Modeling and Forecasting long-term Natural Gas (NG) consumption in Iran, using Particle Swarm Optimization (PSO)

Ebrahim Kamrani

2010
# DEGREE PROJECT

**Computer Engineering**

<table>
<thead>
<tr>
<th>Programme</th>
<th>Reg number</th>
<th>Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masters Programme in Computer Engineering - Applied Artificial Intelligence</td>
<td>E3714D</td>
<td>15 ECTS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of student</th>
<th>Year-Month-Day</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ebrahim Kamrani</strong></td>
<td><strong>2010-11-17</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supervisor</th>
<th>Examiner</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dr. Pascal Rebreyend</strong></td>
<td><strong>Dr. Hasan Fleyeh</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Company/Department</th>
<th>Supervisor at the Company/Department</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Title**

Modeling and Forecasting long-term Natural Gas (NG) consumption in Iran, using Particle Swarm Optimization (PSO)

**Keywords**

Modeling, Forecasting, Natural Gas (NG), Particle Swarm Intelligence (PSO), Iran
Abstract

The gradual changes in the world development have brought energy issues back into high profile. An ongoing challenge for countries around the world is to balance the development gains against its effects on the environment. The energy management is the key factor of any sustainable development program. All the aspects of development in agriculture, power generation, social welfare and industry in Iran are crucially related to the energy and its revenue.

Forecasting end-use natural gas consumption is an important factor for efficient system operation and a basis for planning decisions.

In this thesis, particle swarm optimization (PSO) used to forecast long run natural gas consumption in Iran. This approach is relatively simple compared with other forecasting approaches. In some unusual situations, such as abnormal temperature changes, the forecasting error is high. Although this error might seem high, one does not need to be deeply concerned about the overall results since these unusual situations could be filtered out to yield more reliable predictions. Gas consumption data in Iran for the previous 34 years is used to predict the consumption for the coming years. Four linear and nonlinear models proposed and six factors such as Gross Domestic Product (GDP), Population, National Income (NI), Temperature, Consumer Price Index (CPI) and yearly Natural Gas (NG) demand investigated.
Acknowledgment

I am very grateful to my advisor, Dr. Pascal Rebreyend, and my thesis examiner, Dr. Hasan Fleyeh for their inspiring guidance, patient support and constant encouragement throughout my research.

I would like to thank the Swedish Government for giving me an opportunity to study in Sweden and Department of computer science at Dalarna University for providing all the resources to complete my work.

Also I would like to express my gratitude to my professors in Dalarna university, Dr. Hasan Fleyeh, Dr. Siril Yella and Dr. Jerker Westin. Also a special thanks to Prof. Mark Dougherty for bringing enjoying moments on Cook's theorem proof! I have learnt a lot from them, not only about the knowledge relevant to this thesis, but also the professional ethics.

Special gratitude to all my friends back there in my country and also my colleagues here at Dalarna University with whom I had great experience in DU.

Last, but not least, I would like to thank my family, for their understanding and constant support, without which I would not be able to pursue my dream.
Dedications

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

For Bahar who offered me unconditional love and support throughout the course of this thesis.

Also this thesis is dedicated to the memories of Prof. Albert Hoffman for bringing inspiration. Finally, it is dedicated to No.1 DJ in the world, Armin van buuren who shares his passion on music with others and changed my path of life during past 3 years.

Armin van buuren No.1 DJ in the world in 2007, 2008, 2009 and 2010 [DJ]
TABLE OF CONTENTS

CHAPTER 1: PROBLEM DEFINITION .................................................................1

CHAPTER 2: PREVIOUS WORKS.................................................................2

CHAPTER 3: OPTIMIZATION METHODS.......................................................10

3. Heuristic/Metaheuristic Methods..............................................................10
   3.1 Point-Based Metaheuristics (Trajectory Methods)..............................11
   3.1.1 Tabu Search.............................................................................11
   3.1.2 Simulated Annealing.................................................................11
   3.1.3 Greedy Randomized Adaptive Search Procedure (GRASP).............12
   3.1.4 Iterative Local Search...............................................................12
   3.1.5 Guided Local Search................................................................12
   3.2 Population-Based Metaheuristics.......................................................12
   3.2.1 Genetic Algorithms...................................................................12
   3.2.2 Ant Colony Optimization (ACO).................................................12
   3.2.3 Particle Swarm Optimization (PSO).............................................13
   3.2.4 Genetic Programming (GP).........................................................13
   3.2.5 Evolutionary Strategies (ES).......................................................13
   3.2.6 Memetic Algorithms (MA)........................................................13

CHAPTER 4: SWARM INTELLIGENCE (SI) AND PARTICLE SWARM
OPTIMIZATION (PSO) IN DETAILS..............................................................14

4.1 Introduction to Swarm Intelligence.........................................................14
   4.1.1 Definition..............................................................................14
   4.1.2 Specifications..........................................................................14
4.2 Swarm Intelligence Types.......................................................................14
4.3 Introduction to Particle Swarm Optimization (PSO).................................14
4.4 Particle Swarm Optimization (PSO) algorithms.........................................15
   4.4.1 Algorithm description.................................................................15
   4.4.2 PSO parameters initialization.....................................................16
   4.4.3 PSO Pseudo code....................................................................17
4.5 Defined Neighborhoods........................................................................19
4.6 PSO comparison with GA......................................................................20
   4.6.1: Similarities............................................................................20
   4.6.2: Discriminations.....................................................................20
4.7 Pros and cons of using PSO.................................................................20
   4.7.1: Pros....................................................................................20
   4.7.2: Cons....................................................................................20
4.8 PSO extensions.....................................................................................20
   4.8.1 Hybrid algorithm....................................................................20
   4.8.2 Local Algorithm.....................................................................20
   4.8.3 ARPSO Algorithm..................................................................21
   4.8.4 FARPSO Algorithm...............................................................21
   4.8.5 FRANDPSO Algorithm..........................................................21
4.8.6 ARANDPSO Algorithm.................................................................22

