DATA ANALYTIC STUDIES FOR TURKEY’S ENERGY FORECAST

APPROVED BY SUPERVISING COMMITTEE:

________________________________________
Mo Jamshidi, Ph.D., Chair

________________________________________
Hariharan Krishnaswami, Ph.D.

________________________________________
Shuo Wang, Ph.D.

Accepted: _________________________________________
Dean, Graduate School
DEDICATION

This work is dedicated to my family and my fiancee who are always very supportive of my accomplishments. Thank you for providing me with constant motivation.
ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my advisor Dr. Mo Jamshidi for his patient guidance and support during my years of study at the University of Texas at San Antonio. The many lessons that I have learned from him will stay with me for the rest of my academic career.

I would also like to thank my dissertation committee, Dr. Shuo Wang and Dr. Hariharan Krishnaswami for their support, comments and suggestions.

I also wish to acknowledge that support of my friends at UTSA. I would like to thank the staff and faculty of the Electrical and Computer Engineering Department for their support during my master degree. Finally, I would like to thank my family, for their support at all times.

May 2014

DATA ANALYTIC STUDIES FOR TURKEY’S ENERGY FORECAST

iv
The location of Turkey is between 36 and 42 N latitudes, it means close to axis of equator. Therefore, Turkey is very rich in terms of the potential of renewable energy sources although an important part of this potential is not active yet. The share of renewable energy sources in the total production of electrical energy was 25.2% in 2011. The share of wind and PV are 2.1% and 1% in production electricity. The role of PV systems is at infant stage. Energy gap between production and consumption has been increasing in world. The reason for lack of PV and wind systems is financial feasibility of systems. Renewable resources such as PV systems and wind power systems are built considering to their efficiency of output power. Solar irradiation is one of the major renewable energy sources, but the forecasting of solar irradiation depends on meteorological parameters such as air temperature, cloud base height, relative humidity, and wind speed and air pressure. Data analytic tools help to forecast output power of renewable systems.

In this dissertation, data analytic tools are used to provide an increment in the share of renewable resources in production electricity. An artificial neural network (ANN) model was created to estimate the hourly solar irradiation and wind speed. Dataset was recorded in Antalya in 2013 by the Turkish State and Meteorological Service. Furthermore, this study is purposed to improve the accuracy of energy forecasting. This improvement could be realized by finding the best structure of ANN for forecast energy, and by developing the performance of ANN using data analytics tools such as Genetic Algorithm (GA) and Principal Component Analysis (PCA).
Analysis of big data can distinguish relevant information and extract useful knowledge from apparently unrelated data that is formed in a massive volume. This meaningful information can be used to forecast environmental behavior, which helps maximize the value of wind and solar output power estimations.

In addition PCA is chosen for reducing the dimension of datasets to save time and memory cost with a better network performance. GA is chosen to improve network performance by finding and fixing the best weight for ANN.
# TABLE OF CONTENTS

Acknowledgements........................................................................................................ iv  
Abstract .......................................................................................................................... v  
List of Tables ................................................................................................................... vii  
List of Figures ................................................................................................................ viii  

Chapter One: Introduction .............................................................................................. 1  
   1.1 Background of Turkey’s Energy Sector ................................................................. 2  
   1.2 Background of Turkey’s Solar and Wind Power .................................................... 2  
      1.2.1 Turkey Wind Energy Potential ..................................................................... 2  
      1.2.2 Turkey Solar Energy Potential ..................................................................... 2  
   1.3 Problem Statement and Formulation .................................................................... 2  
   1.4 Contributions of Thesis......................................................................................... 2  

Chapter Two: Data Analytics Tools and Modeling ......................................................... 4  
   2.1 Introduction ........................................................................................................... 2  
   2.2 Data Analytics Tools............................................................................................. 2  
      2.2.1 Genetic Algorithm (GA) ............................................................................. 2  
      2.2.2 Artificial Neural Network (ANN)................................................................. 2  
         2.2.2.1 Multilayer Perceptron and Back Propagation (BP) ............................. 2  
      2.2.3 Principal Component Analysis (PCA) ......................................................... 2  
   2.3 Turkey Energy Data-sets......................................................................................... 2  
   2.4 The Approach of Implementation......................................................................... 2  

Chapter Three: Solar Forecasting Using Neural Network ............................................ 18  
   3.1 Database ................................................................................................................ 2
3.2 Implementation Model

3.2.1 Selection Input Variables

3.2.2 Setting of Neural Network

3.2.3 Mining of Training Data and Validation of the Designed Networks

3.2.4 Applying One Year Data to ANN

3.3 Conclusion

Chapter Four: Optimization the Performance of Neural Network with Using GA and PCA

Chapter Five: Conclusion and Future Work

Appendices

References

Vita
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Primary energy balance in Turkey in 2012 (Mtoe) (MENR, 2012) .................3</td>
</tr>
<tr>
<td>Table 2</td>
<td>Annual development of Turkey’s gross electricity generation by share of primary energy sources, (%).................................................................5</td>
</tr>
<tr>
<td>Table 3</td>
<td>Hourly meteorological dataset Antalya/Turkey ........................................8</td>
</tr>
<tr>
<td>Table 4</td>
<td>Effects of neuron on the performance of ANN...........................................10</td>
</tr>
<tr>
<td>Table 5</td>
<td>Effects of training percentage on the performance of ANN .........................20</td>
</tr>
<tr>
<td>Table 6</td>
<td>The comparing results between ANN and GA optimized ANN .......................10</td>
</tr>
<tr>
<td>Table 7</td>
<td>The comparing the results of ANN and PCA-ANN ...................................20</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Turkey energy balance between production and consumption ..................................................3</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Renewable sources in production electricity (2011)...........................................................4</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Potential wind energy atlas of Turkey ......................................................................................4</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Solar map of Turkey ..................................................................................................................10</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Structure of a Neural Network.................................................................................................10</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Feed-Forward Neural Network.................................................................................................22</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Life-cycle of populations and the process of crossover and mutation..................................10</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Process of Genetic Algorithm.................................................................................................22</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Correlations between Air Temperature and Solar Irradiance ...................................................22</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Correlations between Relative Humidity and Solar Irradiance ...............................................22</td>
</tr>
<tr>
<td>Figure 11</td>
<td>The view of structure of a neural network ................................................................................22</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Artificial Neural Network with 1 hidden neuron in a hidden layer (SI) .................................22</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Artificial Neural Network with 5 hidden neurons in a hidden layer (SI) ..............................10</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Artificial Neural Network with 10 hidden neurons in a hidden layer (SI) ..........................22</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Artificial Neural Network with 20 hidden neuron in a hidden layer (SI) ............................22</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Artificial Neural Network with 5 hidden neurons in a hidden layer (WS) ............................22</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Artificial Neural Network with 20 hidden neuron in a hidden layer (WS) ..........................22</td>
</tr>
<tr>
<td>Figure 18</td>
<td>The training data is the % 70 percentages of dataset (SI) ......................................................10</td>
</tr>
<tr>
<td>Figure 19</td>
<td>The training data is the % 50 percentages of dataset (SI) ......................................................10</td>
</tr>
<tr>
<td>Figure 20</td>
<td>The training data is the % 30 percentages of dataset (SI) ......................................................10</td>
</tr>
<tr>
<td>Figure 21</td>
<td>One year solar irradiance prediction results ............................................................................10</td>
</tr>
<tr>
<td>Figure 22</td>
<td>One week hourly solar irradiance prediction by using only ANN ...........................................10</td>
</tr>
<tr>
<td>Figure 23</td>
<td>One week hourly solar irradiance prediction by using GA-ANN hybrid model ...10</td>
</tr>
<tr>
<td>Figure 24</td>
<td>One month hourly solar irradiance prediction by using only ANN...................10</td>
</tr>
<tr>
<td>Figure 25</td>
<td>One month hourly solar irradiance prediction by using GA-ANN hybrid model .10</td>
</tr>
<tr>
<td>Figure 26</td>
<td>One year hourly solar irradiance prediction by using only ANN.......................10</td>
</tr>
<tr>
<td>Figure 27</td>
<td>One year hourly solar irradiance prediction by using GA-ANN hybrid model.....10</td>
</tr>
<tr>
<td>Figure 28</td>
<td>One month hourly solar irradiance prediction by using reduced dataset (200) .....10</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION

1.1 Background of Turkey Energy Sector

During the past decade, the importance of energy efficiency has been the most crucial topic for government to reduce a country’s dependence on foreign energy resources. Energy is one of the most fundamental requirements for economic and social development of the countries. Energy consumption has increased rapidly because of economical and social growth in recent years. Energy efficiency issue is more important for countries which are developing countries [1].

Approximately 75% of the world’s natural gas and products of petroleum have been reserved in the Middle-East, Europe, Russia and Central-Asia [2]. Turkey is located between Europe and Asia. It bordered by the Black Sea, the Aegean Sea and the Mediterranean Sea. Its geographical position provides to play important role to transfer oil and natural gas from producing countries in the Middle East to European countries [3]. However Turkey’s geographical position looks eligible to have petroleum products or natural gas, its domestic sources like oil and natural gas reserves and their production is low but consumption is high. In the 1990s energy consumption was 4.4% per year, and after the 1990s the annual average of consumption rate was 4.6% while the EU member country’s consumption rate was 1.6%. Turkey also had the highest energy demand increase rate in the OECD countries [4].

