITY

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

MODELLING OF ATMOSPHERIC SO_2 POLLUTION IN SEYDİŞEHİR TOWN BY ARTIFICIAL NEURAL NETWORKS

Zeynep Cansu ÖZTÜRK

MASTER OF SCIENCE
Environmental Engineering Department

December-2015
KONYA
All rights reserved

THESIS ACCEPTANCE AND APPROVAL

Thesis namely "Modelling of Atmospheric SO_2 Pollution in Seydişehir Town by Artificial Neural Networks" prepared by Zeynep Cansu ÖZTÜRK was accepted as MASTER OF SCIENCE THESIS by below jury members with unanimity/plurality on the date of 29/12/2015 at Environmental Engineering Department of Selçuk University the Graduate School of Natural and Applied Sciences.

Jury Members	**Signature**
Chair Prof. Dr. Ali TOR	.....
Supervisor Assoc. Prof. Dr. Şükrü DURSUN	.....
Member Assoc. Prof. Dr. Dünyamin GÜÇLÜ	.....

I approve the result above.

Prof. Dr. Aşır GENÇ
Head of FBE

This thesis was supported by Selçuk University BAP with project no of 15101016.

TEZ KABUL VE ONAYI

Zeynep Cansu ÖZTÜRK tarafından hazırlanan "Seydişehir İlçesi Atmosferik SO_2 Kirliliğinin Yapay Sinir Ağları ile Modellenmesi" adlı tez çalışması 29/12/2015 tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü Çevre Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Prof. Dr. Ali TOR	.....
Danışman Doç. Dr. Şükrü DURSUN	.....
Jüri Doç. Dr. Dünyamin GÜÇLÜ	.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Aşır GENÇ
FBE Müdürü

Bu tez çalışması Selçuk Üniversitesi BAP tarafından 15101016 nolu proje ile desteklenmiştir.

DECLARATION PAGE

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

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Zeynep Cansu ÖZTÜRK

Tarih: 22/12/2015

ABSTRACT

M. Sc. THESIS

MODELLING OF ATMOSPHERIC SO_2 POLLUTION IN SEYDİŞEHİR TOWN BY ARTIFICIAL NEURAL NETWORKS

Zeynep Cansu ÖZTÜRK

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF SELÇUK UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE
IN ENVIRONMENTAL ENGINEERING

Advisor: Assoc. Prof. Dr. Şükrü DURSUN

2015, 61 Pages

Jury
Prof. Dr. Ali TOR
Assoc. Prof. Dr. Şükrü DURSUN
Assoc. Prof. Dr. Dünyamin GÜÇLÜ

Air pollution has become a major environmental problem since last century because of the effects of fast population growth and industrial developments. Sulphur dioxide is considered as one of the major and most common air pollutant with using fossil fuels causing severe health problems such as disrupting tissues and mucous membranes of the eyes, disturbing nose and throat because of the irritating toxic odour, and affecting badly to upper part of respiratory system and bronchi. Seydişehir town of Konya was selected as working area for this study because heavy industrial activities are very wide in many fields such as mining and manufacturing industry. Also, usage of fossil fuels for heating system in winter period is other important atmospheric pollutants source. Eti Aluminium facility is the biggest industrial unite for SO_2 pollution source in Seydişehir town. In this study, SO_2 pollution in Seydişehir town was modelled with Artificial Neural Networks (ANN) which uses characteristics of biological neurons and capable of solving highly complex problems constructing parallel computations. Meteorological factors and previous day's SO_2 concentrations were integrated to model as input parameters and next day's SO_2 concentration was tried to be predicted. Two seasons were selected for model development namely winter and summer. Prediction performances of develop models are 67% for winter season and 81% for summer season. These values are compatible compared with previous studies using ANN modelling and can be improved with larger data sets.

Keywords: SO_2, ANN, Air Pollution, Modelling, Prediction

ÖZET

YÜKSEK LİSANS TEZİ

SEYDİŞEHİR İLÇESİ ATMOSFERİK SO_2 KİRLİLİĞİNİN YAPAY SİNİR AĞLARI İLE MODELLENMESİ

Zeynep Cansu ÖZTÜRK

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
Çevre Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Şükrü DURSUN

2015, 61 Sayfa

Jüri
Prof. Dr. Ali TOR
Doç. Dr. Şükrü DURSUN
Doç. Dr. Dünyamin GÜÇLÜ

Hava kirliliği, hızlı nüfus artışı ve endüstrinin gelişmesiyle birlikte geçen yüzyıldan buyana problem olmaya devam etmektedir. Kükürtdioksit fosil yakıt kullanımıyla birlikte ciddi sağlık sorunlarına neden olan önemli gazlardan biri olarak nitelendirilir. Bu sağlık sorunları arasında dokuların ve gözlerdeki mukus tabakasının zarar görmesi, rahatsız edici kokusu nedeniyle kulak ve burunda oluşan rahatsızlıklar ve üst solunum yolları ve akciğer problemleri sayılabilir. Bu tezde çalışma yeri olarak Konya'nın bir ilçesi olan Seydişehir seçilmiştir. Bölgede endüstrinin bir çok farklı alanda gelişmiş olması örneğin madencilik ve üretim sektörleri, hava kirliliğine katkıda bulunan faktörler arasındadır. Bunun yanı sıra ısınma amaçlı fosil yakıtların kullanılması da bölgedeki hava kirliliğini olumsuz etkilemektedir. Özellikle Türkiye'nin en büyük alüminyum tesisi olan Eti Alüminyum Fabrikası da SO_2 kirliliğine katkıda bulunmaktadır. Bu tez çalışmasında Seydişehir İlçesindeki SO_2 kirliliği yapay sinir ağları (YSA) ile modellenmiştir. YSA biyolojik nöronların karakteristik özelliklerini kullanır ve parallel bağ kurabilme yeteneği sayesinde kompleks problemleri çözüme ulaştırır. Meterolojik faktörler ve bir gün öncesine ait SO_2 değerleri modele girdi parametreleri olarak entegre edilmiş, bir gün sonra salınacak SO_2 konsantrasyonları tahmin edilmiştir. Model geliştirilirken yaz ve kış olarak iki mevsim dikkate alınmıştır.Geliştirilen modellerin tahmin performansları kış mevsimi için %67, yaz mevsimi için %81 çıkmıştır. YSA ile yapılan önceki çalışmalara bakıldığında bu sonuçlar tatmin edicidir fakat daha büyük veri setleri kullanılarak, geliştirilen modellerin tahmin yetenekleri arttırılabilir.

Anahtar Kelimeler: SO_2, YSA, Hava Kirliliği, Modelleme, Tahmin

PREFACE

I would like to express my deepest gratitude and appreciation to my supervisor Assoc. Prof. Dr. Şükrü Dursun for his excellent guidance, attitude and patience. The trust that he put on me always kept me confident and motivated. I feel very lucky to have a supervisor whom I see as my mentor. Also, I want to thank to Dr. Fatma Kunt for her support and valuable information during preparation of my thesis and her friendship.

Kiraz Kök should be mentioned in here for her kind help during measurements of pollution data which is the vital part of this thesis. I would like to thank to her for supporting me preparing my study.

General Directorate of Meteorology supported this thesis by providing meteorological data is greatly acknowledged. Moreover, I want to thank to Selçuk University BAP for their financial support with project no of 15101016.

I thank sincerely to my close friends Merve Soğancıoğlu and Hamza Aysan for their support and help. Also, I want to thank my dear friend Akın Y. Ayturan for his helps on MATLAB and all my lovely friends whose give me motivation.

My parents Mustafa and Melek Öztürk deserve very special thanks, as they have been supportive in every step of my education. I also thank to my brother Murat and his lovely wife Nilüfer for their support. Of course my special thanks go to my dear sister Göksu for her endless energy and joy which give me all the time.

Zeynep Cansu ÖZTÜRK
KONYA-2015

TABLE OF CONTENTS

ABSTRACT ..... **iv**

ÖZET ..... **v**

PREFACE ..... **vi**

TABLE OF CONTENTS ..... **vii**

SYMBOLS AND ABBREVIATIONS ..... **iv**

1. INTRODUCTION **1**

2. LITERATURE REVIEW **4**

2.1. Air Pollution 4
2.1.1. History of Air Pollution 4
2.1.2. Sources of Air Pollution 5
2.1.3. Main Air Pollutants 6
2.1.4. Effects of Air Pollutants 10
2.2. Artificial Intelligence 13
2.2.1. Artificial Intelligence and Engineering 13
2.3. Previous studies using ANN for air pollution modelling 22
3. MATERIAL AND METHOD **25**

3.1. Working Area 25
3.1.1. Climate of working area 26
3.1.2. Geographical situation of working area 30
3.1.3. Air pollution sources of working area 31
3.2. Data gathering 31
3.2.1. SO_2 gas analyser device 31
3.2.2. Provision of meteorological data 36
3.3. Data analysis 36
3.4. Modelling with artificial neural network 38
3.4.1. Modelling with winter data 40
3.4.2. Modelling with summer data 42
4. RESEARCH RESULTS **44**

4.1. Results of winter model 44
4.2. Results of summer model 45
4.3. Evaluation of the effects of input parameters to SO_2 pollution 49
5. DISCUSSION **52**

6. CONCLUSION AND SUGGESTIONS **55**

REFERENCES **57**

CURRICULUM VITAE **62**

SYMBOLS AND ABBREVIATIONS

AI : Artificial Intelligence
AMOS: Automatic Meteorological Observation Station
ANFIS: Adaptive Neuro-Fuzzy Inference System
ANN : Artificial Neural Network
As : Arsenic
BP : Back Propagation
C_6H_6 : Benzene
Cd : Cadmium
CFCs : Chlorofluorocarbons
CO : Carbon monoxide
CO_2 : Carbon dioxide
dm^3 : cubic decimetre
E : East
EPA : Environmental Protection Agency of United States
H_2O : Water
H_2S : Hydrogen sulphide
H_2SO_3: Sulphurous acid
H_2SO_4: Sulphuric acid
HCl : Hydrogen chloride
HCs : Hydrocarbons
HF : Hydrogen fluoride
Hg : Mercury
HNO_3 : Nitric acid
hv : Energy of the photon
IARC : International Agency for Research Cancer
km : kilometre
kW/m^2: kilowatt per meter square
L : litre
m/sec : meter per second
m_2 : meter square
mb : millibars
mm : millimetres
MSE : Mean Square Error
N : North
N_2 : Nitrogen
N_2O : Nitrous oxide
NH_3 : Ammonia
NH_4 : Ammonium
nm : nanometre
NN : Neural Network
NO : Nitrogen monoxide
NO_2 : Nitrogen dioxide
NO_3 : Nitrate
NO_x : Nitrogen oxides
O_2 : Oxygen
O_3 : Ozone
^oC : Celsius degree
OH : Hydroxyl

Pb	: Lead
PM	: Particulate Matter
PMT	: Photo Multiplier Tube
ppb	: part per billion
ppm	: part per million
R	: Correlation coefficient
R^2	: Coefficient of determination
S	: South
SO_2	: Sulphur dioxide
SO_3	: Sulphur trioxide
SO_4^{2-}	: Sulphate
UV	: Ultraviolet
VOC	: Volatile Organic Compound
W	: West

1. INTRODUCTION

Air pollution and its effects have become global issues since middle of 19s. Air pollution is transported to long distances with air movements and has global effects. The major concerns should be cautious that greenhouse effect major cause of global warming and depletion of ozone layer with effect of many primary and secondary pollutants (Zannetti 1990). Major primary air pollutants which are contaminants causing some adverse effects on environment are particulate matter (PM), sulphur compounds (e.g., SO_2, H_2S), nitrogen compounds (e.g., NO, NH_3), carbon compounds (e.g., HCs, CO, CO_2), halogen compounds (e.g., fluorides, bromides, chlorides) (Zannetti 1990). Moreover, chemical reactions happening in the atmosphere causes the transformation of primary pollutants to secondary ones, for example, SO_2 gas transforms to SO_4^{2-} (Zannetti 1990).

Sulphur dioxide emissions are considered as one of the major and most common air pollution problem in the world. SO_2 emissions are responsible from London winter-type smog event and other many lethal smog events. Because of acid deposition problem, water bodies and animal life is also affected from SO_2 gas. Main anthropogenic sources of SO_2 are fuel combustion, petroleum, mining, industrial processes like paper production (Zannetti 1990).