CHAPTER 5: FORECASTING METHODOLOGY ........................................23

CHAPTER 6: MODELING AND DATA FEATURES STRUCTURE .........................25

CHAPTER 7: RUNNING MODES AND RESULTS........................................29

CHAPTER 8: DATA ANALYSIS AND DISCUSSION....................................32

CONCLUSION....................................................................................40

FUTURE WORKS..............................................................................40

REFERENCES..................................................................................41

APPENDIX A. Particle Swarm Intelligence (PSO) MATLAB Implementation............45

APPENDIX B. Mean Absolute Percentage Error (MAPE) function Excel Implementation1.47
LIST OF FIGURES

Fig 1: Worldwide Energy Sources graph in 2004 (World Energy Council)...............................5
Fig 2: World NG consumption, 2007-2035 (tcf) (International Energy Outlook 2010)..............6
Fig 3: Puzzle of optimization methods.....................................................................................10
Fig 4: Search Space and optimums.........................................................................................10
Fig 5: Particle position update in Search Space.....................................................................15
Fig 6: Particle vectors update in Search Space........................................................................16
Fig 7: PSO Flowchart.............................................................................................................18
Fig 8: Circle...........................................................................................................................19
Fig 9: Wheel...........................................................................................................................19
Fig 10: Star..............................................................................................................................19
Fig 11: Model 1 Results (Linear Model).................................................................................33
Fig 12: Model 1 Results (Underestimation, Average Error Percent is % 85.7819025).............33
Fig 13: Model 2 Results (Linear Model)...............................................................................35
Fig 14: Model 2 Results (Underestimation, Average Error Percent is % 61.2057075).............35
Fig 15: Model 3 Results (Linear Model)...............................................................................37
Fig 16: Model 3 Results (Underestimation, Average Error Percent is % 4.029355)...............37
Fig 17: Model 4 Results (Non-Linear Model).........................................................................39
Fig 18: Model 4 Results (Underestimation, Average Error Percent is % 5.237114)...............39

LIST OF TABLES

Table 1: Natural Gas Production in cubic meters (cu m). [CIA World Factbook]......................6
Table 2: World Natural Gas reserves by country (International Energy Outlook 2010).........7
Table 3: Data for 34 years of Iran Natural Gas (NG) and six features....................................27
Table 4: Model 1 Results.......................................................................................................29
Table 5: Model 2 Results.......................................................................................................30
Table 6: Model 3 Results.......................................................................................................30
Table 7: Model 4 Results.......................................................................................................31
Table 8: Model 1 Analysis.....................................................................................................32
Table 9: Model 2 Analysis.....................................................................................................34
Table 10: Model 3 Analysis....................................................................................................36
Table 11: Model 4 Analysis....................................................................................................38
Chapter 1
Problem Definition

End-use natural gas consumption can be grouped in three main categories: residential, commercial and industrial. In Iran residential and commercial natural gas consumption by end use is primarily linked, as usual, with heating (including hot water) and cooking. Therefore these two natural gas consumption categories are clearly influenced by temperature. However, natural gas used by industrial consumers is, in general, not clearly influenced by outdoor air temperature, and follows very different patterns. Natural gas is used by industry to produce different basic materials such as steel, glass, paper or bricks.

Due to limitations of the gas infrastructure (mainly storage facilities and terminals), large disequilibria between supply and demand cannot be compensated in the short term operation (hours). As a result demand and supply have to be matched by the system operator in advance (days). Thus, accurate medium-term forecasting of natural gas demand makes a crucial contribution to the ability of balancing demand and supply. Furthermore, from the operational point of view, it is important to know not only the future total demand but also the daily spatial distribution of it in order to check future pressures, temperatures and gas flows through the pipeline. Due to this last operational requirement, it is interesting to forecast the demand at the metering points with a daily resolution.

For a realistic problem formulation, most analytical, numerical programming or heuristic methods are unable to work well. In recent years, artificial intelligence based methods, such as genetic algorithms (GAs), have been applied to similar problems with promising results. Meanwhile, some new AI-based methods are introduced and developed. Although these AI-based methods do not always guarantee the globally optimal solution, they provide suboptimal (near globally optimal) solutions in a short CPU time. This thesis employs a modern AI-based method, Particle Swarm Optimization (PSO), to solve the problem.

PSO is a nature-inspired evolutionary stochastic algorithm developed by Kennedy and Eberhart [1]. This technique, motivated by social behavior of organisms such as bird flocking and fish schooling, has been shown to be effective in optimizing multidimensional problems. PSO, as an optimization tool, provides a population-based search procedure in which individuals, called particles, change their positions (states) with time. In a PSO system, particles fly around in a multidimensional search space. During flight, each particle adjusts its position according to its own experience, and the experience of neighboring particles, making use of the best position encountered by itself and its neighbors.