Turkey is one of the members of the G-20 major economies and classified developed country by the CIA, the energy efficiency issue of Turkey is not like developed countries. Turkey has become the world’s 17th largest economy. Turkey’s energy demand has risen rapidly as a result of industrialization and rapid population growth. In 2013, the population of Turkey was 76 million. Based on information from the Turkish Statistical Institute (TUIK), between 1990 and 2000, the average growth rate of population was 1.55% per year and between 2000 and 2010, it was 1.41%
The population growth rate is estimated to slow down to 1.2 and the population will reach 82 million in 2020. Turkey has been growing in the way of both its population and economy.

According to MENR; in 2012 Turkey’s primary energy supply was 126,530 Mtoe; 31,527 Mtoe had supplied from domestic energy sources, 95,003 Mtoe which is equal to 75.08 % of total energy supply had supplied with importing. The gap between production and consumption has been increased and this gap has been supplied with importing energy. The dependence of foreign source is shown in figure1.

![Energy Imports](image)

**Figure 1:** Turkey energy balance between production and consumption

Turkey does not have enough indigenous resources, and imports more than 75% of their energy consumption. Considering Turkey natural gas demand, almost 99% of the natural gas has provided by imports in 2012. Moreover, the biggest share in energy supply belongs to oil (31.85%), natural gas (30.38%), hard coal (16.07%) and lignite (14.12%) (Table1).
Table 1: Primary energy balance in Turkey in 2012 (Mtoe) (MENR, 2012)

<table>
<thead>
<tr>
<th></th>
<th>Production</th>
<th>Import</th>
<th>Total Primary Energy Supply</th>
<th>Total Energy Supply (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oil</strong></td>
<td>2,440</td>
<td>37,856</td>
<td>40,296</td>
<td>31.85</td>
</tr>
<tr>
<td><strong>Natural Gas</strong></td>
<td>533</td>
<td>37,910</td>
<td>38,443</td>
<td>30.38</td>
</tr>
<tr>
<td><strong>Hard Coal</strong></td>
<td>1,095</td>
<td>19,237</td>
<td>20,332</td>
<td>16.07</td>
</tr>
<tr>
<td><strong>Lignite</strong></td>
<td>17,860</td>
<td>0</td>
<td>17,860</td>
<td>14.12</td>
</tr>
<tr>
<td><strong>Hydroelectric</strong></td>
<td>4,976</td>
<td>0</td>
<td>4,976</td>
<td>3.93</td>
</tr>
<tr>
<td><strong>Geothermal and other geothermic</strong></td>
<td>2,236</td>
<td>0</td>
<td>2,236</td>
<td>1.77</td>
</tr>
<tr>
<td><strong>Animal and Vegetables Wastes</strong></td>
<td>1,115</td>
<td>0</td>
<td>1,115</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>Wind</strong></td>
<td>504</td>
<td>0</td>
<td>504</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>Solar</strong></td>
<td>768</td>
<td>0</td>
<td>768</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>31,527</td>
<td>95,003</td>
<td>126,530</td>
<td>100.00</td>
</tr>
</tbody>
</table>

In parallel with its consumption for energy, especially electricity demand has been increasing fast. Electricity energy is a vital input for technical, social and economic development of Turkey as the other countries. The electricity consumption had increased by 4.4% in 2012 which is 239.497TWh compared to 229.395TWh in 2011. In 2020 the average annual growth rate of total electricity consumption is expected increase of about 6.9% to 392 TWh for high scenario and increase of 5.5% to reach 357.4 TWh for baseline scenario. The total installed power generation capacity was reached by 64MW in 2013[5].

In early 1980s when the gap between Turkey electricity demand and production was not too much, this demand was supplied with domestic energy sources as lignite, fuel oil, and hydropower. Since the demand is increasing fast and production is not enough to provide it, natural gas and oil has become a solution to generate electricity. The major using of natural gas in Turkey has been for electricity generation.

In 1985, the share of natural gas in Turkey’s annual gross electricity generation was only 1.1%, which exponentially rose to, 25.2 % in 2001, and finally 36.3% in 2011, as shown in Table 2 [6]. In addition, natural gas has been used as premier residential heating and cooking. According
to MENR, Turkey has purchased more than 20.0% of Turkey’s annual trade deficit which is estimated at US$ 20.0 billion to import natural gas in 2011 [4].

**Table 2:** Annual development of Turkey’s gross electricity generation by share of primary energy sources, (%) [5].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lignite</td>
<td>27.9</td>
<td>31.4</td>
<td>29.9</td>
<td>28.9</td>
<td>23.9</td>
<td>18.4</td>
<td>16.6</td>
<td>15.5</td>
</tr>
<tr>
<td>Hard Coal</td>
<td>2.6</td>
<td>2.4</td>
<td>2.0</td>
<td>1.6</td>
<td>1.2</td>
<td>0.9</td>
<td>0.9</td>
<td>1.3</td>
</tr>
<tr>
<td>Imported Coal</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>4.3</td>
<td>6.6</td>
<td>7.3</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>0.0</td>
<td>1.1</td>
<td>13.5</td>
<td>14.0</td>
<td>25.8</td>
<td>35.5</td>
<td>36.7</td>
<td>36.3</td>
</tr>
<tr>
<td>Geothermal</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Fuel Oil</td>
<td>16.1</td>
<td>15.5</td>
<td>9.6</td>
<td>7.4</td>
<td>6.1</td>
<td>7.0</td>
<td>4.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Diesel Oil</td>
<td>7.4</td>
<td>6.9</td>
<td>3.3</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Other</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>Renewable and Wastes</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Hydro</td>
<td>45.8</td>
<td>42.5</td>
<td>41.5</td>
<td>47.1</td>
<td>41.0</td>
<td>33.2</td>
<td>32.0</td>
<td>32.2</td>
</tr>
<tr>
<td>Wind</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>2.7</td>
<td>3.2</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Considering amount of imported natural gas and limited storage capacity for delivery of natural gas will have affects Turkey economy. There are other sources like hard coal and fuel oil which is less expensive than natural gas but it is easy to understand the damage of environment. Therefore, the countries who undertake management, energy should be to find ways like uninterrupted, reliable, clean and cheap. Turkey’s energy strategy is aimed at satisfying with protecting economic growth.

In Turkey, projections have made officially by Ministry of Energy and Natural Resources of Turkey (MENR). Projections indicate a continuing increase in demand for energy, especially for electricity According to MENR Resource Strategic Plan (2010-2014), improving efficiency of domestic source, maximum use of the renewable energy resources and the redesigning of Turkey
energy sector which is not widely depended natural gas to reducing the import dependence and balance between production and consumption have been purposed.

The share of renewable energy sources in electricity generation of 19.6% (2010) is expected to reach level 22.8% in 2020 and 23.8% in 2035 [7]

Turkey needs to improve especially renewable energy sources. Turkey is on very low level for renewable energy plants, however Turkey’s location is convenient for solar and wind energy plants. Targets for the electricity sector on the basis of MENR energy policies are using more lignite and bituminous coals for the production of electrical resources, the share of renewable energy sources in energy supply to be increased to 30%, installed wind power capacity will be increased to 20,000 MW.

To achieve balance between demand and production with minimum dependence foreign energy source to prevent economic growth, improving efficiency of renewable energy and building renewable energy plants should be a first step for Turkey. Across the World, forecasting for renewable energy sources is becoming more important. In Turkey, MENR has used Model for Analysis of Energy Demand (MAED) simulation technique for energy demand forecasting. In addition, data analytics tools have begun to be used for forecasting. Section two will discuss Turkey’s renewable energy source background and research to clarify the status of Turkey and what can do to improve renewable energy efficiency of Turkey.

1.2 Background of Turkey’s Renewable Energy Sources

Alternative energy sources are becoming more desirable in the world because the primary energy sources such as oil, natural gas and coal are decreasing and pollution is becoming a global threat. Fossil (hydrocarbons) as compared with the energy source of renewable energy sources
globally is seen that the utilization rate is quite low but the role of renewable sources in the global has become more crucial issue. Therefore, energy consumers of global actors who are trying to solve the problem of security of energy supply are beginning to turn to renewable energy sources as required. Globally, renewable generation is estimated to rise to 25% of gross power generation in 2018, up from 20% in 2011. Renewable electricity generation is estimated to rise to 40%, from 4860 TWh in 2011 to 6850 TWh in 2018 [8].