In Turkey with the development of the industry and population growth air pollution has started to become a problem since beginning of 1980s. Seydişehir is a town of Konya which is one of the biggest cities of Turkey. In Seydişehir industrial activities are very wide in many fields such as mining and manufacturing industry. Especially mining has significant importance because of the minerals, bauxite, chromium and lignite, extracting from the area. Both extraction and manufacturing steps of minerals produce huge amount of emissions. Moreover, other manufacturing industries like chemical production, agricultural material production, and food production emits SO_2 gas to the atmosphere (URL 1). Also, fossil fuel combustion from residential buildings and other institutions for heating purposes largely contributes to SO_2 pollution in Seydişehir.

Eti Aluminium Facility is the biggest metal production and the only aluminium manufacturing facility of Turkey. It was constructed in closed area of 12,000,000 m²

(URL 2). During aluminium production Sulphur in the anodes is oxidized, releasing SO_2 from the potlines which is a row of electrolytic cells used in the production of aluminium as the anodes are consumed. SO_2 is also released from the anode bake furnace as pitch used to help form the anodes are oxidized during the baking process (URL 3). As a result of this anode process during aluminium production, Eti Aluminium facility contributes SO_2 pollution in Seydişehir town.

All of these sources of SO_2 in Seydişehir make it a problem for town. SO_2 is not lethal in all concentrations but it decreases life quality for human in many manners such as giving harm to tissues and mucous membranes of the eyes, disturbing nose and throat because of the irritating odour, and affecting badly to upper part of respiratory system and bronchi (Hussainn 2011). Also, it affects the animal life, vegetation and environment badly. All these negative effects of SO_2 pollution are considerable.

Nowadays prediction of air pollutants is increasing trend because with the help of future prediction of air pollutants establishing emission control legislations, evaluating the future emission control impacts, selecting possible locations in which it can be possible source of air pollutants, controlling air pollutant episodes and taking preventive precautions, and assessing present air pollutant sources and their responsibility in future events may be achieved (Zannetti 1990). In this study a model was established in order to predict SO_2 pollution in Seydişehir town. There are many techniques used for prediction modelling such as statistical models, neural networks, fuzzy logic. Artificial neural network (ANN) was used for this study and MATLAB (R2011a) software was used for ANN development.

ANNs are structures which consist of interconnected simple adaptive elements having ability to make parallel computations for processing and representation of knowledge (Basheer and Hajmeer 2000). It has many advantages when it’s compared with traditional models because these models only require known input data without any assumption (Gardner and Dorling 1998). With the help of suitable connecting weights and transfer functions, approximation of a multilayer function can be measurable function between input and output (Hornik et al. 1989, Gardner and Dorling 1998). Most of the parallel structured ANN models include intense interconnected adaptive

units. ANN models have very significant characteristic that highly nonlinear problems may be solved (Zurada 1997).

Back propagation (BP) feedforward algorithm used for training the multilayer neural networks is the most famous learning algorithm used in ANN models (Basheer and Hajmeer 2000). In this study Multilayer Feedforward Neural Network including input layer, hidden layer and final layer consisting of neurons was used.

Meteorological parameters which are hourly average temperature (^{o}C), hourly average pressure (bar), hourly relative humidity (%), hourly wind speed (m/sec), hourly rainfall (mm), hourly cloudiness, hourly sun duration, hourly sun radiation (kW/m^2) and previous day's hourly SO_2 pollution concentration (ppb) was integrated to model as input parameters and next day's SO_2 concentration was tried to be predicted. SO_2 data used in this model was gathered between 2012-2013 years just on particular months by M100E UV Fluorescence SO_2 Analyser device and meteorological data measured with automatic meteorological observation station (AMOS) was provided from General Directorate of Meteorology. Gathered data was separated to two datasets namely, winter and summer and one model was developed for each.

Next section of thesis literature was reviewed and previous studies with similar content were investigated. Detailed information related data gathering and model development is given in Section 3. Last sections include information related to research results.

2. LITERATURE REVIEW

2.1. Air Pollution

Air pollution is waste remaining from the ways humans produce their goods, transport themselves and their goods, and generate energy to heat and light the places they live (Vesilind et.al 2010). The amount of gaseous and particulate materials in the air which harms to the living organisms increases because majorly of the combustion of fossil fuels. Impurities in the fuel, poor air-to-fuel ratio, too high and low combustion temperatures lead to pollutants (Boubel et.al 1994).

2.1.1. History of Air Pollution

In early history air pollution firstly became a problem in 1157 because of the burning of the wood in Tutbury Castle in Nottingham. In 1273 coal burning is prohibited in London. After this prohibition, in 1306 sea coal started to be used in furnaces according to the proclamation of King Edward I. In between 1377 and 1399 years Richard II and between 1413 and 1422 years Henry V took steps to regulate the use of coal in England. King Charles II was published a booklet related to air pollution in 1661. After industrial revolution which was a consequence of the power production using steam to move machinery and pump water, coal combustion increased in order to provide necessary boiled steam for steam engines (Boubel et.al 1994). The main pollutants of coal combustion process were smoke and ash in nineteenth century. Reduction of these pollutants was considered as a health issue by first Public Act of 1848. In United States the first regulations related to smoke and ash were prepared im1880s considering reduction of the pollutants from industry, locomotives and domestic sources. Moreover, the first technological developments against air pollution were seen in nineteenth century. Scrubbers for removing acid gases from effluent gas streams, cyclones and bag house collectors were designed for collection of big size particulate matters (Boubel et.al 1994).

In first part of the twentieth century there were great changes in technology. The replacement of steam engine with electric motor was the most important development. Also, very effective air pollution control device, electrostatic precipitator was invented.

However, there were no significant changes related to the legislation, regulation and understanding of the problem. During period between 1925 and 1950 the present air pollution problems and their solutions were started to be investigated. In this period there were episodes causing death of thousands of people. The technical foundation for searching air pollutants and meteorological events was established. The major technological change during that period was the construction of natural gas pipelines. This reduced the amount of coal combustion emissions and improved air quality. Also, diesel locomotives were started to be used instead of steam locomotives (Boubel et.al 1994).

After 1950s both in European countries and United States some precautions had taken related to air pollution and within company of many countries first air pollution control act was signed. Also, national air pollution research centers opened in different countries by 1980. The importance of meteorology was understood and first mathematical models of the pollution of the atmosphere began to be developed. In the United States Clean Air Act Amendments was accepted in 1990 thus the importance of stratospheric ozone depletion and global climate changes was accepted (Boubel et.al 1994).

2.1.2. Sources of Air Pollution

Sources of air pollution are separated as natural and anthropogenic. Natural air pollution sources are volcanic eruptions, forest fires, dust storms, the oceans, plants and trees of the earth. Volcanic eruptions produces huge amount of particulate matter and other pollutant gases like SO_2, H_2S, and methane. These gases and particulate matter stay in the air very long time. Forest fires are another main source of natural air pollution. Because of the uncontrolled forest fire, smoke, hydrocarbons, carbon dioxide, carbon monoxide, nitrous oxides, and ash are emitted with large amounts. Dust storms are the main source of particulate matter pollution which can cause problems to the householders, industry and mobile transporters. Oceans also create air pollution problem since they emits salt form aerosols to the atmosphere that materials lead to corrosion problems. The effect of wave on shells and the surface of rocks erode them to sand particles and make them airborne particulate matter. Moreover, the green plants and trees are major source of hydrocarbons (Boubel et.al 1994).

Anthropogenic sources of air pollution are classified as the industrial sources, utilities, and personal sources. The main source of industrial air pollution originates from raw materials in manufacturing processes. For example, during mining activity dust and SO_2 emissions, during briquetting & coking of coal dust, gases and impurities of coal, during metal smelting SO_2 and various volatile metals such as Hg, As, Pb, Cd, and from chemical industry HCl, HF, H_2S, NO_x, NH_3, HCs, VOC are emitted. Utilities are the important source of anthropogenic air pollution because most of them produce electricity by converting energy. This procedure emits huge amount of carbon dioxide, nitrous and sulphur oxides to the atmosphere. Moreover, personal sources such as mobile vehicles, furnaces and stoves in homes, barbeque grills, and burning of leaves in open area contribute to anthropogenic air pollution (Boubel et.al 1994).

2.1.3. Main Air Pollutants

2.1.3.1. Particulate Matter

Particulate matter (PM) which is also known as particulate pollution composed of very small particles and droplets of liquid. Many constitutions such as acids of sulphide and nitrate, organic compounds, metals and dust particles produce particulate pollution. Moreover, particle size is important factor affecting the health of all living creatures. Especially particles equal or smaller than 10 micron are very important for investigators and they has to be removed since these particles can pass through the nose and throat and affect the lungs and heart. This can lead to very serious health problems (URL 4).

Particulate matters or pollutants are classified in two categories according to Environmental Protection Agency of United States (EPA). These are named as inhalable coarse particles and fine particles. The diameter of inhalable course particles is bigger than 2.5 microns and smaller than 10 microns. Fine particles have diameter smaller than 2.5 microns. These particles are found in the atmosphere as a form of smoke and haze and emitted directly form forest fires and power plants (URL 4).

2.1.3.2. Sulphur-containing compounds

The most important sulphur containing gases in the atmosphere consist of sulphur dioxide (SO_2), sulphur trioxide (SO_3) and hydrogen sulphide (H_2S). Along with these gases, sulphuric acid (H_2SO_4) and sulphate ($SO_4^{=}$) salts are found in the atmosphere as sulphur compounds (Kırımhan 2006). The first group of gases constitutes primary pollutants and the second ones are named as secondary pollutants which are somehow produced from primary ones (Vesilind 1975).

Sources of sulphur containing compounds are combustion of fossil fuels, combustion and decomposition of organic materials, volcanic eruptions, and interference of sulphate salts with atmosphere from oceans by powerful winds and emission of sulphuric compounds because of mining activities. Sulphuric compounds emitted by these sources can return back to the earth surface with wet or dry deposition (Kırımhan 2006). Moreover, sulphuric compounds in the atmosphere are oxidized by atomic oxygen, oxygen and ozone and they can produce new compounds as a result of interaction with some materials by photochemical and chemical process. Because of this transformation, pollutants emitted from different sources enter continuous cycle with air, water and soil (Kırımhan 2006).

Importance of SO_2 is higher than the other sulphur containing compounds for air pollution because SO_2 amount in the atmosphere is always associated with the public health. Besides, SO_2 is evaluated with particulate matters and together with PM produces air pollution index causing health issues. SO_2 creates respiratory problems and especially lethal for lung failure and respiratory insufficiency patients (Müezzinoğlu 1987).

According to national ambient air quality standards of EPA that are prepared for pollutants harmful for public health and environment, SO_2 amount in the atmosphere should be smaller than 75 ppb hourly as a primary pollutant and it should be less than 0.5 ppm hourly as a secondary pollutant (URL 5).

Ambient Air Quality Standards of Turkey is given in the Regulation on Air Quality Assessment and Management. In the Annex 1 of this regulation Ambient Air Quality

Standards of Turkey are listed clearly. The main purpose of these standards is reaching the long term targets on the determined date and not exceeding the determined limit values. In order to reach this goal, governments use air pollution control mechanisms and try to decrease the limits for the further years. Regulation consists of SO_2, NO_2, N0x, PM_{10}, Pb, C_6H_6, CO, O_3 together with Arsenic, Cadmium, Nickel and Benzo(a)pyrene measurements, limit values and assessment methods. According to this regulation, hourly SO_2 amounts in Turkey should be less than 900 $\mu g/m^3$ (URL 6).

2.3.2.1. Sulphur Dioxide (SO_2)

SO_2 is a member of high reactive group of gases which are named as Sulphur oxides. 66% of SO_2 emissions are originating from fossil fuel combustion in power plants and 29% of these emissions are coming from other industrial processes, for example, metal extracting, and burning of substances including high sulphur content. The bonds in the structure of SO_2 molecule are covalent and it is a planar molecule with 120° angel and v-shaped. In Figure 2.1 molecular structure of SO_2 is seen (Hussain 2011).