The main advantages of PSO are: it is very easy to implement and there are few parameters to adjust. In addition, PSO has been successfully applied in many different areas, such as artificial neural network training, fuzzy system control and mainly function optimization. Here, the possibilities of using a Particle Swarm Optimization (POS) as a Meta Heuristic model for forecasting of natural-gas consumption in Iran examined.
Chapter 2

Previous Works

Various deterministic and stochastic models have been applied to describe and forecast the evolution of natural gas consumption in different situations, for example, Maddala et al. [2] proposed a dynamic linear regression (DLR) model to estimate short-run and long-run elasticity’s of residential demands for natural gas in the USA for each of 49 states, as a function of the real per capita personal income, the real residential natural gas price, the real residential electricity price and the heating and cooling degree days. Subsequently, Batalgi et al. [3] discussed problems that might arise in evaluating forecasts produced with the above model, particularly as regards shrinkage estimators. Sarak and Saturau [4] described a deterministic model to forecast natural gas consumption for residential heating in certain areas of Turkey, based on previous studies performed by Durmayaz et al. [5]. Other studies have examined the total (domestic and industrial) consumption of natural gas and other fuels in large geographic areas. For example, Siemek et al. [6] consolidated earlier studies by Hubbert [7] and by Al Fattah et al.[8], proposing a deterministic model based on the logistic growth curve to describe and forecast natural gas consumption in Poland, taking into account the macroeconomic context and the economic cycles affecting the country. Stochastic logistic growth models have also been used in relation to the consumption of various fuels, with special attention to that of electricity. For example, Giovannis and Skiadas [9] described and forecasted total energy consumption in the USA and in Greece, developing estimation methods that were based on statistical inference by Continuous sampling for this type of logistic diffusion, and obtained good results. Other models of non-logistic deterministic growth, notably the Gompertz curve, have been widely used to describe phenomena such as the diffusion of technological innovations and the marketing of new products. A representative example is the model proposed by Franses [10] concerning the sale of new cars in the context of co integrated tri variate systems fitted to market data from the Netherlands. Darrat [11] discussed statistical problems associated with the Franses model [10], analyzing questions related to cointegration, and suggested the possibility of using other explanatory variables. Skiadas and Giovannis [12] applied the stochastic version of Bass’s classical growth-model to the study of electricity consumption in Greece.

It should be noted that efficient use of energy resources require accurate prediction of future energy demand [13]. Numerous researchers and practitioners have analyzed various energy issues and focused on developing appropriate energy demand models to reduce forecasting errors. Herbert [14] has analyzed monthly natural gas sales to the residential consumers in America [14]. Liu and Lin [15] have used time-series models to forecast residential natural gas consumption in Taiwan [15]. Eltony [16] has forecasted natural gas demand in Kuwait by means of econometric models [16]. Siemek [17] have estimated natural gas consumption in Poland based on the logistic curve interpretation. Sarak and Satman [18] have described a deterministic model to forecast natural gas consumption for residential heating in certain areas of Turkey [18] based on previous studies performed by Durmayaz et al. [19]. Kaboudan and Liu [20] have forecasted quarterly US demand for natural gas in short term using combination of genetic programming with a two-stage least squares (2SLS) regression system of equations [20]. Aras and Aras [21] have forecasted natural gas demand for residential sector in Turkey using auto-regression approach [21].
As energy is a vital input for social and economic development of any nation, planning for the energy supply to cover demand is an important issue and forecasting the demand is the starting point for energy planning in a country. Accurate forecasting of short-term Natural Gas (NG) consumption is of great importance for economical and reliable operation of distributive network. NG consumption in Iran in October 2009 has increased by 7.14% compared with October 2008 [22]. The share of residential sector is about 40.2% in total consumption (NIGC, 2009) [23]. According to increasing growth rate of NG consumption in Iran and indispensable need for distribution planning especially in residential areas, forecasting of demands on daily and weekly basis is particularly interesting in the cases of high demands, when the accumulation ability of network itself is decreased. There is an array of methods that are available today for forecasting energy demand. An appropriate method is chosen based on the nature of the data available and the desired nature and level of detail of the forecasts. An approach often used is to employ more than one method and then to compare the forecasts to arrive at a more accurate forecast.

In short-term forecasting, an alternative modeling approach is time series forecasting [22]. This approach uses a detailed analysis of patterns in historical demand to produce a forecast of future demand. Time series models are useful for short-term forecasting and for developing a picture of underlying patterns in data (changes in half-hourly load patterns at individual points of supply are good examples). However, they are of limited use in some situations, such as where there are underlying short-term changes in the key drivers of demand. Gonzales Chavez et al. (1999) [24] used Box–Jenkins time-series analyses (ARIMA) models to formulate the forecasting model for the prediction of energy production and consumption in Asturias, Northern Spain. The trend in current and near-future energy consumption from a statistical perspective by considering two factors, namely, increasing population and economic development had been discussed by Kadoshin et al. (2000) [25]. Saab et al. (2001) [26] had investigated different univariate modeling methodologies for the forecasting of monthly electric energy consumption in Lebanon. Three univariate models were used namely, autoregressive, autoregressive integrated moving average (ARIMA) and a novel configuration combining an AR with a high pass filter [22].