Turkey is very rich in terms of the potential of renewable energy sources although an important part of this potential is not active yet. In company with the increasing fears related to global warming and climate change, the MENR aims to take control of the dependence on foreign energy source by local and renewable energy sources. While the average share of renewable energy in total primary energy supply of OECD countries was 6.1% in 1900s and 9.0 in 2012, the share of renewable energy in total primary energy supply for Turkey in 1990 was around 18.3%, and fell to 10.2% in 2012 [8]

Bio-mass and hydro principal types of renewable energy sources in Turkey and geothermal, wind and solar energy are the types of renewable energy currently used in low rates. The renewable energy source in Turkey is still at a ground level with the majority of electrical generation capacity still coming from hydrocarbons as 74.8 percent of total installed capacity in 2011. Hydro power plants contribute 22.8 percent, wind power plants contribute 2.1 percent, and the remaining 0.3 percent is coming from other renewable sources.

While the share of renewable sources in primary energy supply is decreasing, in 2010 the share of renewable energy sources in the production of electrical energy has reduced to the level of 26.3%. However, due to building new wind power plants, the share of renewable energy sources in the total production of electrical energy was 25.2% in 2011. In addition, the rate of renewable
resources in total supply and share of renewable resources in production electricity are shown at Figure2 [9].

**Figure 2:** Renewable Sources in Production Electricity (2011)

### 1.2.1 Turkey Wind Energy Potential

Turkey is surrounded by sea on three sides has a coastline of about 3,500 km. Turkey in terms of wind energy potential is one of the richest countries in Europe. Turkey introduces a desirable geography to generate wind energy. The total wind energy potential is estimated at 48000 MW and the installed capacity of wind energy is 1792.7 MW in 2012. According to MENR, the total existing capacity of 9.454.90 MW under construction and in the licensing process is one fifth of the total potential, so 80% of the potential is still expected to be utilized. Wind Energy Potential Atlas (REPA) above 100 meter average height was produced in 2006 by EIE in order to determine the distribution and the characteristics of Turkey wind resources (Figure 3).
1.2.2 Turkey Solar Energy Potential

Turkey can be introduced eligible to produce photovoltaic voltage or solar power because its location is at North latitude between 36-42. In Turkey, Solar energy has been used for domestic hot water production in spite of the potential for electricity generation is considerable. However, Turkey, in the production of solar collectors for water heating has become the third largest manufacturer and the fourth largest the market situation in use in the world, the photovoltaic generation application is at an infant stage. The total photovoltaic generation capacity in Turkey is 6 MW, however the solar power potential is 300 TWh/year [5, 9].

The Turkey Atlas of Solar Energy Potential (GEPA), in order to determine the characteristics and the intensiveness of irradiance of solar energy resources was produced in
2010 by the General Directorate of EIE. This Atlas provides us with estimates of the amount of efficiency for solar power plants.

Figure 4: Solar map of Turkey

Average annual insolation time is total of 2640 hours and daily total of 7.2 hours per square meter. The average annual solar radiation is 1,311 kWh/m² per year and it is equal to 3.6 kWh/m² per day [10]. Although Turkey has good irradiance, they have not established large solar power plants except several small capacity plants with the power of 300W. In Turkey, the photovoltaic energy has been used for traffic signaling purposes in rural areas and monitoring towers or lighthouses.

To summarize the status of Turkey for renewable energy; in recent years, the production of electricity from wind power efforts has increased, although the production of electrical energy from solar prominence was not enough. Turkey's electricity while planning within this period, the change in prices should be taken into consideration. In Turkey, there is a great potential for solar energy. The costs of these energy sources can be covered when considerable energy need is
lifelong. Moreover, these sources are renewable and inexhaustible, and unlike conventional fuels constitute a significant threat to the environment and to human health.

Planning for renewable energy in electricity production in excess of 30% of the share which is now 25.2% is aimed by the Ministry of Energy and Natural Resource of Turkey. To reach this goal, MENR was promoted by installing new renewable energy resources. By installing new the solar power system and wind power system, and gathering reliable meteorological data sets of renewable energy resources, the efficiency of Turkey renewable energy resources’ output power can be increased.

1.3 Problem Statement and Formulation

The gap between energy consumption and production is increasing and global environmental pollution is becoming harsher because of the reduction of fossil energy. Therefore the application and improvement of renewable energy resources has been becoming more significant. It is now widely accepted that the renewable energy resources are very important for the future of Turkey. Electricity generation from solar and wind energy resource will become more appealing in renewable sources of energy.

The issue when increasing shares of renewable source in energy production is that controlling and determining efficiency of solar and wind energy plants is not clear. This kind of electricity generation is vague and depends on massive volume of meteorological variables like solar irradiance, air temperature, cloud insensitivity, relative humidity, wind speed, gust speed, and air pressure. The data-sets are by formed the vast amount of meteorological data. In addition, generally data-sets may have missing values. Accompanying efficiency with vagueness and having a high setup fee, renewable energy resources is backup plan for fossil fuels.
To solve this problem, the robust and reliable data sets of renewable energy resources should be developed and continuity and updating of these data sets should be provided. In addition, researching for improvement efficiency for renewable energy resources using data analytics tools should be realized to increase the share of wind and solar source in electricity generation. Data analytics tools such as the Artificial Neural network (ANN), Principal Component Analysis (PCA) and Genetic Algorithm (GA) help us to make decision on efficiency of renewable sources with analyzing and training large data-sets and checking accuracy of these nonlinear control systems.

1.4 Contributions of Thesis

Forecasting of renewable energy sources has two steps. The first step includes collecting meteorological data-sets like solar irradiance (SI), air temperature (T), relative humidity (RH), wind speed (WS), air pressure (P) and pre-processing which means treatment of missing data-sets in massive meteorological data-sets and preparing to be convenient for using in network using PCA.

The second is the forecasting of PV and wind power output with creating Artificial Neural Network which is used to develop forecasting models by using real climate data from State of Meteorology at Turkey. Also checking the accuracy of network is done. The forecasting network is evaluated with standard error metrics, RMSE to verify the accuracy of forecast. In the forth chapter of the thesis, using data analytics tools such as GA minimized errors in network performance by finding best weights for Artificial Neural Network (ANN).

Renewable energy resources with applying data analytics tools can provide more efficiency which means of more output power. Before installing plants, we have the chance to have information about how much power will be accommodated using learning methods of neural
networks like Back Propagation. It helps us to make economic decisions in terms of the efficiency of renewable resources.
CHAPTER TWO: DATA ANALYTICS TOOLS AND MODELING

2.1 Introduction

With the advances of establishing renewable energy resources, such as photovoltaic systems and wind power systems, the forecasted power production of these energy resources is done by researchers using data analytics tools. As one of the most popular forms of solar energy application, PV generation, including grid-connected PV plants and PV systems independent from utility, has increased rapidly around the world in recent years and will keep increasing in the future.

The accuracy of forecasting models is needed to check. It is possible to check accuracy by comparing experienced results and predicted results. Since the PV and wind turbine systems are considered vague meteorological variables, such as temperature, irradiance, wind speed, the results of forecasting models are vital for knowledge about electricity generation. There are many studies related to energy forecasting in the literature.

In recent research, artificial intelligence (AI) techniques are frequently used for energy forecasting. Banos (2011) mentioned that, renewable energy resources have many advantages; renewable energy resources need to control systems due to discontinuous generation because of climate changes [11]. Metaxiotis (2003) stated AI techniques, such as artificial neural networks (ANN) and genetic algorithms (GA), can be used for forecasting short term energy supplies [12]. The study of Mellit and Kalogirou (2008) involves a review of the artificial intelligence techniques relating to PV systems [13]. Bhattacharya and Basu (1993) used the Kalman filter for forecasting [14]. Sen (1998) utilized a fuzzy logic algorithm (FL) for estimating solar irradiation using sunshine duration measurements [15]. Iniyan (2006) created the Optimal Renewable Energy Mathematical (OREM) model to show this model can help for effective application of renewable energy resources [16].

These studies of Bakirci (2009) and Bilgili (2007) were presented a comprehensive review about successful applications of ANNs in renewable energy systems [20, 21]. Using artificial neural networks have proved its efficiency as an estimation tool for predicting factors through other input parameters, which have no any specified relationship. Also the capabilities of ANN methods provide more reasonable results. Their literature covers the methods applied in control of renewable energy systems.

In this thesis, Turkey’s net electricity energy generation and consumption, and solar irradiance and wind power capacity of Antalya which is the one of the main cities of Turkey, are forecasted. ANN is chosen as the forecasting tool. The reasons behind this choice are the ability of ANN to forecast future values of many variables simultaneously, and to model the nonlinear relation in the data structure. In addition, PCA is chosen for pre-processing in completing missing data and reducing dimensions of datasets to get a better performance. GA is chosen to improve network performance by finding the best weight for ANN.

2.2 Data Analytics Tools
Scientific data obtained in our environment is the essential step used to build scientific theories and models of the universe. Scientific data is increasing fast not only in terms of volume but also in terms of variety. Handling vast amount of data is needed to improve human analysis capabilities, so it becomes significant to manage and use database to increase efficiency, and provide more valuable knowledge for behavior of environment to research. Therefore the analysis of a massive volume of data is increasingly becoming central to scientific research. The data not analyzed is meaningless symbols. Between data sets, there is not any formula or rule to understand relationships. When data is interpreted using data analytics tools, it can be considered as information, which can help scientific research.