Figure 2.1 Molecular Structure of SO_2 (Hussain 2011)

SO_2 gas is colourless and two times heavier than air. It has suffocating and pungent characteristics when it is burned. The melting point -72.7 ^oC, boiling point -10 ^oC, and relative density of SO_2 gas is 1.43. SO_2 gas also has high solubility in water at 0 ^oC 1 L of water can dissolve 80 dm^3 of SO_2 gas (Hussain 2011).

In order to form SO_2 gas, burning reaction of sulphur should occur shown in below.

$$S + O_2 \rightarrow SO_2 \qquad \text{(Equation 2.1)}$$

It can also be produced in laboratory by heating of copper pieces with sulphuric acid or reaction of sulphuric acid with concentrated hydrogen sulphide (Hussain 2011).

When it is in moisture environment, it makes reaction with water as a reducing agent and produce sulphurous acid (H_2SO_3) which is reduced in two ways such as addition of nascent hydrogen and removal of oxygen atom. Final molecule produced from these reduction reactions of sulphurous acid is sulphuric acid (H_2SO_4) (Hussain 2011).

The reactions of SO_2 with atmosphere are very complex and many different aspects like wind speed, cloudiness, rainfall and temperature affect this relationship (Dursun 1994).

2.1.3.3. Nitrogen-containing compounds

Nitrogenous compounds include gases like nitric oxide (NO), nitrogen dioxide (NO_2), nitrous oxide (N_2O), ammonia (NH_3) and salts like nitrate (NO_3) and ammonium (NH_4) (Kırımhan 2006). Nitrous oxide is colourless compound and emitted from natural sources. It is made up of bacteriological activity and reaction of nitrogen (N_2) with oxygen and ozone. Nitric oxide (NO) is formed by both natural events and anthropogenic activities. Especially combustion of fossil fuels huge amount of nitric oxide is emitted and nitrogen dioxide (NO_2) is emitted with nitric oxide because when NO contacts with oxygen in the air, it produces NO_2. These two gases are very important pollutants (Kırımhan 2006).

Furthermore, chemical alteration of these gases in the atmosphere produces nitrate and ammonia salts which return to the earth surface by wet and dry deposition. NOx species cause the formation of acid rains (Kırımhan 2006).

2.1.3.4. Other Pollutants

Other pollutants that could be considered as important can be listed as carbon monoxide (CO) which is colourless, odourless, non-irritating but very poisonous gas, carbon dioxide emitted from combustion, volatile organic compounds (VOC) which are vapours and solvents of hydrocarbon fuels, chlorofluorocarbons (CFCs) emitted from products currently banned from used, radioactive pollutants produced by nuclear explosions, war explosives, and natural processes such as the radioactive decaying of radon, and odours from garbage, sewage, and industrial processes (Vesilind 2010).

2.1.4. Effects of Air Pollutants

Air pollutants have adverse effects on human health and welfare, vegetation and animals, materials and structure, and atmosphere, soil and water bodies (Boubel et.al 1994) The information related to health effects of air pollutants come from three different research paths named as epidemiological, clinical and toxicological. Mostly these effects occur in respiratory, circulatory, and olfactory systems. Air pollutants enter to the human body directly by respiratory systems and cause the alteration in the function of lungs. The main air pollutants and their effects on human health can be sort as CO leads to problems in circulation of oxygen in the body, NO_2 makes respiratory system more susceptible to pathogens, O_3 causes asthma attacks, SO_2 and particulate matters rise the acute respiratory disease problems (Boubel et.al 1994). The detailed information about effects of SO_2 in human health is given in next sections.

Moreover, air pollutants affect vegetation in two ways which are visible and nonvisible symptoms seen on the plants. Visible ones include collapse of tissue and colour loss while nonvisible ones cause reduction of plant growth and physiological and biochemical process alteration. Animals are affected from air pollution with respect to the acidification process in water bodies or heavy metal deposition in various environments in which animals are lived (Boubel et.al 1994).

Oher significant effect of air pollutants is on the materials and structures. Especially stone, metals, fabrics and dye, leather, paper, pain and glass, rubber are main materials affected from air pollutants. Air pollution increases the rate of corrosion on metals, causes deformation on the stone parts of buildings because of acid gases, causes loss in tensile strength on fabrics, and decomposes composition of the leather by SO_2 absorption (Boubel et.al 1994).

Also, the effects of air pollutants to atmosphere, soil and water bodies have great importance. The visibility problems and acid deposition because of the formation of H_2SO_4 and HNO_3 by the contact of SO_2 and NO_3 gases with water bodies are the main problems could be seen in these environments (Boubel et.al 1994).

2.1.4.1. Effects of SO_2

Effects of SO_2 depend on gas amount emitted by volcano eruption, atmospheric layer in which the gas is found, wind and weather conditions affecting transportation of the gas. Mostly known effects of SO_2 are it may give harm to tissues and mucous membranes of the eyes, nose and throat because of the irritating odour and the upper part of respiratory system and bronchi are seriously influenced by SO_2 gas. In addition to these main effects, SO_2 has negative impacts on human health, vegetation and animals, and environment (Hussain 2011).

2.1.4.1.1. Effects of SO_2 on Human Health

Health effects of SO_2 gas is changing according to the dose and the exposure time. If exposure time is short and dose is high, there could be problem threatening life for people who already have lung disease. If exposure time is long and dose is high, diseases such as chronic bronchitis, emphysema, and respiratory illness could be formed (Hussain 2011). Moreover, sulphur dioxide gas may react with chemicals and particulates in the air and produce sulphate particles that cause respiratory illness and breathing problems by accumulating in the lungs (Hussain 2011). Fertility decreases in men and women for long term exposure of industrial SO_2 gas. People who have diseases like asthma, heart or lung problems are more sensitive to SO_2 gas than other people. Whether there is a relationship between cancer and SO_2 gas but some studies done in animals show that SO_2 gas may produce cancer on mice but according to classification of International Agency for Research Cancer (IARC), SO_2 gas has not capable to make cancer on humans (Hussain 2011).

2.1.4.1.2. Effects of SO_2 on Vegetation and Animals

Effects of SO_2 on leaves mainly occur with penetration into stomata and injury depends on uptake rate of SO_2. Also, severity of the injury depends on resistance of the leaves. SO_2 with high concentrations and short exposure can create acute problems in the plants while the long term exposure of low concentrations can be much more problematic. As a result of that kind of exposure, chronic injuries occur in the plants such as decrease in growth and yield, increase in senescence, and colour problems (Hussain 2011).

Moreover, lichens and bryophytes are affected from SO_2 gas largely while they are in metabolically active and moist condition. Decrease in nitrogen fixation ability, increase in membrane leakiness, destruction in photosynthetic pigment and activities are the main effects of SO_2 gas on lichens and bryophytes (Hussain 2011).

Furthermore, acid mists another important problem occurs on the plants, especially on forests. SO_2 gas is one of the main pollutants that are capable of producing acid mist causing serious injuries on the leaf surface and plant growth (Hussain 2011).

Long term exposure of SO_2 and its' effects on animals have searched only on the experimental animals. Concentrations higher than 10 ppm with long term exposure may produce harm on the epithelial area of animals. Besides, bronchoconstriction and asthma problems can be developed by animals according to the exposure time and concentration (Hussain 2011).

2.1.4.1.3 Environmental Effects of SO_2

Acid rain produced by SO_2 gas contribution gives harm to plants and buildings, causes acidification problems on lakes, rivers and streams and damages to the aquatic life. When the exposure from this gas become continuous, ecosystem change could be inevitable for plants and animals (Hussain 2011). Volcano eruption is one of the most important contributor of SO_2 pollution. It could emit SO_2 to the atmosphere smaller ranging smaller than 10 ton/day to higher than 20 million ton/day (Hussain 2011).

SO_2 released from volcano eruption in huge amounts is easily oxidized to sulphuric acid. In this reaction primary oxidation agent is hydroxyl radical (OH) which is formed in higher layers of the atmosphere and ozone (O_3) is used for its' formation. If SO_2 and other pollutants which have greenhouse effect are emitted in large amounts, OH and ozone depletion occur and global warming problem is constituted (Hussain 2011).

Also, hydrolysis reaction occurs between SO_2 and H_2O which produce sulphuric acid and causes ozone depletion in oxidation reaction with ozone, hydrogen peroxide and oxygen (Hussain 2011).

2.2. Artificial Intelligence

According to Ginsberg (1993) simple definition, “Artificial Intelligence is the enterprise of constructing an intelligent artifact.” In this definition artificat means physical symbol system consisting of set of symbols that have different physical characteristic by Ginsberg (1993). The main concern of AI is intelligent behaviour which includes communication, perception, and learning, reasoning and behaving in problematic environment. With AI these behaviours can be solved by machines as good as human does (Nilsson 1998). Moreover, AI can understand these behaviours in wherever it occurs like in machines, in humans or in animals. Both scientific and engineering aims are important in AI perspective (Nilsson 1998).

2.2.1. Artificial Intelligence and Engineering

Although starting point of artificial intelligence is computer sciences, there have been developed several practical tool in order to solve engineering problems which are severely complicated for human mind (Pham and Pham 1999). These tools are fuzzy logic, knowledge based systems, inductive learning, neural networks, and genetic algorithms (Pham and Pham 1999).

Knowledge based systems includes domain related to that the knowledge is interlarded in a computer program using IF-THEN rules. In this method problems are solved by inference mechanism making changes in the loaded knowledge. Applications of this method are broad such as signal interpreting, fault diagnosis, machine design, selection of materials and machine and process control (Pham and Pham 1999).

Fuzzy logic is the other important tool of artificial intelligence used in engineering problem solving. It has one disadvantage which is not capable of handling new situations not found in the knowledge bases of the system. However, it shows better performance when it is used for the problems having a qualitative and inexact bases. In these systems variable’s precise value is displaced with description. “For instance, ‘20°C’ can be replaced by ‘normal’ as the linguistic description of the variable ‘room temperature’.” (Pham and Pham 1999) Fuzzy logic uses same IF-THEN rules used in knowledge based system.

Inductive learning includes set of examples to produce a result. In this method the example sets are categorized as attributes and a value is attained to each attribute. After value attenuation step, process called dividing and conquering in which special strategy is used for selection of an attribute in order to achieve division of example set to subsets is applied. Inductive learning produce decision tree which include knowledge gained by specific examples to classify the example set correctly. Thus, inductive learning may be very useful for handling some problems not given in the example set (Pham and Pham 1999).

Neural networks have computational features of human brain. Models developed with neural networks make computation by using neurons which simple interconnected units are working in parallel structure (Pham and Pham 1999). Taxonomy of neural networks is given in Figure 2.2 (Gardner and Dorling 1998). The multilayer perceptron is the mostly known and used neural network. Perceptron modelling mainly uses feedforward network where inputs and outputs are connected with each other in a single line of signals. Precise knowledge of neural network is provided by training process. Weights of the connections between neurons are modified by using error among actual and expected output values. The back propagation algorithm used in multilayer perceptron determines the weight modifications in the hidden layer of network and spread the error from output vectors. There are many application of neural networks used for engineering problem solving such as prediction, classification and pattern recognition, clustering, optimisation and signal processing (Pham and Pham 1999). Detailed information related to artificial neural networks is given in next section.

Figure 2.2 Taxonomy of neural networks (Gardner and Dorling 1998)

Genetic algorithms are the last mostly used artificial intelligence technique. Natural evolution is the inspiration point of genetic algorithm. The most important feature of genetic algorithm is solving the complex problem without requiring any specific information related to issued problem. Nevertheless, strings of numbers must be used for representative of given problem so that algorithm is applicable to give potential solutions. These strings of numbers known as chromosomes must show the goodness or badness of each (Pham and Pham 1999).

2.2.1.1. Artificial Neural Networks

In order to model complex problems related to many aspects, artificial neural networks (ANN) which is computational modelling have appeared. ANN models have interconnected structures that produce parallel computations during data processing. The main idea behind the ANN modelling is using complex problem solving ability of biological networks (Rozlach 2015). The remarkable characteristics of biological systems such as high parallelism, fault and failure tolerance, nonlinearity, leaning, handling ability of fuzzy information, and generalization ability are significant features of ANN models (Basheer and Hajmeer 2000).