An oil and gas supply model (OGSM) has been solved and the projections of oil and natural gas supply and demand to the year 2020 for Canada have been presented in Jai Persaud and Kumar (2001) [28]. Chow (2001) [27] has discussed the seected energy consumption in Hong Kong for the period 1984–1997 with special emphasis on the household sector. Ediger and Tatlldil (2002) [29] used semi-statistical technique to formulate the forecasting model to predict the primary energy demand in Turkey and analysis of cyclic patterns. Reddy and Balachandra (2003) [30] looked at various factors that influence the energy demand in India and develop the energy and environmental outlook in the year 2010. This was done by developing an integrated mathematical model incorporating various factors such as GDP, population growth. The model presented in Siemek et al. (2003) [31] estimates natural gas demand, based on average trend of the economy development; the model considered natural production/demand maxima of energy carriers. The prognosis was loaded with an error resulting from the use of average data related to yearly increases of the national gross product. Gorucu and Gumrah (2004) [32] forecasted short-term gas consumption by multivariable regression analysis for the capital city of Ankara, Turkey. Gutierrez et al. (2005) [33]
examined the application of a Gompertz-type innovation diffusion process for stochastic modeling and capturing the growth process of natural gas consumption in Spain. Sanchez-Ubeda and Berzosa (2007) [34] forecasted industrial end-use natural gas consumption in a medium-term horizon (1–3 years) with a very high resolution (days) based on a decomposition approach. The forecast was obtained by the combination of three different components: one that captures the trend of the time series, a seasonal component and a transitory component. Parikh et al. (2007) [35] estimated demand projections of petroleum products and natural gas in India. They considered GDP and population as inputs of their NG estimation model. Yoo et al. (2009) [36] estimated households’ demand function for natural gas by applying a sample selection model using data from a survey of households in Seoul. There are also alternative approaches available such as neural network and hybrid models. These typically use multiple techniques and inputs, and produce forecasts based on the mix that produces the best results given the input data available at the time. Hybrid and neural network models have the potential to produce forecasts that perform well compared to the more traditional modeling approaches. Their main disadvantage is their “black box” nature because a forecasting model needs to be intuitive and easily explained to “non-experts”. Khotanzad et al. (2000) [37] focused on combination of artificial neural network (ANN) forecasters with application to the prediction of daily natural gas consumption needed by gas utilities. Gorucu et al. (2004) [38] trained the ANNs to decide the optimum parameters to be used in forecasting the gas consumption for short-term applications.

Also [39] employed an ANN approach for annual electricity consumption in high energy consumption industrial sectors. Azadeh et al. (2008) [40] developed an integrated algorithm for forecasting monthly electrical energy consumption based on ANN, computer simulation and design of experiments using stochastic procedures.

Azadeh et al. (2009) [41] proposed a new hybrid ANFIS computer simulation for improvement of electricity consumption estimation.

The proposed algorithm uses the most important relative error estimation method which is mean absolute percentage error (MAPE). All error estimation methods except MAPE have scaled output. As input data used for the model estimation have different scales, MAPE method is the preferred method to estimate relative errors. MAPE is calculated via:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{x_t - \hat{x}_t}{x_t} \right|$$

Such that $x_t$ is the actual data, $\hat{x}_t$ is the estimated value for $x_t$ and $n$ is the sample size. Excel implementation of mean absolute percentage error (MAPE) function is in Appendix B.

Natural gas (NG) is one of the most important energy sources in the world [13]. The natural gas consumption growth has been the fastest of all the fossil fuels in recent years. While the share of oil in the world's total energy produced declined to 38% in 2004 from 45% in 1970 [42], the share of natural gas has gone up to 23% from 17.2% for the same period. In the last
20 years, global production of natural gas has increased about 1.7 times and the US Energy Information Administration predicts its use to double by 2020 [42].

Fig 1: Worldwide Energy Sources graph in 2004 (World Energy Council).

Iran’s primary energy sources include oil, natural gas, electric power (three-quarters or more was natural gas fired, with the remainder either being hydroelectric or oil fired), solar, wood, animal and plant waste [43]. Oil and natural gas are the major source of primary energy in Iran. The known oil reserves are estimated to be about 90 billion barrels [43]. The major oil field of Iran is located in the southwest region and in the Persian Gulf. Iran produced about 3.5 million bbl/d of oil during past years. Iran’s current sustainable crude oil production capacity is estimated to be around 3.75 million bbl/d. The domestic oil consumption is about 1 million bbl/d in 2002. Iran’s oil export was about 2.5 million bbl/d in 2002, of which 28% went to Europe, 20% to Japan, and under 10% to South Korea, and other large customers including China and India [44]. Iran’s domestic oil consumption has been increasing rapidly as the economy and population grow (about 7% per year). The price of oil products is heavily subsidized to the tune of $3 billion or so per year [44]. This causes a large amount of waste and inefficiency in oil consumption; in addition, a substantial amount of petroleum products (mainly gasoline) is also smuggled to the neighboring countries since petroleum products are much cheaper in Iran than in neighboring countries [43]. Iran also imports around $1 billion per year worth oil products (mainly gasoline). Gasoline price is raised by 30–35% every year in early April. This is a part of an effort to curtail the rise in gasoline subsidy expenditure, gasoline consumption and imports (both are growing rapidly). The consumption trend is expected to continue in an unsustainable manner [45]. Besides public awareness of energy conservation methods, domestic supply and demand may be balanced by changing the price structure of oil products in Iran. Since Iran has nationalized oil resources, there is a general feeling among people that the government should ensure that oil products are available to the public at lower prices [43].
Table 1: Natural Gas Production in cubic meters (cu m). [CIA World Factbook]