Analysis of big data can distinguish relevant information and extract useful knowledge from apparently unrelated data that is formed in a massive volume. This meaningful information can be used to forecast environmental behavior, which helps maximize the value of wind and solar output power estimations.

There is a crucial need for a new generation of computational tools to assist researchers in extracting useful knowledge from the rapidly growing volumes of digital data.

The data mining process can be divided into four steps.

1. Data preparation that is called pre-processing to prepare data for analysis,
   - In this study, PCA was used for pre-processing

2. Creation of a model, which can provide predictions using big data,
   - Artificial Neural Network is used for learning and prediction,

3. Verification of results using real data and evaluated data with standard error metrics know as Root Mean Square Error (RMSE)

4. Interpretation of results with graphical representations.
2.2.1 Artificial Neural Networks

An artificial neural network (ANN) is often called a neural network (NN). ANN had been used to refer to a network of biological neurons. Neural networks are means to explain brain functioning and to simulate human brain’s learning processes and organization [22]. While simulating the operations of biological neural systems, neural network use a general mathematical computing paradigm [23].

ANN appeared over seven decades ago in the literature with the study of McCulloch and Pitts (1943) designed the first neural network by cell model, called a Perceptron [24]. In their study the ANN demonstrated a set of logical statements and operators. Hebb (1949) concentrated on the learning ability of humans and its modeling which can be simulated in ANNs [25]. Hebb improved the learning algorithm. After 1980, improvements on the learning ability of ANN were focused [23]. In 1982 Hopfield studied especially the mathematical base of ANN for understanding the structure of networks [26]. Kohonen (1984) focused on regular arrays of neurons into the property unsupervised learning networks developed for characteristic mapping [27]. Rumelhart and McClelland (1986) established the back-propagation learning algorithms for multilayer networks that are capable to get results with more diverse features of inputs [13, 28].

ANN can be divided into two parts as structure and functional properties. Neural network structure includes interconnected neurons or processing elements as inputs, weights, bias, and output (See Figure 5). The functional properties determines to how the neural network learns using mathematical functions [23]. In neural network, two type functions are used. The first one is net function between inputs and weights/bias, and the second is activation function that use updated input to reach desired output.
Principally, inputs from other sources reach to biological neuron. The inputs are combined and then a nonlinear operation is applied on the inputs to get an output as a result. A typical ANN consists of three layers: the input layer, hidden layer, and output layer [19].

\[ \sum_j f(-) \]

Figure 5: Structure of a Neural Network

The summing function (net function) determines how the network inputs are combined inside the neuron. In this figure, \( m \) is the number of the input signals, \( x_j \) is an input signal (for \( j = 1, 2, \ldots, m \)), parameter \( w_j \) are known as weights. The quantity \( \theta_j \) is called the bias (or threshold) and is used to model the threshold.

\[ I_j = \sum_{j=1}^{m} w_j x_j + \theta_j \]

Then the products are summed, so the parameter \( I_j \) is output of net function is obtained.

Also, parameter \( I_j \) can be called updated network input with weight. Afterward, updated input passed through a nonlinear activation function to generate an output. The output of the neural
network, illustrated by $y_j$ in the figure below, is related to the network input $I_j$ passing through a linear or nonlinear transformation called the activation function:

$$y_j = f(I_j)$$

The activation function is an algebraic equation that can be linear or nonlinear. Both sigmoid (‘logsig’) and linear (‘purelin’) activation functions are commonly used functions to compute desired output from the knowledge of updated input in multilayer neural network. The derivatives of both sigmoid and linear tangent activation functions, which represents the output of network, can be computed by using the knowledge of $f(I_j)$ [29].

After the architecture and theory of neural network are examined, the learning ability of ANN can be understood better. ANNs are very important in control system especially nonlinear control applications because of their learning ability. ANN training process can be considered as a black-box modeling with a set of input factors and output variables [13]. The learning means that updating and determination of the synaptic weights. There are different connection styles and learning algorithms. Several architectures and algorithms of ANN have been developed for solving different problem in nonlinear systems [13]. Back-propagation algorithm (BP) and feed-forward network is popular in estimation technologies. Because of the self learning ability which means the characteristics of ANN can be trained, ANN can build a relationship between inputs and output from the historical data. This learned relationship can be used in the new prediction process [30].

2.2.1.1 Multi Layer Networks and Back propagation
Multi layer perceptron (MLP) is one of the most widely used ANN structure type for data mining, pattern recognition or forecasting. Multilayer perceptron (feed-forward) networks have only forward connections to subsequent layer in network, and hidden layer which provide to improve the learning capacity [13]. A diagram of typical multilayer feed forward neural network architecture is shown in Figure 6.

![Feed-Forward Neural Network](image)

**Figure 6:** Feed-Forward Neural Network [19]

This algorithm determines the weights that can achieve the best mapping between the input and the output [23]. Furthermore, the weights are adjusted randomly (between -1 and 1) so they do not represent the knowledge about input or output. But, after training is completed the weights contain meaningful information [31]. Back-propagation training is a gradient descent algorithm. It tries to improve the performance of the neural network by reducing the total error by changing the weights along its gradient. Also, this learning algorithm eliminates some of the disadvantages such as long time processing because it uses standard numerical optimization techniques. The error is expressed by the root-mean-square error (RMSE) value, which can be calculated by:
\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{\text{actual}} - P_{\text{predicted}})^2}
\]

\(P_{\text{predicted}}\) is output of neural network that is predicted solar irradiance \(P_{\text{actual}}\) is the measured solar irradiance power for desired time horizons and \(n\) is the number of dataset.

2.2.1.2 Forecasting Using ANN and its Accuracy

The output variables highly depends on the inputs, network produces outputs by setting correlations between them. So, the correlation among experienced data has a vital role to succeed better forecast. On the other hand, small prediction error can be reached with strong correlation. In addition, to achieve better prediction, all the information the network needs to learn is supplied to the network as a data set. While creating an ANN model, their performance should be verified for the problem. According to study of [32], eventually, the Levenberg–Marquardt, which is counted in backpropagation algorithm, is observed to produce more accurate results in the sense of lower MSE [32].

In this thesis, mean square error (MSE) metrics will be used to check error. The error will be checked by using the Levenberg–Marquardt training algorithm. When the error reached a tolerance level, the network keeps the weights constant to use the trained network to make prediction.

The following sections were designed to determine the data analytics tools such as GA and PCA. Solar energy forecasting will be achieved and the accuracy of network will be checked with MSE in Chapter three. Furthermore, GA and PCA will be used to get more accuracy.

2.2.2 Genetic Algorithm

In this section, we introduce the nature and structure of genetic algorithm (GA) and review on GA theory and practice. In the past decade, genetic algorithms have become as practical, robust
optimization and search methods. GA was first proposed by Holland in the 1960s and were developed by Holland and his students and colleagues at the University of Michigan in the 1970s. GA has improved for scientific research and engineering applications. [33]

We need to draw the similarity between genetic algorithms and the search processes in nature for optimal solution. Genetic algorithms are implemented as an optimum direct search technique based on principles of natural genetics and natural selection. [13] GA is a stochastic algorithm such as natural process of biological evolution. Furthermore, the GA is not like a mathematically guided algorithm. The optimal solution is developed from generation to generation without any mathematical formulation. GAs are a part of evolutionary computing, which is a rapidly growing area of artificial intelligence.

This model provides the possible solutions to the optimization problem as binary strings of 0s and 1s. In GA, a solution vector is called an individual or a chromosome which is made of discrete units called genes. Also, each gene is represented with binary digits, and controls one or more characteristics of the chromosome. The collection of chromosomes is called a population. The population contains solutions, and converges which means that it is dominated by a single solution.

A GA has the ability to solve complex problems was verified in [34, 35]. According to these studies, the application of genetic operators like selection, crossover, and mutation are used to modify and generate new a population. This process allows finding best chromosomes (solutions) and selecting the most appropriate offspring to transfer to the next generation. The life cycle of populations and the crossover and mutation are illustrated in Figure 7 [13].
Algorithm is started with a set of solutions which are represented by chromosomes and called population. The initial population of individuals is generated randomly. Second, the chromosomes are evaluated by a defined fitness function. Third is a selection which means that some of the chromosomes are selected to reproduce. Selecting parent chromosomes from a population according to their fitness (the better fitness) is achieved. The forth step includes mutation and crossover on the individuals whose fitness has just been measured to create new population from old one.
The crossover is the most important operation of GA. In the crossover, two chromosomes are combined together to create new chromosomes (offsprings). GA selects the parents in terms of fitness so that offspring is selected in good genes. It means the parents become fitter. By iteratively applying the crossover operator, the good chromosomes are expected to survive more frequently in the population, so finally it can reach convergence to a good solution when the fitness function can not longer be improved [33].

The next step, mutation process, is effective on characteristics of chromosomes. The mutation does not make differences in population as much as crossover. The mutation is applied to each offspring individually after the crossover process. The crossover makes the population to converge by producing the chromosomes alike the old chromosomes. While leading convergence to find best solution, there is need for genetic diversity to increase and produce new good genes. The mutation

Figure 8: Process of Genetic Algorithm
may cause delay to find global optima but it prevents to find local optima instead of global optima [33].