There are three ways which uses ANN models in below:

- Used as biological nervous system models
- Used as controller in robotic systems or signal processing tool
- Used as analytical data methods (Sarle 1994)

In this thesis artificial neural network was used as analytical data method. Detailed of this process will be discussed in next sections.

2.2.1.1.1. Fundamentals of ANN

In order to understand artificial neural networks clearly, fundamental information related this subject is very useful. Firstly, understanding of biological neuron model is necessary for solving the connection logic of ANN. Then, advantages and possible

problem solving ability of ANN have been investigated. Finally, classification of ANN has been done and multilayer neural networks have been studied in this section.

2.2.1.1.1.1. ANN and Biological Neuron Model

Biological neuron in human brain is the important part of the nervous system so investigating the operation of biological neuron is useful to understand ANN. In human brain there are millions of neurons in different features and types (Basheer and Hajmeer 2000).

Figure 2.3 a) Biological neuron, b) Mechanism of signal transfer between two biological neurons (Basheer and Hajmeer 2000)

In Figure 2.3 schematic representation of biological neuron is given in (a) and signal transfer mechanism between two biological neurons is given in (b). Main units of neuron are dendrites, axon, and cell body. Dendrites take signals from other neurons and transmit them to the cell body. Cell body contains information necessary for future uses and provides equipment for the neuron. Axon receives signals coming from cell body and transports them to the synapse (Basheer and Hajmeer 2000).

There is a significant similarity between biological neuron and artificial neuron that connection between the nodes in artificial neuron represents the dendrites and axons and the weights of connections are similar analogy with synapses (Basheer and Hajmeer 2000).

2.2.1.1.1.2. Advantages of ANN

Artificial neural network is one of the mostly used modelling techniques nowadays. It has many advantages when it's compared with traditional models because these models need known input data without any assumption (Gardner and Dorling 1998).With the help of suitable connecting weights and transfer functions, approximation of a multilayer function can be measurable function between input and output (Hornik et al. 1989, Gardner and Dorling 1998). Owing to its capability of using many variables and making connection between different variables, its' importance is getting increase.

Most of the parallel structured ANN models include intense interconnected adaptive units. The ANN models have very significant characteristic that is adaptivity of the network of highly nonlinear problems (Zurada 1997). In order to train any measurable function, the ANN modelling technique may be used properly (Simpson 1989, Hornik et al. 1989).

Moreover, ANN can train large amount of datasets, use parallel structure so that the proceeding time is decreasing drastically, and has capability of capturing nonlinear and complex underlying characteristics of any physical process (Basheer and Hajmeer 2000). ANN models show better results according to prediction and data fitting abilities than the classical statistical modelling methods especially for the data showing nonlinearity (Bourquin, Schmidli et al. 1998).

2.2.1.1.1.3. Problem solving capability of ANN

There are five main problems that can be observed during model development and ANN is capable of solving these problems efficiently. The problems solved by ANN are given in the Figure 2.4.

Figure 2.4 Solved problems by ANN a) pattern classification, b) clustring, c)function approximation, d)forecasting, e)imege completion (Basheer and Hajmeer 2000)

Forecasting ability of ANN has been used in this thesis. Training of the given parameters provides the prediction of future data by using the given scenarios. Network predicts the target data by using historical datasets (Basheer and Hajmeer 2000).

2.2.1.1.1.4. Classification of ANN

According to the features of ANN, classification is made in many different ways. This classification is made with respect to consider the function related to modelling purpose, connectivity degree of the neurons, flow direction of the information, type of the learning algorithm, and learning rule required for training step of ANN modelling (Basheer and Hajmeer 2000). The mostly known ANN models are hopfield networks,

adaptive resonance theory networks, kohonen networks, backpropagation networks, counter propagation networks, and radial basis function networks. (Basheer and Hajmeer 2000)

Backpropagation ANN

Back propagation (BP) feedforward algorithm is the most famous learning algorithm used in ANN models. Networks in which BP is used for learning algorithm called backpropagation networks. Error is searched by back propagation using gradient descent points. Iterations in BP include forward step in which solution is provided and backward step in which error is computed for modifying weights (Basheer and Hajmeer 2000).

Back propagation algorithm is used for training the multilayer neural networks (Haykin 2001). The speed up or slow down of the error convergence are associated with the 'η' and 'μ' symbols respectively in back propagation training (Rumelhart et al. 1986). Moreover, the gradient descent method is used to compute the back-propagation algorithm which approximates to the trajectory in weight space (Battiti 1992). 0 and 1 or -1 and 1 is used to identify the 'η' and 'μ' (Rege and Tock 1996, Rumelhart and McClelland 1995). There are several guidelines for 'η' and 'μ' to determine their optimal values (Haykin 2001).

2.2.1.1.1.5. Multilayer Neural Networks

Multilayer neural networks include input layer, hidden layer and final layer consisting of neurons. Every connection in neural network is related to a weight that is a numeric number (Wang 2003). Output formula of neural network is given in the below;

$$h_i = \sigma\left(\sum_{j=1}^{N} V_{ij}\, x_j + T_i^{hid}\right) \qquad \text{(Equation 2.2)}$$

h_i: the output of i neuron in the hidden layer

$\sigma(\,)$: the activation function

N: the number of input neurons

V_{ij} : the weights

x_j: inputs to the input neurons

T_i^{hid}: threshold terms of hidden neurons

The activation function in this formula is used to bind the values of neurons in order to prevent collapse of the model because of divergent neurons. Moreover, it promotes the nonlinearity in the network model. One of the mostly used examples of activation factions is logistic function. The other probable ones are tangent and hyperbolic tangent that give the same response with logistic function but different output (Wang 2003).

Moreover, the number of input neurons does not depend on the function while the output neurons depend on the activation function approximated by neural network. Also, the input and output variables in activation function can be represented with binary terms such yes or no. Also, they can be represented with symbols such as red or blue if the encoding of the data is appropriate. With this feature the feasibility of neural network increases (Wang 2003). In Figure 2.5 achitecture of simple multilayer neural network is given.

Figure 2.5 Architecture of multilayer neural network (Dehkordi 2012)

The input variables and the input layers are equal to each other with respect to number. The output layer includes only one neuron. In addition, the input/output neurons,

training patterns, network architecture, activation function type and training algorithm are affected by the neurons in hidden layer (Alsugair and Al-Qudrah 1998, Sarle 1997). In order to approximate any nonlinear function, one hidden layer is enough (Hornik et al. 1989). To determine the neuron number in the hidden layer is provided by training different networks and approximating the errors on the test data. If the neurons are fewer, the hidden layer gives results with high training and testing errors because of the statistical bias and under-fitting. On the other hand, if the neurons are greater, the errors in training procedure decreases while the ones in testing procedure increases because of the high variance and over fitting (Sarle 1997). In much study using ANN modelling the iterative approach is used to find hidden layer neurons.

In order to find the non-linear relationship between input and output variables, the activation function is needed. This function connects the corresponding parameters (Sarle 1997). To determine the appropriate activation function, the previous works should be examined and according to the environmental problem working on it, best resulting one should be selected. There are several activation functions such as linear, sigmoidal, hyperbolic tangent, squashing, and linear threshold in that mostly use done is logistic (sigmoid) function.

The initialization of weights and bias (free parameters) of neural network is necessary before training process is starting. The starting values of weights and free parameters contribute to the training process being fast. In order to initialize the free parameters and weight, the random numbers can be used and distributed within known range. It is suggested that the distribution range of free parameters should be smaller so that occurrence of the errors become less in the network (Wasserman 1989).

The mostly used training method for neural networks is the back-propagation learning algorithm. The input and output data supplied with an ordered form in this learning algorithm (Rumelhart and McClelland 1995). The weighting of network helps the model indicating the important parameters in the training data. This is provided with minimizing the errors of input and output data (Gardner and Dorling 1998). Moreover, the under and over training makes algorithm fail and produces high prediction errors (Gardner and Dorling 1998, 1999, Comrie 1997).

Thus, while deciding the number of training iterations, over and under training should be avoided. The data is divided into three parts named as training data set, test data set, and evaluation data set in BP algorithm. Majority of the data is used in the training data set for training purpose. The other data is used in the testing data set in which trained model performance is checked. When the model error is minimized on test data set, the training step is stopped. Lastly, predicted model is validated by evaluation data set (Gardner and Dorling 1998).

2.3. Previous studies using ANN for air pollution modelling

When the recent past is investigated, there are so many studies have been developed using artificial neural networks related to branches of environmental engineering. In these studies air quality and pollution is one of the mostly dwelled on subjects. Prediction of particulate matter, modelling of volatile organic compounds, predictions of nitrous oxides, methane, sulphur oxides are the mostly studied subjects.

First air modelling related to air quality was developed by Boznar in 1993. He and his team predicted SO_2 concentration from stack gas of power plant for certain distance in complex topography. In this study as an input parameter SO_2 concentrations measured in different locations in power plant and wind direction and speed, temperature and time was used. Results of this model was very good despite the complex topography of studying area (Boznar, Lesjak et al. 1993).

In 1994 studies related to models which can be used in air quality modelling and neural networks were made. Satisfactory results of the applicability of these models in air quality modelling in future were taken (Mlakar, Božnar et al. 1994).

Daily ozone concentration prediction was made using data of 8 cities of United States and compared the results of statistical regression model and artificial neural network (Comrie 1997).

Effects of seasonal climate change to atmospheric dissolved organic carbon and nitrogen amounts were modelled with artificial neural networks by using data gathered from 14 basins in North America and Canada (Clair and Ehrman 1998).

Multilayer perceptron modelling was applied to air pollution modelling and the features of this method were evaluated in 1998. Also, the strengths and advantages of ANN over other statistical modelling techniques in many scientific fields were investigated (Gardner and Dorling 1998).

Atmospheric hourly NO_x and NO_2 concentration was evaluated by using multilayer perceptron modelling in London. Results showed that there is a strong relationship between NO_x and NO_2 concentrations and meteorological factors and this relationship can modelled with artificial neural network very efficiently (Gardner and Dorling 1999).

For Delhi city in India SO_2 concentrations modelled with artificial neural networks including wind speed, temperature, relative humidity, and the wind direction index as input parameters and good results were taken (Chelani et.al 2001).

Adaptive neuro-fuzzy inference system (ANFIS) was applied in order to predict effects of meteorological factors on SO_2 concentrations for Erzurum city. Developed model had performance greater than 60 – 70% of prediction ability (Çelebi 2002).

In study done in Zonguldak city particulate matter (PM) and SO_2 concentrations were predicted for two measurement stations by using ANN. During model development meteorological parameters are used as ınput parameters along with PM and SO_2 data. As a result of the study, model developed with ANN gave successful results (Tecer 2007).

Göztepe station was selected as measurement area for measuring SO_2 concentration and meteorological data in İstanbul. Previous days SO_2 concentration averages, daily temperature, relative humidity, cloudiness, average daily wind speed and dominant direction were indicated to the model as input parameters and next days' SO_2 concentration was tried to be estimated by ANN. Performance of model was higher than 70% (Akkoyunlu et.al 2010).

ANN models were developed which predicts SO_2, CO, CO_2, NO_X, O_2 concentrations emitted from stacks of 16 brick and tile factories in Çorum city. In these models vent diameter, cross sectional area, fuel calorie, fuel amount, thermal power, stack height,

ambient temperature, gas temperature, average humidity, gas speed, gas flow, and dust concentrations of brick and tile factories were used as input parameters for prediction of SO_2, CO, CO_2, NO_X, O_2 concentrations. With respect to these parameters, experimental studies were conducted in three different period of a year and data was gathered for model development. Some part of gathered data was used for training purpose and rest of them was used for testing of the model. The model predicted SO_2 concentrations gave R values 0.9915 and 0.9807 respectively for training and testing (Arabacı et. al 2010).

ANN model was developed for Malesia by considering SO_2, CO, NO_2, NO, temperature, relative humidity, and air velocity parameters. Performance of developed model was evaluated according to mean square error (MSE) and R^2 values. Best prediction of this model was 0.57 R^2 and 0.062 MSE (Ghazali 2012).