<table>
<thead>
<tr>
<th>RANK</th>
<th>COUNTRY</th>
<th>[cu m]</th>
<th>DATE OF INFORMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>United States</td>
<td>593,400,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>2</td>
<td>Russia</td>
<td>546,800,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>3</td>
<td>European Union</td>
<td>131,600,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>4</td>
<td>Canada</td>
<td>161,300,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>5</td>
<td>Iran</td>
<td>116,900,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>6</td>
<td>Norway</td>
<td>103,000,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>7</td>
<td>Algeria</td>
<td>86,500,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>8</td>
<td>China</td>
<td>82,900,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>9</td>
<td>Netherlands</td>
<td>79,500,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>10</td>
<td>Saudi Arabia</td>
<td>77,100,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>11</td>
<td>Qatar</td>
<td>75,900,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>12</td>
<td>Indonesia</td>
<td>70,000,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>13</td>
<td>Uzbekistan</td>
<td>67,800,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>14</td>
<td>Egypt</td>
<td>62,700,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>15</td>
<td>Mexico</td>
<td>60,900,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>16</td>
<td>United Kingdom</td>
<td>58,500,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>17</td>
<td>Malaysia</td>
<td>57,300,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>18</td>
<td>United Arab Emirates</td>
<td>50,200,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>19</td>
<td>Australia</td>
<td>42,300,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>20</td>
<td>Argentina</td>
<td>41,300,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>21</td>
<td>Trinidad and Tobago</td>
<td>39,300,000,000</td>
<td>2008 est.</td>
</tr>
<tr>
<td>22</td>
<td>India</td>
<td>39,600,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>23</td>
<td>Pakistan</td>
<td>37,600,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>24</td>
<td>Kazakhstan</td>
<td>35,600,000,000</td>
<td>2009 est.</td>
</tr>
<tr>
<td>25</td>
<td>Turkmenistan</td>
<td>34,000,000,000</td>
<td>2009 est.</td>
</tr>
</tbody>
</table>
Iran owns the second largest oil and gas reservoir in the world. Iran is also the fourth largest producer and consumer of gas and the fourth largest producer and exporter of oil in the world [46]. A strong dependency has been created between the oil and gas sectors and the economic growth in the country during the hundred years of oil production and the forty years of development of gas consumption network. Oil revenue is also the biggest provider of the government’s budget and the growth of all sectors of the economy directly and indirectly is effected by petro-dollars (Central Bank of Iran Report, 2010) [46].

![World natural gas reserves by country as of January 1, 2010](image)

<table>
<thead>
<tr>
<th>Country</th>
<th>Reserves (trillion cubic feet)</th>
<th>Percent of world total</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>6,699</td>
<td>100.0</td>
</tr>
<tr>
<td>Top 20 Countries</td>
<td>6,003</td>
<td>90.8</td>
</tr>
<tr>
<td>Russia</td>
<td>1,680</td>
<td>25.4</td>
</tr>
<tr>
<td>Iran</td>
<td>1,046</td>
<td>15.8</td>
</tr>
<tr>
<td>Qatar</td>
<td>899</td>
<td>13.6</td>
</tr>
<tr>
<td>Turkmenistan</td>
<td>265</td>
<td>4.0</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>263</td>
<td>4.0</td>
</tr>
<tr>
<td>United States</td>
<td>245</td>
<td>3.7</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>210</td>
<td>3.2</td>
</tr>
<tr>
<td>Nigeria</td>
<td>185</td>
<td>2.8</td>
</tr>
<tr>
<td>Venezuela</td>
<td>176</td>
<td>2.7</td>
</tr>
<tr>
<td>Algeria</td>
<td>159</td>
<td>2.4</td>
</tr>
<tr>
<td>Iraq</td>
<td>112</td>
<td>1.7</td>
</tr>
<tr>
<td>Australia</td>
<td>110</td>
<td>1.7</td>
</tr>
<tr>
<td>China</td>
<td>107</td>
<td>1.6</td>
</tr>
<tr>
<td>Indonesia</td>
<td>106</td>
<td>1.6</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>85</td>
<td>1.3</td>
</tr>
<tr>
<td>Malaysia</td>
<td>83</td>
<td>1.3</td>
</tr>
<tr>
<td>Norway</td>
<td>82</td>
<td>1.2</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>65</td>
<td>1.0</td>
</tr>
<tr>
<td>Kuwait</td>
<td>63</td>
<td>1.0</td>
</tr>
<tr>
<td>Canada</td>
<td>62</td>
<td>0.9</td>
</tr>
<tr>
<td>Rest of World</td>
<td>606</td>
<td>9.2</td>
</tr>
</tbody>
</table>

Source: Oil & Gas Journal

Table 2: World Natural Gas reserves by country (International Energy Outlook 2010) [104]

Numerous elements have caused the nature and structure of the relationships in Iran energy sector to be different from other countries. For example [46] has mentioned some of the characteristics of some developing countries that should be taken into account when modeling energy policies. Another study was carried out by Urban et al. (2007) [Also in 46] on modeling of Energy systems for developing countries. They have indicated how to improve energy models for increasing their suitability for developing countries and have given advice on modeling techniques and data requirements.

The structural difference in Iran’s energy sector is different from the industrialized countries that most energy models have been developed for. In Iran the causal relationships have some major differences [46]. For example, the energy price for consumers is not a function of
production costs or the rate of exploration and oil production is not a function of global oil price but is according to OPEC’s share. Therefore the energy model for Iran must be developed based on particular relationships that exist in this country [46].