The produced offspring is became the initial population by replacing their parents. In this reproduction process, only the selected parents in the third step will be replaced by their corresponding offspring. The important point is that each generation individuals are selected according to their performance with respect to the fitness function for generating new population. This selection provides a higher chance of survival to better individuals; it provides a rapid convergence to a near optimum solution in many types of problems.

This GA process repeats until the end condition is satisfied. It means the genetic algorithm processes are applied to form new and better offspring. The algorithm is terminated when an indicated number of chromosome generations or the optimal solution has been found.

By taking into consideration the feature of GA that finding optimal solution by calculating global optima, GA can be use to find optimal weights in ANN to get more accurate prediction results.

### 2.2.3 Principal Component Analysis

The Principal Component Analysis (PCA) is one of the most widely used statistical techniques in data analytics applications. PCA provides a roadmap for how to reduce a complex data set to a lower dimension by keeping the statistical feature of dataset [37]. This reduction means that get scientific knowledge from meaningless set of large data in a reasonable time. PCA reduces the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variations present in the original dataset. The reduction of dimension is realized by identifying the maximum value that called principal component in the dataset. Using fewer components, each sample can be represented with relatively small number as
an alternative by thousands of variables [38]. Also, PCA gives us the opportunity to work for applications which have suitable linear models in linear domain such as control theory.

To clarify the process of Principal Components Analysis, we should to determine the mathematical background of PCA. When we consider the mathematical background, there are two sections we will be required to know. The first one is statistics which looks into how the data is spread out. In the many research area, we have huge dataset, and the dataset is needed to analyze the in terms of relationship between different points in this data set, so the statistics tools are the best solution to analyze. In statistics section there are three types of measurement methods such as Standard Deviation, Variance and Covariance [39].

The second one is the matrix algebra that provides a background for the matrix algebra required in PCA such as eigenvectors and eigenvalues of a given matrix. The eigenvectors and eigenvalues are important properties of matrices that are fundamental for PCA to calculate best principal components which represents better others. The next paragraph will continue with a short review of the PCA process.

The first process is subtracting the mean from each of the data dimensions since the mean subtracted is the average across each dimension and calculating covariance matrix to obtain eigenvalues and eigenvectors. This process help us to be able to extract lines characterize the data. Second, determining components to form a feature vector provides us that with components can be ignored.

Once eigenvectors are found from the covariance matrix, the next step is to put them in order by eigenvalues and significance.PCA applies this process and ignores the components of lesser significance. The significance can be determined by take into consideration the affect of eigenvalues on eigenvectors. The final data set will have fewer dimensions than the original.
According to Jackson (1991), there are many ways to determine the optimum dimension of dataset [40]. One of the widely adopted rules is checking the amount of explained and unexplained variability. In this rule, eigenvalues and eigenvectors are obtained until the amount of unexplained variability, or the remaining, has been reduced to a desired number [40]. Individual variance explained by each principal component and cumulative variance can be calculated to help determine the number of PC that should be retained. The advantage of PCA is its potential ability to represent an n-variable data set in a smaller dimension (< n). The important point is that, however, there is no standard rule for determining the number of PCs to remain in a PCA model; it has the ability to recognize the principal components, which are linear combinations of the original variables. So, dimension reduction capability of PCA makes it fundamental in data mining activities.

The mission of PCA in this study is reducing the massive of dataset which is used for forecasting. Reducing dimension of dataset has advantages such as less time and less memory to forecasting of energy parameters.

2.3 The Status of Renewable Energy Forecast in Turkey

In Turkey, forecast of energy has been done by MENR. Forecasting of PV and wind power is a trend topic in the world because of limited fossil fuels and environmental pollution. In previous section several studies on Turkey renewable energy forecast are mentioned. The share of Turkey’s wind power systems is 2.2%, and PV systems are less than 1% in production electricity, however the capacity of solar irradiance and wind is high [9]. Turkey needs a comprehensive forecast model to increase the share of renewable energy resources. In addition; Turkey should provide forecast results for all possible location for PV and wind systems. These needs can be provided by collecting more data relevant PV and wind systems, and by getting better performance in
forecasting model. Collecting more data and forecasting energy with using this data mean more cost. Data analytic tools can reduce this cost by decreasing a memory needs to collect massive amount of data and forecast with lower error using these datasets.

2.4 The Approach of Implementation

The main objective of this study was to estimate the solar irradiance building the artificial neural network using meteorological dataset such as air temperature (T), relative humidity (RH), wind speed (WS), air pressure (P), and cloud base height(CH). In the study, data obtained by the Turkish State of Meteorological Service to train and test the model. According to Jiang (2008), the ANN performs more suitable to predict solar radiation compare to other regression models [41].

Various training algorithms are available for the training of neural networks. In chapter 3, ANN is created in Matlab software by using the Levenberg–Marquardt backpropagation algorithm. ANN’s structures and properties will be tried and built to obtain more accurate results for output power prediction. Mean square error (MSE) and Regression values (correlation coefficients) are calculated to illustrate the performance of the network [42].

In chapter 4, the optimized algorithm of incorporating the GA and PCA optimization with neural network is used, which incorporates with artificial neural network, and improves the learning ability of neural network to get a more precise performance estimate results. This optimized algorithm model solves the problem of artificial neural network’s learning ability, which has a high dependency of the network initial weights. This model improves the learning ability of the original model with the advantages of GA and PCA [30].

CHAPTER THREE: SOLAR FORECASTING USING NEURAL NETWORK

3.1 PV and Wind Databases
The Photovoltaic (PV) and Wind Power input data are comprehensive and complex. Data from different sources was combined to get more variables and better results.

The weather and photovoltaic datasets are recorded by the Turkish State Meteorological Service [43]. Also, these dataset include hourly recorded solar conditions and environmental parameters between 2012 and 2013.

The second source of data was obtained from the Iowa Environmental Mesonet [44]. Their Automated Surface Observing System (ASOS) has recorded weather data sets which provide a chance for estimation for wind power.

The azimuth and zenith angles are obtained from solar position calculator by Solar TOPO [45].

Solar irradiance, air temperature, wind speed, cloud altitude and air pressure are quantified by Watt-hours per square meter, Celsius degree, miles per second, feet, Pascal, respectively. Hourly one day meteorological dataset is given in the Table 3.
Table 3: Hourly meteorological dataset Antalya/Turkey