Moreover, SO_2 concentrations in center of Konya city were modelled with indication meteorological parameters by ANN and adaptive neuro-fuzzy inference system (ANFIS). In this study ANFIS was found more successful than ANN model but both could be used for prediction of air quality (Dursun, Kunt et al. 2015).

3. MATERIAL AND METHOD

3.1. Working Area

Working area of this study is Seydişehir town which is in south part of Konya City. It is the biggest city in Turkey with surface area of 40,814 km^2. This amount covers 5% of total surface area of Turkey (URL 8). Surface area of Seydişehir town is 2207 km^2 and it is surrounded with mountains. Average altitude of the town is 1113 m. Population of Seydişehir is 63773 at 2014. Half of the population lives in town center and rest lives in villages. In Figure 3.1 the map of Konya city and Seydişehir town is given.

Figure 3.1 Map of Konya and Seydişehir town (URL 9)

Seydişehir Eti Aluminium Facility is one of the biggest industrial factories of Konya. Bauxite and lignite are the main minerals extracted from this town and they are processed in Eti Aluminium Facility (URL 10).

3.1.1. Climate of working area

Climate of Seydişehir is warm and mild. Rainfall amount in winter months are higher than summer months. Yearly average temperature of Seydişehir is 11.7 ^{o}C and average rainfall amount of town is 750 mm. Rainfall difference between most arid and rainy month is 126 mm and average temperature changes around 21.8 degrees. (URL 11) Figure 3.2 shows the change in temperature and rainfall during one year period in Seydişehir town. According to this figure, high temperatures and minimum rainfall are seen in July month of year.

Figure 3.2 Change in temperature and rainfall during one year period in Seydişehir town (URL 11)

According to study done by Sarı and İnan (2011), climates of Seydişehir and Beyşehir towns were compared. Information related to climatic factors of Seydişehir town is given next sections by making use of this study. For developing ANN model, meteorological factors were integrated to the model as input parameters.

3.1.1.1. Temperature

Yearly average temperature of Seydişehir town is 11.8 °C and minimum average temperature happens on January, maximum average temperature happens on July months. The average temperature during spring is lower than fall season because winter and spring passes rainy, surface is covered with green plants, cloudiness increases in spring months while arid summer season increases fall temperature. Around Seydişehir, altitude increases between short distances so 5 – 6 °C temperature difference is seen in between bottom land and mountain areas. In Table 3.1 temperatures of seasons in Seydişehir town is given (Sarı and İnan 2011).

Table 3.1 Temperatures of seasons in Seydişehir town (°C) (Sarı and İnan 2011)

Winter	Summer	Spring	Fall	Average
1.1	22.4	10.7	12.9	11.8

In winter season inner parts of Turkey enters the effects of cold air masses coming from Siberia, therefore, very low temperatures can be seen in the town. For example, on February 1991 absolute minimum temperature of Seydişehir was measured as -27 °C. That kind of low temperatures may be harmful for agricultural activities, transportation, power transfer lines, and water supply networks. Absolute maximum temperatures reach to 39.2 °C in Seydişehir town (Sarı and İnan 2011).

3.1.1.2. Atmospheric Pressure

There is a sharp increase seen in the pressure values of fall season. Yearly average pressure of Seydişehir town is 887.3 mb. Monthly pressure between September and December is higher than yearly average pressure. Table 3.2 gives the distribution of yearly average pressure of Seydişehir (Sarı and İnan 2011).

Table 3.2 Distribution of yearly average pressure of Seydişehir (mb) (Sarı and İnan 2011)

Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.	Avg.
885.5	886.9	886.2	885.7	887	886.8	885.6	886.4	887.9	891	890.4	889.1	887.3

3.1.1.3. Wind

Average wind speed of Seydişehir town is 2.1 m/sec and average wind speeds of spring season in measurement stations is higher. Seydişehir town and closed settlements are affected from strong winds (Sarı and İnan 2011).

Figure 3.3 Dominant wind direction and frequencies of Seydişehir town (Sarı and İnan 2011)

Direction and frequency of wind is very important with respect to many activities. Especially for determining construction area of factories and city planning wind direction and frequency should be considered. Dominant wind direction and frequencies are determined according to Rubinstein formula. There are two dominant wind directions which are shown in Figure 3.3 for Seydişehir town. First one is N to 62.9 W with 40.7% frequency and second one from S to 24.3 E with 32.5% frequency (Sarı and İnan 2011).

3.1.1.4. Relative Humidity

Yearly average relative humidity of Seydişehir is 64.4%. Relative humidity of town is high in winter season because humid air masses coming from Mediterranean region enters easily to Seydişehir and its surroundings. Moreover, relative humidity decreases in summer season since frequency of humid air masses is low in this season. Table 3.3 represents distribution of yearly average relative humidity of Seydişehir (Sarı and İnan 2011).

Table 3.3 Distiribution of yearly average relative humidity of Seydişehir (%) (Sarı and İnan 2011)

Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.	Avg.
79.2	75.7	68.7	62.9	60.9	55	49.1	49.6	54.5	64.7	72.8	79.3	64.4

3.1.1.5. Rainfall

Yearly average rainfall amount of Seydişehir is 750.3 mm which is greater than Turkey's yearly average rainfall amount. Number of the rainfall events in winter season is higher than summer season for Seydişehir town (Sarı and İnan 2011).

In Seydişehir the highest yearly rainfall amount between 1975 and 2006 was 1126.6 mm and the lowest yearly rainfall amount between same years was 474.9 mm. Rainy days starts on the beginning of October and reaches its' maximum level on December. After January it starts to be decreasing regularly. Distribution of yearly rainfall according to the months is 46.9% for winter, 24.4% for spring, 6.2% for summer and 22.5% for fall season. Figure 3.4 shows this distribution (Sarı and İnan 2011).

Figure 3.4 Distribution of yearly rainfall of Seydişehir according to the months (Sarı and İnan 2011)

3.1.1.6. Cloudiness

Cloudiness reaches its' peak values on April and May months in Seydişehir. Increase in convectional movements and mobile depressions affect this increase. The number of cloudy day on May is 19.7. August is the month on which cloudiness is the lowest level around 8 days. On winter months cloudiness is in high degree. Table 3.4 shows the distribution of average cloudy day number in Seydişehir according to months (Sarı and İnan 2011).

Table 3.4 Distribution of average cloudy day number in Seydişehir according to months (day) (Sarı and İnan 2011)

Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.	Avg.
16.6	16.2	16.7	18.6	19.7	13.8	10	8.2	10.3	13.7	15.4	16.3	14.6

3.1.2. Geographical situation of working area

Seydişehir town is in the central Anatolia region and ancillary of Konya city. It is 107 km far from Konya city center and in 37.7 north latitude, 31.57 degrees eastern longitude. Town is circulated by Beyşehir from north, by Yalıhöyük, Ahırlı, Akseki provinces and Antalya city from south, by Derebucak from west, by Akören from east.

Town which locates at Suğla plain that is a fertile valley extended with Çarşamba stream at north foot of Toros Mountains is surrounded by Küpe Mountain from west to south and by Gidengelmez Mountain from south which has very rich bauxite mineral deposits. On the other hand, Küpe Mountain has many natural water reserves. These huge mountains around Seydişehir create a different aesthetic silhouette. Besides, Pınarbaşı, Kuğulu, Beldibi ponds used for sporting and entertaining activities are natural beauties which are taking place in the city and fed from spring at Küpe Mountain. There is Beyşehir Lake at northwest and Suğla Lake at southeast part of the town.

In Seydişehir town there are 136,256 ha usable area encompassing 3.34% of Konya's usable area and 9.84% of lakes basin. This usable area of the town includes 45.64% of forest area, 29.6% of agricultural area, and 3.52% of forage-grass area. Forest area of Seydişehir generates 11.51% of Konya's forest area and 12.74% of lakes basin. Agricultural area of town comprises 9.81% of agricultural area of lakes basin while it comprises 11.79% of Konya's agricultural area.

Seydişehir has connection with every part of Turkey by highway of Ankara-Konya-Antalya. Town is 85 km far from Konya, 208 km from Antalya, 135 km from Manavgat. Also, there is 50 km to Bozkır, 37 km to Ahırlı, 31 km to Yalıhöyük, 36 km to Beyşehir from the town.

There is 217 km long village road network under responsibility of General Directorate of Rural Services. 31 village roads which are 156 km part of total roads are made by asphalt and 17 km of other roads used by neighbourhoods and pedestrian are stabilized, 38 km of them are levelling, 6 km of them are raw. Also, there is 135 km road network in all Seydişehir under responsibility of General Directorate of Highways (URL 1).

3.1.3. Air pollution sources of working area

Industry of Seydişehir town is very wide in variety of fields such as mining and manufacturing industry. Especially mining in this area has significant importance because of the minerals, bauxite, chromium and lignite, extracting from the area. Both extraction and manufacturing steps of minerals cause air pollution. Moreover, other manufacturing industries like chemical production, agricultural material production, and food production emits some air pollutants to the atmosphere (URL 1). Also, fossil fuel combustion from residential buildings and other institutions for heating purposes largely contributes to SO_2 pollution in Seydişehir.

Eti Aluminium Facility is the biggest metal production and the only aluminium manufacturing facility of Turkey. It was constructed in closed area of 12,000,000 m² (URL 2). During aluminium production Sulphur in the anodes is oxidized, releasing SO_2 from the potlines which is a row of electrolytic cells used in the production of aluminium as the anodes are consumed. SO_2 is also released from the anode bake furnace as pitch used to help form the anodes are oxidized during the baking process (URL 3). As a result of this anode process during aluminium production, Eti Aluminium facility contributes SO_2 pollution in Seydişehir town.

Furthermore, air pollution is transported long distances along with air movements so air pollution in Seydişehir can be increased by the effects of that kind of transportation mechanism.

3.2. Collection of data

Pollution data of thesis was gathered by SO_2 analyser device which was positioned in Seydişehir town between 2012 and 2013 years. Detailed information related to this device is given in the next section.

3.2.1. SO_2 gas analyser device

M100E UV Fluorescence SO_2 Analyser (Figure 3.5) was used for SO_2 measurement in this area. The Model 100E uses the proven UV fluorescence principle, coupled with

state of the art microprocessor technology to provide accurate and dependable measurements of low level SO_2. Measurement range of the device is between 0-50 ppb to 0-20 ppm (M100E Operation Manual).

Analyser device is controlled by microprocessor which indicates SO_2 concentration passing through the device. “It requires that sample and calibration gases be supplied at ambient atmospheric pressure in order to establish a constant gas flow through the sample chamber where the sample gas is exposed to ultraviolet light causing the SO_2 become excited (SO2*).” (M100E Operation Manual) As these SO_2* molecules decay into SO_2 they fluoresce. Instrument measures the amount of fluorescence to determine the amount of SO_2 present in the sample gas. In Figure 3.6 internal gas flow diagram of M100E UV Fluorescence SO_2 Analyser is given.

Calibration of the device is supplied with software program and there is no need to manual operation for this device. The output signal of sensor is measured by microprocessor during SO_2 gas passing through to device with known amounts and these measurements are stored in the memory of microprocessor. With other performance parameters such as ratio of UV lamps, stray light amount and PMT dart offset, and temperature and pressure values of gas sample microprocessor utilizes stored SO_2 calibration values in order to calculate final concentration of SO_2 (M100E Operation Manual).

(a)

(b)

(c)

Figure 3.5 Photo of M100E UV Fluorescence SO_2 Anayser (a) front view (b) inside of the device (c) back view

Figure 3.6 M100E internal gas glow diagram (M100E Operation Manual)

3.2.1.1 Measurement Principle

Method used for measurement with M100E Analyser includes physical process between ultraviolet light and SO_2 within the wavelengths of 190 nm to 230 nm occurred in the fluorescence. There are two steps for happening of this reaction (M100E Operation Manual).