Some of the specifications of Iran that makes the energy relationships structure different from industrial countries [46] are as follows:

1- Governmental management of oil and gas production and its particular decisions and policies.
2- Low energy prices and very high energy intensity and the existence of illogical and uneconomical relation in consumption because of low energy prices.
3- Lack of technical and financial abilities for developing the oil and gas production due to political sanctions, difficulties caused by war, etc.
4- Existence of huge undeveloped resources and the slow development of the resources regardless of their final production capacity.
5- A high potential for oil and gas export in the future with the realization that there is a limit on these resources in the world.
6- Existence of high potential for energy conservation and possibility of decrease in domestic consumption due to the high amount of waste in various stages of oil and gas consumption since long ago, which has been continued to the present.
7- Very low final production costs of oil and gas in comparison to other countries and the possibility of competition with low world oil prices.
8- High effectiveness of oil revenue on the economy and the effects of oil revenue on investments in oil and gas productions.

Knowing the above facts, there are some weaknesses and strengths that affect the energy sector of Iran. Major policies should go towards demand side management and optimal production strategies [46]. The relationship framework and priorities used in the model are based on major policies to increase the productivity in the country. Based on this fact, gas consumption has been prioritized [46] as follows:

1. Domestic consumption.
2. Injection into the oil reservoirs in order to increase oil production.
3. Gas and LNG export.

Oil production is based on the country’s share in OPEC and OPEC’s share of oil production is related to global demand [46]. The policy to substitute oil consumption by gas in different sectors depends on the trend in gas production and the peak demand in the cold season of the country.

The policies [46] defined in this model can be considered in the following groups:

1. How to increase the productivity of energy in different sectors of economy and decrease the energy intensities of these sectors.
2. How to introduce policies related to improving the efficiency of production, refining and distribution of oil, gas and electricity.
3. Policies to introduce nuclear energy, hydroelectric power, decrease the associated flare gas.

The resources that supply peak demand for gas in Iran [46] are as follows:

1- The underground gas reservoirs, which reserve surplus gas during non-peak period.
2- Increasing production at some gas fields (mostly gas fields in central regions of Iran).
3- Substitution of gas with oil in power plants during cold season peak demand.
4- Not injecting gas in the oil fields during cold season and using the gas in residential and commercial sector.
Chapter 3
Optimization Methods

Optimization Methods
Here, some of the optimization methods will briefly introduce. We classify these methods in two major categories where exact methods will find optimal solutions and heuristic methods will find reasonable (not necessarily optimal) solutions in a reasonable computation cost. Although lots of optimization techniques developed in literature, here we address most famous ones in heuristic methods which applied on most combinatorial optimization problems.

Fig 3: Puzzle of optimization methods

Fig 4: Search Space and optimums
3 Heuristic Methods
When we are going to deal with NP-hard problems (problem class that their complexity and computational cost increase exponentially with the size of data), exact methods are impractical, so we use methods called heuristics where they use domain knowledge to solve problems in a more efficient way. However, heuristics will not guarantee optimality of the solution as exact methods do, but they are fast methods which provide good enough solutions. A vast number of Modern heuristics developed in the past years called metaheuristics with capabilities to escape from local optima toward the global optima. Hill climbing or iterative improvement move to a solution which should always be better than the current one until it finds local optima (the solution that is better among other neighborhoods). This method will stick very quickly in local optima that may be much worse than the global optimum. The idea of metaheuristics roughly is to accept solutions even if they are worse than the current one with the hope of finding better solutions on the search process [47]. Metaheuristics should make a balance between two main features known as intensification and diversification where the first one is a thorough check in promising region of search space and the latter discover and explore other parts of search space in hope of finding better solutions. A roughly classification of metaheuristics could be trajectory based and population based metaheuristics where in a trajectory metaheuristic, just one solution will consider and in a population based metaheuristics, a population of solutions will maintain. In the past few years metaheuristics applied on different combinatorial optimization problems successfully.

3.1 Point-Based Metaheuristics (Trajectory Methods)

3.1.1 Tabu Search
Glover [102] proposed this method in 1977 which is an adaptive memory based method. In iterative improvement of local search process, there is possibility of getting stuck in local optima; however tabu search tries to prevent getting stuck in local optima by accepting non-improving moves to even worse neighbors. This procedure may cause a cycle on previously visited solutions. Tabu search will use a short term memory structure (tabu list) to avoid this problem. Length of tabu list and different tabu lists for each type of neighborhood moves proposed in [48]. At some points tabu list prohibit search process from better moves where there is no risk of cycling, to override tabu list, aspiration criteria defined to accept a move even if it is in the tabu list. Tabu search hybridize with other metaheuristics methods.

3.1.2 Simulated Annealing
Kirkpatrick et all. [49] Propose this method in 1983. Origins of idea come from the physical annealing process of metals in Metropolis [50]. Intensification is possible by accepting a solution if it is better than the current solution, however diversification will provide by a probability of accepting a worse solution. An interesting point about this method is that there is a theoretical convergence proof. Also it is easy to implement comparing other metaheuristic methods. The probability of accepting worse solutions decrease during the search and it called cooling schedule implemented as a static schedule which should define before the search process and also in a dynamic way where could be based on information obtained during the search process. Similar methods to simulated annealing proposed in literature like Threshold Accepting [51, 52], noising method [53]. Some recent issues on simulated annealing
addressed in [54, 55]. Also a full review of different problem domains and applications where simulated annealing applied on could be found in [56].