<table>
<thead>
<tr>
<th>Date</th>
<th>Hours</th>
<th>Air Pressure (hPa)</th>
<th>Temp. (°C)</th>
<th>Relative Humidity (%)</th>
<th>Cloud Altitude (Ft)</th>
<th>Wind Speed (m/sec)</th>
<th>Solar Irradiance (Watt/m²)</th>
<th>Cloud Coverage</th>
<th>Zenith Angle</th>
<th>Azimuth Angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/15/2013 0:00</td>
<td>1001.80</td>
<td>12.70</td>
<td>93.00</td>
<td>1500.00</td>
<td>2.20</td>
<td>35.20</td>
<td>6.00</td>
<td>-61.07</td>
<td>321.92</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 1:00</td>
<td>1001.60</td>
<td>11.80</td>
<td>93.00</td>
<td>1500.00</td>
<td>2.60</td>
<td>67.80</td>
<td>6.00</td>
<td>-65.72</td>
<td>353.32</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 2:00</td>
<td>1001.50</td>
<td>11.30</td>
<td>95.00</td>
<td>1500.00</td>
<td>1.90</td>
<td>73.30</td>
<td>6.00</td>
<td>-63.48</td>
<td>27.48</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 3:00</td>
<td>1001.90</td>
<td>10.90</td>
<td>91.00</td>
<td>1500.00</td>
<td>2.20</td>
<td>39.10</td>
<td>6.00</td>
<td>-55.70</td>
<td>52.28</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 4:00</td>
<td>1002.00</td>
<td>11.30</td>
<td>91.00</td>
<td>1500.00</td>
<td>2.00</td>
<td>245.70</td>
<td>6.00</td>
<td>-45.23</td>
<td>68.50</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 5:00</td>
<td>1002.70</td>
<td>11.80</td>
<td>89.00</td>
<td>1500.00</td>
<td>2.00</td>
<td>1749.50</td>
<td>6.00</td>
<td>-33.68</td>
<td>80.19</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 6:00</td>
<td>1003.10</td>
<td>12.70</td>
<td>90.00</td>
<td>1500.00</td>
<td>1.70</td>
<td>5877.30</td>
<td>6.00</td>
<td>-21.72</td>
<td>89.76</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 7:00</td>
<td>1003.30</td>
<td>12.60</td>
<td>89.00</td>
<td>1500.00</td>
<td>2.40</td>
<td>2899.30</td>
<td>7.00</td>
<td>-9.76</td>
<td>98.54</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 8:00</td>
<td>1003.60</td>
<td>11.70</td>
<td>70.00</td>
<td>1500.00</td>
<td>1.00</td>
<td>6679.60</td>
<td>6.00</td>
<td>2.22</td>
<td>107.38</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 9:00</td>
<td>1003.40</td>
<td>11.30</td>
<td>71.00</td>
<td>1500.00</td>
<td>2.00</td>
<td>32233.90</td>
<td>7.00</td>
<td>13.12</td>
<td>117.00</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 10:00</td>
<td>1003.30</td>
<td>10.00</td>
<td>64.00</td>
<td>2500.00</td>
<td>1.80</td>
<td>30495.10</td>
<td>6.00</td>
<td>23.22</td>
<td>128.14</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 11:00</td>
<td>1003.40</td>
<td>10.30</td>
<td>78.00</td>
<td>1500.00</td>
<td>2.10</td>
<td>11280.90</td>
<td>5.00</td>
<td>31.74</td>
<td>141.57</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 12:00</td>
<td>1003.90</td>
<td>10.10</td>
<td>86.00</td>
<td>1500.00</td>
<td>1.00</td>
<td>6815.30</td>
<td>6.00</td>
<td>37.83</td>
<td>157.81</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 13:00</td>
<td>1004.30</td>
<td>10.10</td>
<td>86.00</td>
<td>1500.00</td>
<td>2.40</td>
<td>7876.90</td>
<td>6.00</td>
<td>40.53</td>
<td>176.40</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 14:00</td>
<td>1004.40</td>
<td>10.30</td>
<td>83.00</td>
<td>1500.00</td>
<td>0.80</td>
<td>1729.80</td>
<td>6.00</td>
<td>39.30</td>
<td>195.45</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 15:00</td>
<td>1005.00</td>
<td>9.80</td>
<td>84.00</td>
<td>3000.00</td>
<td>2.00</td>
<td>181.30</td>
<td>6.00</td>
<td>34.40</td>
<td>212.73</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 16:00</td>
<td>1005.80</td>
<td>10.30</td>
<td>84.00</td>
<td>3000.00</td>
<td>1.60</td>
<td>233.90</td>
<td>4.00</td>
<td>26.69</td>
<td>227.20</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 17:00</td>
<td>1006.20</td>
<td>9.60</td>
<td>89.00</td>
<td>3000.00</td>
<td>2.00</td>
<td>37.70</td>
<td>4.00</td>
<td>17.11</td>
<td>239.14</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 18:00</td>
<td>1006.40</td>
<td>9.30</td>
<td>89.00</td>
<td>3000.00</td>
<td>1.80</td>
<td>41.60</td>
<td>1.00</td>
<td>6.40</td>
<td>249.27</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 19:00</td>
<td>1006.60</td>
<td>9.10</td>
<td>94.00</td>
<td>3000.00</td>
<td>1.40</td>
<td>21.00</td>
<td>2.00</td>
<td>-5.24</td>
<td>258.34</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 20:00</td>
<td>1006.80</td>
<td>9.10</td>
<td>93.00</td>
<td>3000.00</td>
<td>1.90</td>
<td>16.30</td>
<td>0.00</td>
<td>-17.13</td>
<td>267.08</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 21:00</td>
<td>1007.00</td>
<td>9.40</td>
<td>96.00</td>
<td>3000.00</td>
<td>1.60</td>
<td>40.10</td>
<td>0.00</td>
<td>-29.10</td>
<td>276.28</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 22:00</td>
<td>1007.40</td>
<td>9.80</td>
<td>96.00</td>
<td>3000.00</td>
<td>1.10</td>
<td>41.70</td>
<td>0.00</td>
<td>-40.84</td>
<td>287.02</td>
<td></td>
</tr>
<tr>
<td>2/15/2013 23:00</td>
<td>1007.70</td>
<td>10.50</td>
<td>96.00</td>
<td>3000.00</td>
<td>1.90</td>
<td>45.50</td>
<td>0.00</td>
<td>-51.79</td>
<td>301.22</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Implementation Model

All simulations made in this work were performed using MATLAB from MathWorks.

MATLAB is an efficient tool for this type of simulation as it can combine toolboxes in ANNs and optimization.

Our main goal was to develop a model of artificial neural networks capable of:

- Learning power output in a reasonable amount of time
- Performing with certain accuracy for all future load conditions

The first objective was accomplished by architecting of a neural network using an efficient training algorithm, that is related with complication of the forecasting problem, and selection of correct input variables.

The second requirement was achieved by the selection of useful part of representative training data in dataset and finding the best percentages of training and test data by the forecasting model.

Therefore, the implementation of the forecasting system using neural network required carrying out several tasks:

- Selection of input variables
- Neural networks structure
- Mining of training data and validation of the designed networks

### 3.2.1 Selection input variables

The first step of the creation model to data analytics is identifying relevant inputs which are needed for data analytics tool. Using all variables in the data analytics will be not useful because of process time and more memory requirement. So, selection of the set of the input variables that are significantly should be correlated with the output variable. Generally, in mentioned papers, the power output of PV was based on solar irradiance. In addition to these, papers related with load prediction have been taken by a lot of authors such as Drezga and Rahman (1999) and Senjyu (2002). [46, 47].
Furthermore, we need to determine a relationship between inputs and output in while choosing input variables to get better results. The network may perform poorly when applied to unrelated data.

Figure 9 shows us to the variability of the solar irradiance and air temperature from February 1\textsuperscript{st} to February 28\textsuperscript{th} 2013. The main point is that correlation between the solar irradiance and air temperature should be the same characteristic through all dataset. It is clear the respond of solar irradiance to air temperature is in harmony. The shapes do not have to be same completely.

![Figure 9: Correlations between Air Temperature and Solar Irradiance](image)

There can be a harmony between inputs and output, however the correlation opposite of each other. When relative humidity is increasing, solar irradiance will decrease. Figure 10 represents opposite correlation between relative humidity and solar irradiance.

![Figure 10: Correlations between Relative Humidity and Solar Irradiance](image)
Consequently, inputs such as relative humidity, air pressure, wind speed, cloud coverage, and air temperature have a harmony with solar radiation in terms of forecasting. The number of significant input variables is chosen by determining the correlation between solar irradiance.

3.2.2 Neural networks structure

A learning algorithm is usually used to update the parameters of the network architecture. To get better forecasting result, training should be done better. Better training can be achieved by setting better structure of network. It is often difficult to decide how complex a structure is actually necessary for the desired control. Moreover, the designer cannot express an opinion about operating of the networks unless see results.

The key issue in optimal network design is the selection of the number of hidden layers and the number of neurons per layer. The hidden-layer neurons determine a network's learning capabilities. A network with too few hidden neurons will not be capable of accurately modeling load. On the other hand, too many hidden layer neurons may not help to get more accurate model, and unnecessary loss of computational time.

In this thesis, Feed-Forward Network with BP is used. To make sure about the best amount of hidden layer and its neuron, several choices are tried and checked in terms of RMSE/MSE value and Regression(R). Figure 11 shows a representation of the feed-forward neural network generated when training using 8 inputs variables and 1 hidden layer consisted of 10 neurons. To set better structure for ANN, several numbers of neurons will be tried and discussed the accuracy. Solar irradiance and wind speed are chosen as outputs which can help us to have an idea about the output power.
Figure 11: The view of structure of a neural network

The networks to forecast solar irradiance have 8 inputs such as air temperature(T), air pressure(P), relative humidity(RH), wind speed(WS), cloud base altitude(A), cloud coverage(CC), azimuth angle, and zenith angle.

Figures 12-17 presents the results with different number of hidden neurons per one hidden layer in ANN while predicting the PV and wind output power.

Figure 12: Artificial Neural Network with 1 hidden neuron in a hidden layer (SI)
Figure 13: Artificial Neural Network with 5 hidden neurons in a hidden layer (SI)

Figure 14: Artificial Neural Network with 10 hidden neurons in a hidden layer (SI)
The network created to predict wind speed (WS) used inputs such as, (P), (T), (RH), (A), and (SI).

**Figure 15:** Artificial Neural Network with 20 hidden neurons in a hidden layer (SI)

**Figure 16:** Artificial Neural Network with 5 hidden neurons in a hidden layer (WS)
Figure 17: Artificial Neural Network with 20 hidden neurons in a hidden layer (WS)

Table 4 shows the effects of numbers of neuron by comparing each network with values of MSE and Regression

<table>
<thead>
<tr>
<th>Number of Neurons</th>
<th>MSE</th>
<th>Regression (R)</th>
<th>Epoch (consideration of Time)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind Speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13.37</td>
<td>0.602</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>4.02</td>
<td>0.85</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>3.95</td>
<td>0.802</td>
<td>13</td>
</tr>
<tr>
<td>20</td>
<td>3.98</td>
<td>0.8507</td>
<td>9</td>
</tr>
<tr>
<td>Solar Irradiance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.30E+07</td>
<td>0.8184</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>5.00E+07</td>
<td>0.9478</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>1.70E+07</td>
<td>0.9668</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>5.00E+07</td>
<td>0.9348</td>
<td>16</td>
</tr>
</tbody>
</table>

Finally, the number of hidden layer’s hidden neurons has a significant effect on forecasting networks. In this study 10 neurons hidden layer is used to forecasting of photovoltaic power and wind power.
3.2.3 Mining of training data and validation of the designed networks

The neural networks are trained in the supervised manner for relevant input and the actual output. When neural network achieve desired output, weights and bias adapted to minimize an error function that calculate the differences between the desired output and the computed output by the network.