At first step SO_2 molecules are captured by photons of suitable ultraviolet wavelength. M100E device limits the wavelength of the light around 214 nm because there is a band pass filter between affected gas and UV light source. SO_2 molecules captured by UV light absorb energy so affected molecules give one electron to higher energy orbital state. (Equation 3.1)

$$SO_2 + hv_{214nm} \rightarrow SO_2^* \qquad \text{(Equation 3.1)}$$

Average intensity of UV light (I_a) affects the SO_2 amount which is converted to SO_2* in chamber of sample. The peak intensity of UV light is not important here because this, intensity does not distributed constantly in each part of chamber. During the travelling of light through the gas sample, SO_2 gas absorbs some photons. The average intensity of light is defined in below equation. (Equation 3.2)

$$I_a = I_0 [1 - \exp(- ax(SO_2))] \qquad \text{(Equation 3.2)}$$

Where:

I_0 : Excitation UV light intensity

a : SO_2 absorption coefficient

SO_2 : SO_2 concentration in sample chamber

x : Distance between UV light and affected SO_2 molecules

At second step reaction shown below happens. (Equation 3.3) SO_2^* molecule returns to its' ground state by giving the excess energy in the photon (hv) form since the system wants to pass the lowest stable energy state. This light has longer wavelength centered at 330 nm.

$$SO_2^* \rightarrow SO_2 + hv_{330\ nm} \qquad \text{(Equation 3.3)}$$

Detectable UV amount taken out by SO_2^* decay reaction is affected the rate at which this reaction occurs. (Equation 3.4)

$$F = k\,(SO_2^*) \qquad \text{(Equation 3.4)}$$

Where:

F : Fluorescent light amount taken out by SO_2^* decay reaction

k : The rate of decay reaction

SO_2^* : Excited SO_2 amount in the chamber

Therefore,

$$k(SO_2^*) \xrightarrow{F} SO_2 + hv_{330\ nm} \qquad \text{(Equation 3.5)}$$

M 100E UV Fluorescence SO_2 Analyser's main features are given in below;

- Making the light path (x) very short,
- Removing the effects of lamp drift by measuring the excitation UV light intensity with reference detector,
- Making the rate of decay (k) constant by measuring and controlling the temperature of the sample gas with heaters attached to wall of chamber,
- Removing the common interfering gases from sample by hydrocarbon scrubber,
- Reducing the effects of stray light by optical geometry and spectral filtering of chamber (M100E Operation Manual).

3.2.2. Provision of meteorological data

Meteorological data was provided by General Directorate of Meteorology. Data including average hourly temperature, average hourly pressure, hourly relative humidity, hourly wind speed, hourly rainfall, hourly cloudiness, hourly sun duration, hourly sun radiation values for 2012 and 2013 years was taken. Meteorological measurements of Seydişehir are carried with 17898 no of automatic meteorological observation station (AMOS). Figure 3.7 show typical AMOS.

Figure 3.7 Typical automatic meteorological observation station (URL 12)

3.3. Data analysis

Gathered data was firstly separated two datasets namely, winter and summer. Both datasets include meteorological parameters which are hourly average temperature (^{o}C),

hourly average pressure (bar), hourly relative humidity (%), hourly wind speed (m/sec), hourly rainfall (mm), hourly cloudiness, hourly sun duration, hourly sun radiation (kW/m^2) and previous day's hourly SO_2 pollution concentration (ppb) as input parameters. Next day's SO_2 concentration was tried to be predicted and integrated model as target parameter. Winter dataset has 284 data of each parameter while summer dataset has 130 data of each parameter.

Before using data for modelling, it should be normalized in order to get better prediction performance. There are two main normalization methods such as z-score and max-min which can be used to prepare input data for modelling. In order to understand which one give better results, both of them was applied on winter and summer data separately and 10 trials of feedforward neural network with default settings were done to normalized data in MATLAB (R2011a). Default settings separate data into training, testing, and validation subsets as 70%, 15%, and 15%, respectively. Moreover, 10 neurons in hidden layer are used and as a training function Levenberg-Marquardt is applied in default settings of feed forward neural network. After these trials, average of mean square errors (MSE) was taken and the normalization method that gave lowest MSE was chosen as a best option because MSE is defined as the average squared difference between outputs and targets. Lower values are better. Zero means no error (Ozturk and Dursun, 2016).

In order to normalize input data, SPSS statistic programme was used. First of all, z-scores normalization was applied to input data. The definition of z-scores normalization method is given in equation 3.6.

$$z = \frac{x-\mu}{\sigma} \qquad \text{(Equation 3.6)}$$

Where:

z : z-score of value

μ : Mean

σ : Standard deviation

x : Value

Secondly, max-min normalization was applied to input data. The definition of max-min normalization method is given in equation 3.7.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \qquad \text{(Equation 3.7)}$$

Where:

z_i : normalized value

min(x) : Min. value of xdata

max(x) : Max. value of x data

x_i : Value of ith in x data

As a result of trials, z-scores normalization of input data for winter gave better results than max-min normalization while max-min normalization gave better results for summer input data than z-scores normalization. The summary of the results taken from proposed methods is given in Table 3.5.

Table 3.5 MSE values taken by z-scores and max-min normalizaiton approaches

	Winter (Avg. MSE)	**Summer (Avg. MSE)**
Z-scores normalization	**0.0509801**	0.001189285
Max-min normalization	0.055453	**0.000885296**

3.4. Modelling with artificial neural network

In order to develop ANN model, first of all architecture of model should be determined after data optimization process. Every ANN model consists of three layers namely input layer, hidden layer, and output layer. For this study there are 9 neurons in input layer and 1 neuron in the output layer. Hourly average temperature (^oC), hourly average pressure (bar), hourly relative humidity (%), hourly wind speed (m/sec), hourly rainfall (mm), hourly cloudiness, hourly sun duration, hourly sun radiation (kW/m^2) and previous day's hourly SO_2 pollution concentration (ppb) were integrated to the model as input parameters and SO_2 concentration (ppb) of next day was tried to be predicted.

The neuron number in hidden layer has significant importance for model development because it affects the model accuracy and performance. “The best network topology corresponds to a feedforward neural network which presents a minimum value of MSE for the validation data set.” (Sousa, Martins et al. 2007) As a result, determination of neuron number in hidden layer was done according to (MSE).

Another important step of model development is determination of activation function which connects the corresponding parameters. Mostly used activation functions are linear, sigmoidal, hyperbolic tangent, squashing, and linear threshold. For this model, hyperbolic tangent activation function was chosen for connecting input layer and hidden layer and linear activation function was chosen for connecting hidden layer and output layer. In Figure 3.8 general architecture of developed ANN model is given.

The range of hyperbolic tangent function is [-1, 1]. Its’ MATLAB function in Equation 3.8 and mathematical function in Equation 3.9 are given.

a = tansig (w*p + b) (Equation 3.8)

$$f(x) = \frac{e^{x}+e^{-x}}{e^{x}+e^{-x}} \quad \text{(Equation 3.9)}$$

The range of linear function is [-∞, ∞]. Its’ MATLAB function in Equation 3.10 and mathematical function in Equation 3.11 are given.

a = purelin (w*p + b) (Equation 3.10)

f(x) = x (Equation 3.11)

Moreover, data is separated in three parts namely training, testing and validation sets. Majority of the data is used in the training data set for training purpose. The other data is used in the testing data set in which trained model performance is checked. When the model error is minimized on test data set, the training step is stopped. Lastly, predicted model is validated by validation data set. Thus, training function should be selected and this selection affects the performance of the model. There are many choices of training

function in NN toolbox of MATLAB which uses backpropagation learning algorithm. In this study 70% of data was used in training set, 15% in testing and 15% in validation sets. Training functions were selected with respect to trials of several functions and comparing average R values.

Figure 3.8 General architecture of developed ANN model

3.4.1. Modelling with winter data

Architecture of model for winter data is same as Figure 3.8. Meteorological factors and previous day's SO_2 value was integrated to the model as input parameters and next day's SO_2 concentration was tried to be predicted. 284x9 matrixes developed by winter data as input and 284x1 matrix as target was used in the model. 198 of winter data were used in training set, 43 in testing and 43 in validation sets. Furthermore, hyperbolic tangent activation function was chosen for connecting input layer and hidden layer and linear activation function was chosen for connecting hidden layer and output layer.

Selection of training function was done according to average of overall R values of trials. If R values are closer to 1, it means target and predicted values are compatible with each other. 25 trials were done for each mostly used training functions which are Levenberg-Marquardt backpropagation (trainlm), Resilient backpropagation (trainrp), Scaled conjugate gradient backpropagation (trainscg), Conjugate gradient backpropagation with Polak-Ribiére updates (traincgp), Conjugate gradient backpropagation with Fletcher-Reeves updates (traincgf), Conjugate gradient backpropagation with Powell-Beale restarts (traincgb), and BFGS quasi-Newton backpropagation (trainbfg) for backpropagation learning algorithm . During these trials 10 neurons were used in hidden layer. As a result of these trials Levenberg-Marquardt backpropagation (trainlm) training function gave better results than other functions so it was selected as training function of this model developed with winter data. Average values of 25 trials of each function are given in Table 3.6.

Table 3.6 Average values of 25 trials of each function for normalized winter data

Training function	Criteria	Avg. Values
trainlm	MSE	0.058039
	R	**0.439187**
	R^2	0.197295
trainrp	MSE	0.055584
	R	0.353764
	R^2	0.130307
trainscg	MSE	0.056656
	R	0.328006
	R^2	0.116135
traincgp	MSE	0.056329
	R	0.322654
	R^2	0.113659
traincgf	MSE	0.052654
	R	0.375284
	R^2	0.145863
traincgb	MSE	0.05201
	R	0.365215
	R^2	0.138878
trainbfg	MSE	0.058156
	R	0.365172
	R^2	0.140549

After determination of training function of the model, neuron number in the hidden layer was designated according to average MSEs. By using selected training function (trainlm), neuron numbers 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 was used in hidden layer and 10 trials for each were done. Consequently, 15 neurons in hidden layer gave the best average MSE result. Summary of these trials is given in Table 3.7.

Table 3.7 Average MSE of 10 trials of each neuron number for normalized winter data

Neuron number	Criteria	Avg. Values
5	MSE	0.0731658
10	MSE	0.065927
15	MSE	**0.0586323**
20	MSE	0.0796901
25	MSE	0.1025047
30	MSE	0.085002
35	MSE	0.0880057
40	MSE	0.1320985
45	MSE	0.1004459
50	MSE	0.1656062

3.4.2. Modelling with summer data

Model architecture, activation functions and separation of summer data was taken in the same manner with winter data. Meteorological factors and previous day's SO_2 value was integrated to the model as input parameters and next day's SO_2 concentration was tried to be predicted like in the winter data. 130x9 matrixes from summer data as input and 130x1 matrix as target was used in the model. 90 of data was used in training set, 20 in testing and 20 in validation sets for summer model. Hyperbolic tangent activation function for connecting input layer and hidden layer and linear activation function for connecting hidden layer and output layer were selected.

Moreover, same procedure used to winter data was applied to the summer data in order to determine best training function and neuron number in the hidden layer. 25 trials for each training function and 10 trails for determining neuron number in the hidden layer were applied. According to these trials, Levenberg-Marquardt backpropagation (trainlm) training function was selected as best option like in the model developed with winter data and 5 neurons in the hidden layer gave the best results differently from

winter model. Average values of 25 trials of training function are given in Table 3.8 and average MSEs of 10 trials of each neuron number in hidden layer are given in Table 3.9.

Table 3.8 Average values of 25 trials of each function for normalized summer data

Training function	Criteria	Avg. Values
trainlm	MSE	0.001119
	R	**0.332397**
	R^2	0.145539
trainrp	MSE	0.004775
	R	0.223746
	R^2	0.067071
trainscg	MSE	0.001374
	R	0.225626
	R^2	0.063927
traincgp	MSE	0.002227
	R	0.191542
	R^2	0.047455
traincgf	MSE	0.001069
	R	0.213416
	R^2	0.06624
traincgb	MSE	0.000966
	R	0.189762
	R^2	0.057706
trainbfg	MSE	0.001452
	R	0.133021
	R^2	0.02547

Table 3.9 Average MSEs of 10 trials of each neuron number for normalized summer data

Neuron number	Criteria	Avg. Values
5	MSE	**0.000896857**
10	MSE	0.001139685
15	MSE	0.001559346
20	MSE	0.001186972
25	MSE	0.00321715
30	MSE	0.0038636
35	MSE	0.00284139
40	MSE	0.00322328
45	MSE	0.003945779
50	MSE	0.00353073

4. RESEARCH RESULTS

4.1. Results of winter model

Developed model with winter data produced mean square error closer to zero which is important because it shows how close a fitted line is to data points and is used to evaluate the performance of a predictor or an estimator. Figure 4.1 shows the MSE results of the model.