3.1.3 Greedy Randomized Adaptive Search Procedure (GRASP)
This method is a multi-start metaheuristic with constructive and improvement phase on each iteration. Constructive heuristic applies on different seeds to generate different random starting points and in improvement phase it will greedy select the best candidate. These solutions after constructive phase are not necessarily optimal, so in improvement phase it uses a local search strategy based on a neighborhood structure. GRASP usually applies these two phases repeatedly to obtain optimal solution. It also calls adaptive because of updating the evaluation of other candidates when an element selected. A recent survey of greedy randomized adaptive search procedure is in [57] where it can be hybridize with other approaches [58].

3.1.4 Iterative Local Search
This method is similar to GRASP because of applying a local search repeatedly on different solutions. It starts with an initial solution which is a local optimum by itself, then perform a perturbation of solution and again apply a local search on new solution and it continues. Acceptance criteria will apply to decide whether the new one or the previous local optima should be accepted. A full investigation on method is in [59].

3.1.5 Guided Local Search
The idea is to change the objective function value when it finds local optima, so it diversifies search process to other parts of search space. It developed by Tsang and Voudouris [60, 61].

3.2 Population-Based Metaheuristics

3.2.1 Genetic Algorithms (GA)
Genetic algorithms mimic the evolution of biological species in nature. It first introduced in late 1950’s, [62, 63, 64] contributes to method and John Holland [65, 66] carried out much of the main work on genetic algorithms. Chromosomes simulated by a population of strings where crossover and mutation operators will recombine them. It guided by the objective function for all strings in maintained population. Individuals with higher fitness value have a better opportunity to breed according to survival of the fittest. From generation to generation, new solutions (offspring) repeatedly will improve to obtain optima. Genetic algorithms need to be set for some parameters like the initial population (mostly obtained at random), population size and mutation and crossover rate and type to apply, the selection mechanism for next generation, the stopping condition, [67], [68] by Davis and [69] by Goldberg are a part of most important literature in the field.

3.2.2 Ant Colony Optimization (ACO)
An early form of Ant Colony Optimization (ACO) called ant systems where developed by Dorigo. Ant Colony Optimization is a nature inspired optimization method that was developed based on real ants searching for food in nature. Ant Colony Optimization is based on indirect communication of ants as artificial agents by artificial pheromone trails
which mimics by numerical values ants use to construct solutions to a given problem probabilistically. This pheromone trails are updating as search continues to reflect experience of ants in search process. Ant Colony Optimization is a constructive heuristic which starts with a population of empty solutions and add new elements to partial solution to form the complete feasible solution for the problem. Usually problem will model by a graph with vertices and edges where ants start from vertices to other adjacent vertices of problem. The ants should decide on which vertex visit next. The amount of pheromone will guide other ants to choose a better path in the graph model. A level of pheromone evaporation defined to prevent early convergence. The evaporation rate and updating rules for pheromone trails is important for the balance between intensification and diversification of ACO method.

3.2.3 Particle Swarm Optimization (PSO)
This method have discussed in details in chapter 4.

3.2.4 Genetic Programming (GP)
This approach uses Darwinian principles of natural selection, crossover, mutation, gene duplication, gene deletion, etc. Genetic programming starts from a randomly created computer programs with composition of the functions and terminals of the problem, then it evaluate each program by executing them based on a fitness function iteratively. The programs will select for the next generation based on their fitness and this process continues until some stopping criteria met.

3.2.5 Evolutionary Strategies (ES)
Basic form of Evolutionary Strategies [70] is very close to genetic algorithms with two main features; first, ES is mostly applied to optimize real-valued variables, second is in its usage of just mutation. ES also applied on discrete variables and used crossover feature [71]. In [72, 73, 74, 75, 76, and 77] more details are discussed.

3.2.6 Memetic Algorithms (MA)
Basically, memetic algorithms are a combination of evolutionary strategies (also genetic algorithms) with one or more local search methods in the hope of more improvement for individuals. [78, 79] are two good starting points for memetic algorithms in literature.

Dalarna University
Röda vägen 3S-781 88
Borlänge Sweden
Tel: +46(0)23 7780000
Fax: +46(0)23 778080
http://www.du.se
Chapter 4
Swarm Intelligence (SI) and Particle Swarm Optimization (PSO) in details

4.1 Introduction to Swarm Intelligence (SI)
4.1.1 Definition
    In many years, researchers of different fields explored many interesting behavior of insects and animals in nature. Inspirations provided by the natural phenomena like bird flocking or fish schooling, which has been an interesting area of study in artificial life for quite some time now [80, 81]. Swarm intelligence theory developed by Dorigo, Bonabeau and Theraulaz. It can be applied in many different applications including but not limited to optimization methods in communication systems, robotics, traffic control systems and other applications [82, 83 and 84].

4.1.2 Specifications
    The most important characteristic of these swarms is their Self Organization behavior where there is no central controlling system for them and they behave as a group or swarm of intelligent agents. Robustness and Flexibility are two other important characteristics of swarms. [85, 86] Some of the general specifications of swarm come as below:
    • each member of group has interactions with other members;
    • each member of group has interactions with the environment;
    • each member of group coordinate its own movement path based on other members;
    • each member of group tries to be near to other neighbors/members;
    • there is no collision between group members;
    • there is no leader to command members;
    • each member of group can move in any space (front, center and rear);
    • Using social behavior, all members can achieve food and defense against hunters.

4.2 Swarm Intelligence (SI) Types
    Swarm intelligence (SI) describes the collective behavior of decentralized, self-organized systems, natural or artificial. There are different types of swarm in nature. Ant colonies, Bee hives, Bird flocking, Bacteria molding and Fish schooling are samples of these swarm. Among them two groups find more attraction between scientists. Bird flocking which Particle Swarm Optimization (PSO) is based on that and Ant colony makes Ant Colony Optimization (ACO) method. [87, 88]
In my work, Particle Swarm Optimization (PSO) used as preferred method.