Out of several existing training algorithms, the Levenberg-Marquardt algorithm was selected due to its effectiveness [13]. This algorithm uses both gradient-descent and Gauss-Newton methods, so this algorithm can make linear combinations and adapt rules. The Levenherg-Marquardt method is a deterministic optimization method used to find a local minimum of the error function.

Training data should satisfy the fundamental requirement of having all possible charge conditions equally represented. This provides us to obtain similar performance of the forecasting for all future output power. The proposed formulation of the forecasting problem lends itself to meet this requirement.

Training data are gained from using part of available historical solar irradiance and meteorological data. Once we set the ANN’s training structure and parameters for one week or one day, we do not need to set again for another day or week power prediction.

Figures 18-20 presents the results of ANN with different rates of training, testing and validation. The best rates can provide best performance to train networks.
**Figure 18:** The training data is the % 70 percentages of dataset (SI)

**Figure 19:** The training data is the % 50 percentages of dataset (SI)
Table 5: Effects of training percentage on the performance of ANN

<table>
<thead>
<tr>
<th>Percentage of Training, Test, and Validation</th>
<th>MSE</th>
<th>Regression</th>
<th>Epoch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training %30, Test %50, Validation %20</td>
<td>6.8E+07</td>
<td>0.6046</td>
<td>10</td>
</tr>
<tr>
<td>Training %50, Test %30, Validation %20</td>
<td>7.9E+07</td>
<td>0.8253</td>
<td>11</td>
</tr>
<tr>
<td>Training %70, Test %15, Validation %15</td>
<td>1.70E+07</td>
<td>0.9668</td>
<td>15</td>
</tr>
</tbody>
</table>

Finally, according to results, the optimum performance could be obtained by using ANN with 10 neurons per 1 layer and 70% of training data.

3.4 Applying one year dataset to ANN

While using one week dataset, the creating optimum ANN is achieved. To spread of the photovoltaic prediction to future, the simulation was achieved again with hourly one year dataset of 2013. The accuracy of network is decreasing when apply one year dataset compare to one week,
however network has still same settings and structure that used with hourly one week dataset in above section.

Figure 21 shows one year prediction, there is some broken line and unsuitable differences between actual and predicted output. These problems can be emerged since big data needs more preprocessing or weights of ANN can be change more compare of smaller dataset. Data analytics tools allow solving these problems. In next chapter, several data analytics tools will be determined.

![Figure 21: One year solar irradiance prediction results](image)

3.3 Conclusions

Artificial neural network with the Levenberg–Marquardt back propagation method is built and trained in several choices to get better results. The mean squared error that is called as the error function is used to check accuracy. The minimizing MSE was achieved with using mentioned settings in the chapter three such as improving network structure and choosing suitable parameters for network. The 10 neurons per hidden layer and 70% training data percentages is used to create
neural network. The created ANN provide us hourly prediction of solar irradiance indirectly solar systems output power using meteorological and geopolitical datasets. When ANN is used with only one week divide hourly dataset, the best performance value MSE and R were 1.7E+07 and 0.9668. But when we applied all year dataset to network, the best performance value (MSE) and regression value (R) are obtained with less accuracy like 19.58E07 and 0.6687.

In chapter 4, several tools such as GA and PCA will be used to improve accuracy of ANN. Also, the hybrid systems which include more than one tool is created to get better performance. GA will be used for optimize weights of ANN. PCA will be applied dataset for reducing dimension of massive dataset like pre-processing tools.
CHAPTER FOUR: OPTIMIZATION THE PERFORMANCE OF NEURAL NETWORK WITH USING GA AND PCA

4.1 GA-ANN hybrid systems

Most of the research on GA-ANN systems has been studied on the utilization of the GA to improve the design and accuracy of networks by finding out the best architectures in terms of connectivity configuration such as weights and bias. A Neural network’s training algorithm allocates random numbers to weights and biases at the beginning of the training. Then, the training algorithm tries to minimize the objective function in each iteration based on assigned numbers for initial weights and updates the weights accordingly. ANN learns by checking error and updating connectivity parameters [13]. The GA is called a global search tools that focus on global optima to find best solution. Also, it provides the opportunity to search in large population through producing constantly new population (set of solution) [48]. ANN is a capable tool for classification. ANN uses gradient search techniques such as the backpropagation algorithms (see section 2.2.1) to obtain weights for building better structure of ANN to provide better results in prediction problems [49]. But, finding minimum MSE sometimes need to more iteration which cause wasting time. In addition, ANN works with unpredictable and not hundred percent certain performance. In this section, Genetic Algorithm (GA) is applied to optimize the architecture of ANN.

4.1.2 Process of GA-ANN hybrid system

GA-ANN hybrid system improves the accuracy of solution (output of network) with arrangement of inputs and network parameters. To more clarify how GA improves the network performance, the feature of learning algorithm should be mentioned. The backpropagation learning is a local optima search algorithm, and GA works by finding global optima. ANN use local optima
search to improve the possibility for implementation complex nonlinear systems. But the global optima search algorithm gives better specific solution in compare to local optima search algorithm. This can be called an adjustment that is meant to restructure the ANN structure. After this adjustment backpropagation algorithm can find the global solution.

First, we need problem definition, which is getting better accuracy for results in network and consistent performance. Subsequently, selection of the objective function to minimize MSE is applied. The objective function helps us determine which individuals (each weights) give the best output (MSE). In the most of GA works for ANN, binary representation of the weights is preferred as initial values to create a population to the best solution in GA [19]. In this study, the real value of weights from ANN is used to produce new population with mutation and crossover to reach the best population (set of weights) for getting lower MSE. The objection function is MSE calculating which use real weight which are obtained trained network. Then, GA reproduces the weights by crossover and mutation process. In GA, the population size is chosen to be 10, and tolerance value which is decision parameter to stop algorithm producing and iteration number is determined. Finally, after GA reaches maximum evolution number or tolerance value, the best weights in according to the best MSE can be found. The next step is using the weights which are derived from GA in ANN and checking the accuracy of ANN in term of MSE. To understand the additives of GA for prediction of renewable energy outputs with the lower differences between desired output and network output, MSE and Regression value (R) will be in consideration.

4.3 Results of GA-ANN
Chapter 3, ANN was used to create model to predict renewable energy resources using meteorological dataset provided by Turkish State Meteorological Service. The prediction model of weekly, monthly and yearly are performed using hourly recorded dataset. After evaluate the results, yearly prediction model performed lesser accuracy compare to the weekly and monthly prediction. So, one of the common data analytic tools that is GA is used to minimize error and optimize the architecture of ANN. The results will be evaluated by figure which is included solar irradiance prediction accuracy with examine desired output and network output. In hourly divided, one week, one month and one year solar prediction will be shown below figures 22-27. Next, the MSE and R values given in the table will be discussed.

Figure 22: One week hourly solar irradiance prediction by using only ANN
Figure 23: One week hourly solar irradiance prediction by using GA-ANN hybrid model

Figure 24: One month hourly solar irradiance prediction by using only ANN
Figure 25: One month hourly solar irradiance prediction by using GA-ANN hybrid model

Figure 26: One year hourly solar irradiance prediction by using only ANN hybrid
**Figure 27:** One year hourly solar irradiance prediction by using GA-ANN hybrid model

**Table 6:** The comparing results of ANN and GA optimized ANN

<table>
<thead>
<tr>
<th></th>
<th>One Week Hourly Dataset</th>
<th>One Month Hourly Dataset</th>
<th>One Year Hourly Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANN</td>
<td>GA-ANN</td>
<td>ANN</td>
</tr>
<tr>
<td>MSE (better is lower)</td>
<td>5.97E07</td>
<td>3.88E07</td>
<td>5.26E07</td>
</tr>
<tr>
<td>R (better is more close to 1)</td>
<td>0.7169</td>
<td>0.9344</td>
<td>0.7936</td>
</tr>
</tbody>
</table>

According to results, the MSE value decreased from 19.58E07 to 16.30E07 in one year solar irradiance prediction model; this means the performance of ANN is increased 16.75% by
optimizing weights with GA. Also, after obtaining the best weights, the results of ANN can be consistent by using best weights for each iteration. ANN needs to retrain the network in each iteration while finding the best solution, but GA does not need to train network for each iteration and uses weights which is obtained after first iteration of ANN.

4.2 Preprocessing using PCA

Increasing the number of inputs generally provide better performance unless the missing value in dataset is increased. More input variables may cause lower performance because of missing values. The most important issue on larger dataset is handling cost and time. Also, there can be out of memory error due to large dataset. Implementation of large dataset needs more powerful computer features. At this point, data analytic tools become more essential to get information by deriving large dataset which include any relationships rule between variables.

PCA applied to dataset is like a filter to compress the value massive number of variables to a smaller dataset. PCA is used for dimension reduction for large meteorological dataset. First, PCA calculates the eigenvalues and create eigenvalue vector, then eigenvalues becomes the principal components (PCs). Subsequently, PCA sorts the eigenvalues from highest to lowest. The eigenvalues with higher value have most information.