Figure 4.1 MSE results of winter model

Other important criterion showing the prediction ability of the model is the R values which gives the correlation between output and target values. If R values close to 1, this means high correlation between target and predicted values. R values taken out by winter model are given in Figure 4.2. R value of training is 0.74569, testing is 0.44229, validation is 0.73291, and overall value is 0.67012.

Figure 4.2 R values taken out from winter model

4.2. Results of summer model

MSE of summer model is given in Figure 4.3. According the results of model developed with summer data, MSE value is closer to zero than winter model that means prediction performance of this model is better than winter.

Figure 4.3 MSE results of summer model

R values of the model are 0.92841 for training, 0.26353 for testing, 0.96069 for validation, and 0.81231 for overall. These R values are better than winter for training, validation and overall but testing results show worse performance than winter.

Predicted SO_2 concentrations according to overall performance of models and actual values are compared in Figure 4.5 and Figure 4.6 for winter and summer, respectively. As seen in the figures predicted and actual values are accurate with respect to model predictions.

Moreover, in Figures 4.7 and 4.8 shows the linear equations taken out by actual and predicted values for winter and summer. R^2 values taken from these graphs show that summer models' prediction performance is better than winter.

Figure 4.4 R values taken out from summer model

Figure 4.5 Comparison of predicted and actual SO_2 values for winter according to overall performance

Figure 4.6 Comparison of predicted and actual SO_2 values for summer according to overall performance

Figure 4.7 Comparison of actual and predicted values with linear equations for winter

Figure 4.8 Comparison of actual and predicted values with linear equations for summer

4.3. Evaluation of the effects of input parameters to SO_2 pollution

In order to evaluate the effects of input parameters to SO_2 concentration, weight estimation that is usable regression technique was used. This regression attempts to model the relationship between two or more explanatory variables and a response

variable by fitting an equation to observed data. Log-likelihood values for each power of the weight source variable tested. Weight Estimation procedure tests a range of weight transformations and indicates which will give the best fit to the data. Every value of the independent variable is associated with a value of the dependent variable. In this study independent variables are input parameters such as temperature, pressure, wind speed, humidity, rainfall, cloudiness, sun duration, sun radiation, and previous day's SO_2 concentration and dependent variable is actual (target) SO_2 concentration.

Weight Estimation equations developed by IBM SPSS Statistics 22.0 for winter and summer seasons are given in Table 4.1.

Table 4.1 Weight estimation regression equations for winter and summer seasons

Season	Weight Estimation Regression Equations	R	R^2
Winter	Target SO_2 = -1.242 + 0.014 PreviousSO_2 + 0.002 Temperature + 1.508 Pressure – 0.002 Windspeed – 0.034 Humidity + 0.008 Rainfall + 0.001 Radiation – 0.003 Cloudiness – 0.25 Sunduration	0,313	0.098
Summer	Target SO_2 = 0.469 + 0.047 PreviousSO_2 – 0.001 Temperature – 0.466 Pressure – 0.003 Windspeed – 0.025 Humidity + 0.013 Rainfall – 1.758E-5 Radiation – 0.001 Cloudiness + 0.00 Sunduration	0,328	0.108

These regression equations show the contribution of input parameters to target SO_2 concentration. In winter most important contributor to target SO_2 is pressure and sun duration, cloudiness, and wind speed negatively contribute to the model while in summer previous day's SO_2 concentration is the main parameter contributing to SO_2, rainfall and sun duration have positive effects and wind speed, humidity and radiation have negative effects on target SO_2 concentration.

R^2 values are important estimator giving the accuracy of these equations because R^2 gives idea about independent variable explains which percent of dependent variable. There are many factors affecting accuracy of these equations such as normalization hypothesis, usage of wrong function, dependency of errors. R^2 values taken out by these equations are 0.098 for winter and 0.108 for summer. These are very low values but they may give idea about what is important for model development. With respect to

assumptions of normalization of error, constant variance, and independency of error, seasons which have higher R^2values should give more accurate results in ANN models and the effects of independent variables should be seen in the dependent variable. Of course considering human factor is important to understand contributing parameters to SO_2 pollution (Kunt 2014).

5. DISCUSSION

Air pollution modelling is increasing trend nowadays. In literature there are many examples of air pollution modelling. SO_2 pollution has significant importance because human health, vegetation, animals, and structures are affected negatively from it. Therefore, SO_2 pollution which is one of the air pollutants should be modelled. In this study SO_2 pollution in Seydişehir town was predicted with artificial neural networks. Two models were developed for winter and summer seasons in Seydişehir and 67% prediction for winter and 81% prediction for summer were achieved in overall.

Prediction performance of models may be affected from many reasons. Because of the problems in SO_2 analyser device, there are some missing values for some months so it was impossible to evaluate SO_2 data for whole year. Thus, model was developed just for winter and summer season with available data. Moreover, meteorological data was provided from General Directorate of Meteorology and integrated to model as input parameter with SO_2 concentrations. For those reason measurement locations of meteorological data and SO_2 data were not enough closer from each other.

Mean square error is important estimator for determining average squared difference between output and target values. Lower values of MSE means good estimation and 0 means no error. According to results of this study, MSE of winter model is 0.021277 and MSE of summer model is 0.00019252. Akkoyunlu et al. (2010) predicted SO_2 pollution conducted for İstanbul. Three models were developed for winter, summer and overall. MSE values taken out by these models were as 0.0042634, 0.33397419 and 0.125757. Furthermore, Yüksek et al. (2007) made SO_2 prediction with ANN approach for Sivas and they found the MSE value as 0.00204885. Therefore, when the MSEs of this study are compared with previous ones, it is clearly seen that these values are acceptable and good.

Another important estimator used for determining model accuracy is R values. In ANN approach data is divided in three subsets called training, testing and validation. After training, ANN model gives figures for these subsets. Evaluation of the performance of the model is done with R values which show prediction performance of the model. R values closer to 1 means that high performance. In this study R values are unstable

between subsets for summer season. There are several reasons of that kind of variance. One of the reasons is memorizing the learning process during training. (Basheer and Hajmeer 2000) Moreover, size of the datasets is important because too many or too low data may produce bad results

Chelani et al. (2002) tried to predict SO_2 concentrations for Delhi city in India with artificial neural networks including wind speed, temperature, relative humidity, and the wind direction index as input parameters. Area was separated into three parts namely commercial, industrial and residential and three models were developed for each. Correlation between target and predicted values for each area were found as 0.68, 0.72, and 0.63. Moreover, Dursun et al. (2015) conducted a study for center of Konya and SO_2 values predicted with indication of meteorological parameters by ANN and adaptive neuro-fuzzy inference system (ANFIS). Study results showed that total response of the ANN model is 0.778.

Developed models for winter and summer can be compared according to their R values taken by validation data set. R value of validation set is 0.73 for winter and 0.96 for summer. Validation set is used in most models for evaluate generalisation performance of ANN that is whether the approximation performance of model is efficient enough (Viotti, Liuti et al. 2002).

Error in the validation set should be decreased during training. If data start to overfit which means generalisation ability of the model does not fit the new data, validation set error start to increase. In order to prevent this situation early stopping was applied to models. “When the validation error increased for a specified number of iterations, the training was stopped, and the weights and biases at the minimum of the validation error were returned.” (Akkoyunlu 2011) The accuracy between test and validation sets shows model working unproblematic. R value of testing set for summer data did not show the same characteristics with validation data while for winter data test set and validation set performances showed similar characteristics. Although validation performance of summer model gave better results than winter season, test accuracy of winter season is better than summer season. Test set is defined as “the group of data, given to the network still in the learning phase, by which the error evaluation is verified in order to effectively update the best thresholds and the weights” by Viotti and Lituti (2002). This

definition explains main purpose of test set in ANN modelling. In this study R values of test do not represent the efficiency of models. They just help weight and bias determination during training step. Test set is used for understanding the overfitting problem. In the case of summer model there might be overfitting problem occurred in validation set. However, when overall prediction performances of the models are compared, results of summer model are more successful.

Furthermore, weight estimation regression applied to winter and summer data show the effects of independent parameters to dependent parameter. The R^2 of summer season is higher than winter season so the contribution of input parameters to SO_2 pollution is more in summer season. ANN model developed in summer have more correlation between input parameters and predicted value that is compatible with the results taken from regression equations. Also, there are other factors that should be considered before determining which actually affects the air pollution mostly human factor.

6. CONCLUSION AND SUGGESTIONS

In this study SO_2 pollution in Seydişehir town was modelled with artificial neural networks. Two models were developed for winter and summer seasons in Seydişehir and 67% prediction for winter and 81% prediction for summer were achieved in overall. Besides, R value of validation set is 0.73 for winter and 0.96 for summer. According to these results, generalisation performance of summer model gave better results than winter model.

While there are many studies focusing on SO_2 modelling, it is a new subject in Turkey and most of the study subjecting to city air pollution modelling but there is no study focusing on town. In that manner this study is the first one focusing on town air pollution modelling and may be baseline of further researches.

Sulphur dioxide emissions are represented as one of the most common air pollution problem throughout the world. There are so many examples of negative effects of it such as lethal London smog event and other lethal effects on vegetation and water bodies with acidic deposition (Zannetti 1990). It is very important to predict future concentration of SO_2 in order to establish emission control legislations, evaluate the future emission control impacts, select possible locations in which it can be possible source of air pollutants, control air pollutant episodes and take preventive precautions, and assess present air pollutant sources and their responsibility in future events (Zannetti 1990). Thus, this study will be useful source for evaluating Seydişehir SO_2 pollution and in future new studies may be done at same town and models which have long term prediction performance may be developed.

Besides, the prediction performance of this study is acceptable when it was compared with the studies in literature. It is possible to take more accurate results using ANN approach but many factors affect the prediction performance during modelling such as abundance or lack of data and under or over training.

One of the problems for developing model is different measurement locations of meteorological data and SO_2 data. In order to overcome this problem, device which does meteorological measurements should be placed at the location of SO_2 pollution

measured. By this means better model results may be taken out. Moreover, measurement duration should be increased and time intervals between each measurement should be decreased so more data may be gathered in order to reach better results.

If more pollutants are measured such as NO_2, CO_2, CO, and PM, prediction capability of models may be increased. Thus, new measurement device which has capability of measuring more pollutants should be provided.

REFERENCES

Akkoyunlu, A., Yetilmezsoy K., Erturk, F. and Oztemel, E., 2010, A neural network-based approach for the prediction of urban SO_2 concentrations in the Istanbul metropolitan area, *Int. J. Environment and Pollution*, Vol. 40, No. 4, 301-321.

Alsugair, A. M., Al-Qudrah, A. A., 1998, Artificial neural network approach for pavement maintenance, *J. Comput. Civil Eng. ASCE*, 2 (4), 249–255.

Arabacı M., Bayram M., Yüceer M. and Karadurmuş E., 2010, Tuğla Ve Kiremit Fabrikalarının Hava Kirliliğine Katkılarının Yapay Sinir Ağı Modellemesi İle Araştırılması, *UKMK- 9 – Gazi Üniversitesi*, 22-25 Haziran.

Basheer, I. A. and Hajmeer, M., 2000, Artificial neural networks: fundamentals, computing, design, and application, *Journal of Microbiological Methods*, 43(1), 3-31.

Boubel, W. R., Fox, D. L., Turner, D. B. and Stern, A. C., 1994, Fundamentals of Air Pollution, Third Edition, *Elsevier,* USA, 3-153.

Bourquin, J., Schmidli, H., van Hoogevest, P. and Leuenberger, H., 1998, Advantages of Artificial Neural Networks (ANNs) as alternative modelling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form, *European Journal of Pharmaceutical Sciences*, **7**(1): 5-16.

Boznar, M., Lesjak, M. and Mlakar, P., 1993, A neural network-based method for short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain, *Atmospheric Environment. Part B. Urban Atmosphere*, 27(2), 221-230.

Chelani, A.B., Chalapati Rao C.V., Phadke K.M. and Hasan M.Z., 2001, Prediction of sulphur dioxide concentration using artificial neural networks, *Environmental Modelling & Software 17*, 161–168.

Clair, T. A. and Ehrman, J. M., 1998, Using neural networks to assess the influence of changing seasonal climates in modifying discharge, dissolved organic carbon, and nitrogen export in eastern Canadian rivers, *Water Resources Research*, 34(3), 447-455.

Comrie, A. C., 1997, Comparing Neural Networks and Regression Models for Ozone Forecasting, *Journal of the Air & Waste Management Association*, 47(6), 653-663.

Çelebi S., Havadaki SO2 konsantrasyonunun yapay sinir ağlı bulanık mantık yöntemi ile modellenmesi: Erzurum şehir örneği, *Yüksek Lisans Tezi*, Gebze Yüksek Teknoloji Enstitüsü.

Dehkordi M. B., 2012, Compressive Sensing Based Compressed Neural Network for Sound Source Localization, *American Journal of Intelligent Systems*, 2(4), 35-39.

Dursun, S., 1994, The Effects of Sulphur Pollution On Soil Fungi And Decomposition of Tree Leaf Litters, Thesis of Philosophiae Doctor, *University of Wales*, Collage of Cardiff, 8.

Dursun, S., Kunt, F. and Taylan, O., 2015, Modelling sulphur dioxide levels of Konya city using artificial intelligent related to ozone, nitrogen dioxide and meteorological factors, *International Journal of Environmental Science and Technology*, 12(12), 3915-3928.

Gardner, M. W. and Dorling, S. R., 1998, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, *Atmospheric Environment*, 32(14–15), 2627-2636.

Gardner, M. W. and Dorling, S. R., 1999, Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London, *Atmospheric Environment*, 33(5), 709-719.

Ghazali, S. I., 3-5 April 2012, Lokman Hakim Air quality prediction using artificial neural network, *In: The International Conference on Civil and Environmental Engineering Sustainability (IConCEES 2011)*, Johor Bahru, Malaysia.

Ginsberg, M., 1993, Essentials of Artificial Intellegence, *Morgan Kaufmann Publishers*, San Francisco, California, 3-10.

Haykin, S., 2001, Neural Networks: A Comprehensive Foundation, *Second ed. Pearson Education Inc.*, New Delhi, India.

Hornik, K., Stinchcombe, M., White, H., 1989, Multi layer feedforward networks are universal approximators, *Neural Netw. 2*, 359–366.

Hussain, S.T., 2011, Sulfur Dioxide: Properties, Applications And Hazards, *Nova Science Publishers, Inc.*, Chapter 3, 49-68.

Kırımhan, S., 2006, Hava Kirliliği ve Kontrolü, *Turhan Kitabevi*, Ankara, 54-70.

Kunt F., 2014, Bulanık Mantık Ve Yapay Sinir Ağları Yöntemleri Kullanılarak Konya İl Merkezi Hava Kirliliği Modellenmesi, Doktora Tezi, *Selçuk Üniversitesi Fen Bilimleri Enstitüsü*, Konya.

Mlakar, P., Božnar, M. and Lesjak, M., 1994, Neural Networks Predict Pollution, *Air Pollution Modeling and Its Application X*, S.-E. Gryning and M. Millán, Springer US, 18, 659-660.

Müezzinoğlu, A., 1987, Hava Kirliliği ve Kontrolü Esasları, *Dokuz Eylül Yayınları*, İzmir.

M100E Operation Manual, 2011, UV Fluorescence SO2 Analyzer, *Teledyne Advanced Pollution Instrumentation 9480*, Carroll Park Drive San Diego, CA 92121-5201, USA, http://www.teledyne-api.com/manuals/04515F_100E.pdf, retrieval date: 02.12.2015.

Nilsson, N.J., 1998, Artificial Intelligence: A New Synthesis, *Morgan Kaufmann Publishers*, San Francisco, California, 1-11.

Ozturk, Z. C., Dursun, S., 2016, Modelling of Atmospheric SO2 Pollution in Seydişehir Town by Artificial Neural Networks. *J. Int. Environmental Application & Science*, Vol. 11(1): 1-7.

Pham, D. T. and Pham, P. T. N., 1999, Artificial intelligence in engineering, *International Journal of Machine Tools and Manufactur*, 39(6), 937-949.

Rege, M. A. and Tock, R.W., 1996, A simple neural network for estimating emission rates of hydrogen sulphide and ammonia from single point source, *J. Air Waste Manage. Assoc. 46*, 953–962.

Rozlach, Z., 2015, Data-driven Modelling in River Channel Evolution Research: Review of Artificial Neural Networks, *J. Int. Environmental Application & Science*, Vol. 10(4), 384-398.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J., 1986, Learning internal representations by error propagation, In: Rumelhart, D.E.,McClelland, J.L. (Eds.), *Parallel Distributed Processing, eighthed. MIT Press*, Cambridge, England, 45–76.

Rumelhart, D. E., McClelland, J. L., 1995, Parallel Distributed Processing: Explorations in the Microstructure of Cognitions, *Vol.1. MIT Press*, Cambridge, England.

Sarı, S. and İnan, N., 2011, Comparison of Climate of Seydişehir and Beyşehir, *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26 / 2011.

Sarle, W. S., 1994, Neural Networks and Statistical Models, *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC, USA, 1-13.

Sarle, W., 1997, Neural network frequently asked questions, ftp://ftp.sas.com/pub/neural/FAQ.html, retrieval date: 07.10.2015.

Simpson, P.K., 1989, Artificial Neural Networks: Paradigms, Application and Hardware Implementations, *Foundations of neural networks. In: Sinencio, E.S., Lau, C. (Eds.)*, IEEE Press, NY, 3–24.

Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M. and Pereira M. C., 2007, Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, *Environmental Modelling & Software*, 22(1), 97-103.

Tecer, L.H., 2007, Prediction of SO_2 and PM Concentrations in a Coastal Mining Area (Zonguldak, Turkey) Using an Artificial Neural Network, *Polish J. of Environ. Stud. Vol. 16*, No. 4, 633-638.

URL 1, 2014, Seydişehir İlçe Raporu, *Mevlana Kalkınma Ajansı (MEVKA),* Konya, http://www.mevka.org.tr/Download.aspx?filePath=yHE9bIuGshyovkGX9Z1VhQ==, retrieval date: 15.11.2015.

URL 2, Seydişehir Eti Alüminyum Tesisi, http://www.etialuminyum.com/tr-tr/Tesisler/Sayfalar/Seydisehir-Eti-Aluminyum-Tesisi.aspx, retrieval date: 15.11.2015.

URL 3, 2007, Bart Determination for Alcoa Intalco Works Ferndale, Washington http://www.ecy.wa.gov/programs/air/globalwarm_reghaze/BART/IntalcoBARTDeterminationFINAL.pdf, *ENVIRON Corporation*, retrieval date: 15.11.2015.

URL 4, Particulate Matter (PM), http://www3.epa.gov/pm/, retrieval date: 20.10.2015.

URL 5, National Ambient Air Quality Standards (NAAQS), http://www3.epa.gov/ttn/naaqs/criteria.html, retrieval date: 20.10.2015.

URL 6, 2008, Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği, http://www.mevzuat.gov.tr/Metin.Aspx?MevzuatKod=7.5.12188&MevzuatIliski=0&sourceXmlSearch=hava%20kalitesi, retrieval date: 20.10.2015.

URL 7, http://www.thestudentroom.co.uk/showthread.php?t=934612, retrieval date: 20.10.2015.

URL 8, 2015, Konya İli 2014 Yılı Çevre Durum Raporu, http://www.csb.gov.tr/db/ced/editordosya/Konya_icdr2014.pdf, retrieval date: 20.11.2015.

URL 9, http://mapsof.net/map/konya-districts, retrieval date: 26.11.2015.

URL10, Konya İlçeleri, http://www.cografya.gen.tr/tr/konya/ilceler.html, retrieval date: 26.11.2015.

URL 11, İklim: Seydişehir, http://tr.climate-data.org/location/8607/, retrieval date: 26.11.2015.

URL 12, http://www.met.ie/about/weatherobservingstations/rochespoint.asp, retrieval date: 14.12.2015.

Vesilind, P.A., 1975, Environmental Pollution And Control, *Ann Arbor Science Publishers, Inc.*, Ann Arbor, Michigan 48106, 155.

Vesilind, P.A., Morgan, S.M. and Heine, L.G., 2010, Introduction to Environmental Engineering, *Cengage Leraning*, Stamford-USA, 415-420.

Viotti, P., Liuti, G. and Di Genova, P., 2002, "Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia." *Ecological Modelling*, 148(1), 27-46.

Wang, S.C., 2003, Artificial Neural Network, *The Springer International Series in Engineering and Computer Science*, Volume 743, 81-100.

Wasserman, P. D., 1989, Neural Computing: Theory and Practice, *VanNostrand Reinhold*, NY.

Yüksek, G. A., Bircan, H., Zontul, M. and Kaynar, O., 2007, Sivas İlinde Yapay Sinir Ağları ile Hava Kalitesi Modelinin Oluşturulması Üzerine Bir Uygulama, *C.Ü. İktisadi ve İdari Bilimler Dergisi*, Cilt 8, 97-112.

Zannetti, P., 1990, Air Pollution Modeling Theories, Computational Methods and Available Software, *Springer Science + Business Media, LLC*, AeroVironment Inc. Monrovia, California, 3-20.

Zurada, J.M., 1997, Introduction to Artificial Neural Systems, *WestPublishing Company*, Mumbai, India.

CURRICULUM VITAE

PERSONAL INFORMATIONS

Name Surname	**:**	Zeynep Cansu ÖZTÜRK
Nationality	**:**	T.C
Birth Place and Date	**:**	ANKARA – 10.09.1990
Telephone	**:**	2232113/05444210664
Fax	**:**	2410635
e-mail	**:**	zcozturk@selcuk.edu.tr

EDUCATION

Degree		**Name, Town, City**	**Graduation Date**
High School	:	Private Science High School of Gazi University Foundation, Çankaya, Ankara	06.2007
University	:	Middle East Technical University, Çankaya, Ankara	06.2013
Master of Science	:	Selçuk University, Selçuklu, Konya	-
Doctorate	:		

WORK EXPERIENCES

Year	**Institution**	**Position**
02.2015-	Selçuk University	Research Assistant
06.2013 -01.2014	EN-ÇEV Energy Environmental Consultancy	Environmental Engineer
06.2012 - 07.2012	Ministry of Environment and Urbanization	Intern
06.2011 - 07.2011	General Directorate of Water and Sewerage Administration	Intern

EXPERTICE AREA

Air Pollution, Modelling, Occupational Safety

FOREIGN LANGUAGES

English

UNDERSTANDING		SPEAKING		WRITING
Listening	Reading	Spoken interaction	Spoken production	
C2	C1	B1	B2	C1

Levels: A1/2: Basic user - B1/2: Independent user - C1/2 Proficient user
Common European Framework of Reference for Languages

OTHER FEATURES

Computer Skills

- good command of Microsoft Office™ tools
- good command of AutoCAD Software
- little command of C program
- good command of WaterCAD, SewerCAD Software
- good command of MATLAB Software
- good command of IBM SPSS Statistics

Certificates

- Environmental Officer Certificate
  Ministry of Environment and Urbanization - 12.2013
- C-Class Occupational Safety Specialist Certificate
  Ministry Of Labor and Social Security – 11.2015

PUBLICATIONS

Articles

Ozturk, Z.C, Dursun, S., (2016) Modelling of Atmospheric SO_2 Pollution in Seydişehir Town by Artificial Neural Networks. *J. Int. Environmental Application & Science*, Vol. 11(1): 1-7.

Proceedings

Ozturk, Z. C., Kunt, F., Dursun, S. (2015) Application of Artificial Neural Network in Environmental Engineering Problems, International Conference of Ecosystems (ICE2015), 109, 368-375.

Kunt, F., Ozturk, Z. C., Dursun, S. (2015) Used Some Modelling Applications in Air Pollution Estimates, International Conference of Ecosystems (ICE2015), 138, 82.

Dursun S, Kunt, F., Ozturk, Z. C., Vrenozi, B., (2015) Global Climate Change Effects on Ecology, 2nd ICSAE 2015 İnternational Conference on Sustainable Agriculture and Environment, 182.