4.2.1 Results of PCA for preprocessing dataset

The one month solar irradiance dataset which has 672x8 variable matrixes, the dimension is reduced to 200 from 672. The one year dataset is reduced to 1000x8 matrix format from 8560x7 matrix format. The dimension reduced dataset applied the ANN and compared with actual size dataset results. Figure 28 shows performance of ANN using PCA to reduce dimension to 200 from 672.
Figure 28: One month hourly solar irradiance prediction by using reduced dataset (200)

Table 7: The comparing the results of ANN and PCA-ANN

<table>
<thead>
<tr>
<th></th>
<th>One Month Hourly Dataset</th>
<th>One Year Hourly Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANN (672)</td>
<td>PCA-ANN (200)</td>
</tr>
<tr>
<td></td>
<td>ANN (8560)</td>
<td>PCA-ANN (1000)</td>
</tr>
<tr>
<td>MSE (better is lower)</td>
<td>5.26E07</td>
<td>4.76E07</td>
</tr>
<tr>
<td></td>
<td>19.58E07</td>
<td>14.55E07</td>
</tr>
<tr>
<td>R (better is more close 1)</td>
<td>0.7936</td>
<td>0.7176</td>
</tr>
<tr>
<td></td>
<td>0.6687</td>
<td>0.5889</td>
</tr>
</tbody>
</table>

According to results which are represented in table 7, PCA can handle the information of value while minimizing the dimension of dataset. However, regression value is decreased and characteristics of figure which represents prediction and actual value is not as good as figure that
is obtained only ANN, the MSE value is decreased 25% in one year data and 10% for one month dataset. To make prediction by using ANN with massive volumes of dataset, PCA has ability to handle minimizing dataset to use it in ANN.
CHAPTER FIVE: CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

In this thesis, we addressed the problem of handling and getting information from a large meteorological dataset to predict renewable energy resources, such as PV and wind power systems. In Turkey, the measurement of data for the prediction of renewable energy output is available for very a limited area, however, Turkey’s location is convenient to produce PV and wind power. In order to solve this problem, the data analytic tools use AI techniques, such as ANN, PCA and GA. It is well known that solar and wind power, depending on meteorological conditions, are intermittent in nature and reliable prediction means.

The first step was building an ANN, which used a learning algorithm (BP), and setting up the inputs related with output to make predictions. The inputs such as air temperature (T), air pressure (P), relative humidity (RH), wind speed (WS), cloud base altitude (A), cloud coverage (CC), azimuth angle, and zenith angle are used to forecast solar irradiance and wind speed. Before the training process, we checked the correlations between inputs and output. Solar irradiance and wind speed can be used to get information about output of PV and wind power systems. Finally, after the architecture of an ANN is ready to train, the network is trained and tested for validation. The results were reasonable and showed that renewable energy outputs can be predicted using an ANN.

The second step is focused on improving the efficiency of ANN using GA and PCA tools. GA provides consistent results, not like ANN which gives results with small differences for each iteration. PCA gives us a chance to handle large dataset which is impossible to apply due to the limited of most computer memories. The results express that the performance of ANN is improved. Also, data analytics tools can be used to save time and memory for the prediction of energy.
The share of Turkey’s wind power systems is less than 3%, and PV systems are less than 1% in production electricity [9]. The average annual solar radiation is 1,311 kWh/m² per year and it is equal to 3.6 kWh/m² per day [10]. Turkey is beginning to build PV and wind power systems. Turkey plans to install PV plants, which have no widespread solar measurements, and need data analytics tools. Applying data analytics tools studied in this thesis can provide more efficiency with contributions of an energy forecasting model.

5.2 Future Work

The results show that ANN can be used for renewable energy prediction. In addition, the simulation results realized in this thesis show that GA and PCA could increase the performance of ANN. Future research can include the optimized ANN with other data analytic tools, such as a Fuzzy logic controller. Another improvement to energy forecasts with ANN could be applying different kinds of training algorithms for neural networks. ANN eliminates the missing value, so future research can contain an approach to replace the missing value with the average of neighboring points in the dataset.

The simulations do not take into consideration inverter switching losses while forecasting PV and wind power. After inverter switching loss had been forecasted, the output power could be estimated more accurately.

Finally, the financial feasibility of solar and wind power systems for each location can be calculated using results of the forecasting model. Also, the interface which provides the amount of power by renewable energy resources can be created for each user of residential PV systems and residential wind turbines. It can encourage people to build residential PV systems and wind turbines to cover their energy consumptions.
APPENDIX A

Code1 ANN for prediction

clc;
clear all;

load('datafebruary.mat');
t=[data(:,7)]; %Target
x=[ data(:,5),data(:,2),data(:,3), data(:,4),data(:,6 ) ]'; % Inputs
net = feedforwardnet(10); % one hidden layer with three neurons;
feedforwardnet(hiddenSizes,trainFcn)
% net.numLayers = 3;
% net.numInputs = 2;
net= configure(net,x,t);
net.trainParam.epochs= 1000;
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
net.trainParam.goal = 1e-5;

net.biasconnect= [1;0]; %biases for each layers : hidden and output
net = init(net);
net.trainParam.lr=0.1 % default = 0.01
net.trainParam.mc = 0.9; %momentum

[net,tr] = train(net,x,t); % training

outputs=net(x);
errors = gsubtract(outputs,t);

%view the results
P = sim(net,x);
figure
plot(A,'b-')
hold on
plot(t,'r')
grid on
view (net)
figure, plotperform(tr)
figure, plottrainstate(tr)
figure, ploterrhist(errors)
% INITIALIZE THE NEURAL NETWORK PROBLEM %
load('datafeb-june.mat');
targets=[data(:,7)]'; %Target inputs=[data(:,2),data(:,3),data(:,4),data(:,5),data(:,6)]'; % X = inputs % number of neurons n = 10;
% create a neural network net = feedforwardnet(n);
% net.trainParam.goal = 1e-5;
%net.dividefcn='divideblock';
% configure the neural network for this dataset net = configure(net, inputs, targets);
view(net)
num_weig = net.numWeightElements;

% create handle to the MSE_TEST function, that calculates MSE h = @(x) mse_test(x, net, inputs, targets);

% Setting the Genetic Algorithms parameters ga_opts = gaoptimset('PopulationSize',10,'Generations',1000,'TolFun', 1e-8,'display','iter');

[x,fval,exitflag,output,population,scores] = ga(h, num_weig, ga_opts);

% To set the weights and biases vector net = setwb(net, x');
net.trainParam.epochs= 1000;
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
net.trainParam.lr=0.1 % default = 0.01
net.trainParam.mc = 0.9; %momentum
net.layers{1}.transferFcn = 'logsig';
net.layers{2}.transferFcn = 'purelin';
%To evaluate the outputs based on the given weights and biases vector y = net(inputs);

% Calculating the mean squared error mse_calc = sum((y-targets).^2)/length(y)
fprintf('MSE: %u\n', mse_calc)
% PCA processing
load('datafebruary.mat');
d=200 % desired dimension to reduce
data=data(:,1:7)';
[V,D]=eig(cov(data));
eigValue=diag(D);
[eigValue,IX]=sort(eigValue,'descend');
eigVector=V(:,IX);
% rebuild matrix
norm_eigVector=sqrt(sum(eigVector.^2));
eigVector=eigVector./repmat(norm_eigVector,size(eigVector,1),1);

% dimension reduction
Y=eigVector(1:d,:); % data';
indices=find(Y<0); % find the elements of Y, which are negative
because of pca process
Y(indices)=Y(indices)*-1;
Y=Y(:,:);
t=[Y(:,7)]';
x=[Y(:,5),Y(:,2),Y(:,3), Y(:,4), Y(:,6)]';
% set the parameters
net.trainParam.lr=0.1 % default = 0.01
net.trainParam.mc = 0.9; % momentum
net = feedforwardnet(10); % one hidden layer with ten neurons;
net= configure(net,x,t);
net.trainParam.epochs= 1000;
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
net.trainParam.goal = 1e-5;
net.biasConnect= [1;0]; % biases for each layers: hidden and output
net = init(net);
[net,tr] = train(net,x,t);
outputs=net(x);
errors = gsubtract(outputs,t);

% view the results
P = sim(net,x);
figure
plot(A,'b')
hold on
plot(t,'r')
grid on
legend('Predicted','Actual');
title('Simulated Data Set');
ylabel('Solar Irradiance');
xlabel('Hours in February 2013')
view (net)
figure, plotperform(tr)
figure, plottrainstate(tr)
figure, ploterrhist(errors)
REFERENCES


[43] Turkish State Meteorological Service, Ankara/Turkey.


VITA
Halid Kaplan was born in Turkey and received his B.S. degree in Electrical and Electronics Engineering from Kahramanmaras Sutcu Imam University in 2011. He is studying at The University of Texas at San Antonio to receive his M.S. degree in Electrical Engineering with concentration in control and systems. He is planning to pursue a Ph.D. degree and contribute to science with academic research.