4.3 Introductions to Particle Swarm Optimization (PSO)
    Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling.
Recently, PSO has been applied successfully in many areas such as continuous function optimization (Eberhart & Kennedy, 1995; Kennedy & Eberhart, 1995) [89, 90], flowshop scheduling (Tasgetiren, Liang, Sevkli, & Gencyilmaz, 2007) [91], job shop scheduling,
resource constrained project scheduling, part-machine grouping, task allocation, and neural network training (Shen, Shi, Yang, & Ye, 2006) [92]. PSO has many desirable characteristics:

- the concept of PSO is very simple and it can be implemented easily for many applications;
- it is versatile, robust and general purpose in that it can be applied to similar versions of a problem with minor modification;
- it is computationally efficient in the sense that reasonable solutions can be obtained in short computational time;
- unlike genetic algorithms it involves only a few parameters so that it is easier to find the best combination of parameter values;
- it is easily integrated with many local search techniques to improve the solution quality;
- the algorithm is well suited to parallelization by implementation on a cluster of workstations.

4.4 Particle Swarm Optimization (PSO) algorithms

4.4.1 Algorithm description

PSO is based on the interaction and the social communication of the group of particles. PSO, like GA or ACO, is a population based random search procedure which starts with a pool of possible particles and finally attempts to determine the best solution vector of variables under optimization, so that the fitness function reaches its satisfactory goal. In PSO, a solution is represented as a particle. Particles fly in the search space guided by their individual experience and the experience of the whole population. Each particle is actually a vector corresponding to a unique position (solution) in the search space. In addition, each particle is also associated with a velocity which is responsible for the motion of the particle. At the beginning of the algorithm, both the particles and their associated velocities are generated randomly. Over each generation, each particle’s position as well as velocity is to be updated until satisfactory solutions are found. The particle flies iteratively through the search space by using the velocity function, which is constantly updated according to its own previous experience and the group’s experience. Each particle tends to adjust the position toward its own previous best position and the group’s previous best position. Tracking and memorizing the best positions encountered builds upon the particle’s experience. PSO possesses a memory as every particle remembers the best position it has reached. The population, called a swarm, represents the set of points that are potential solutions. PSO combines local search (through self-experience) with global search (through neighboring experience), attempting to balance exploration and exploitation. [93, 94]

Fig 5: Particle position update in Search Space
Each particle has a valuation function should be optimized; its value is based on distance to
the goal. First it is initialized by random numbers and search will start. Each member will use
its best experience (pbest) and also best of other experiences (gbest). In some other variation
of PSO, particles will use the best experience of their neighbor (lbest). [93, 86] Particles will
update values using relations below:

\[
v_{t+1} = w_t \cdot v_t + c_1 \cdot \text{rand}() (pbest - present_t) + c_2 \cdot \text{rand}() (gbest - present_t)
\]

\[
\text{present}_{t+1} = \text{present}_t + v_{t+1}
\]

Such that, \(C_1\) and \(C_2\) are learning parameters and usually assumed equal to 2. \text{rand}() is a
random number generator function in \([0,1]\) range. \text{present}_t is the current position and \(v_t\) is the
velocity of particles. \(w_t\) is a control parameter for ratio of current velocity \((v_t)\) on the next
velocity and is a balance parameter between global search and local search and will decrease
to have a global search for first part of search and gradually toward local search. Updating
vectors comes below:

\[\text{Fig 6: Particle vectors update in Search Space}\]

4.4.2 PSO parameters initialization

As mentioned in [93, 84 and 86], Population size (number of particles) usually assign
between \([20, 40]\) for normal problems. For most problems 10 is enough but in complicated
cases \([100, 200]\) is also used. All vectors like present position, pbest, gbest and \(v_t\) can be with
more than one component which is number of parameters should be optimize. Each
component of present position should be in the range of searching space such that:
\[
present_{\text{min}} \leq \text{present}_t \leq present_{\text{max}}
\]

Using relation below, we can update \(w\) in each step:

\[
w = w_{\text{max}} - \frac{[\text{w}_{\text{max}} - \text{w}_{\text{min}} \ast \text{iter}]}{\text{max Iter}}
\]

Also, \(C_1\) and \(C_2\) parameters can have different values, based on that we have different modes
for our PSO as below: [84, 86, 93 and 94]

- Full Model \((c_1, c_2 > 0)\)
- Cognition Only \((c_1 > 0 \text{ and } c_2 = 0)\)
- Social Only \((c_1 = 0 \text{ and } c_2 > 0)\)
- Selfless \((c_1 = 0 \text{ and } c_2 > 0, \text{ and } g \neq i)\)
4.4.3 PSO Pseudo code
For each particle
   Initialize particle
END
Do
   For each particle
      Calculate fitness value
      If the fitness value is better than the best fitness value (pBest) in history
         set current value as the new pBest
   End
   Choose the particle with the best fitness value of all the particles as the gBest
   For each particle
      Calculate particle velocity
      Update particle position
   End
While maximum iterations or minimum error criteria is not attained
Fig 7: PSO Flowchart

Begin

Initialize The Parameters Of Problem Such as:
Dimension – Particle Number – C1, C2, ..... 

Initialize The Position and Velocity of particles using
Random Functions.

Initialize The Position and Velocity of particles using
Random Functions.

Iteration Number i

Calculate the Objective (Fitness) Functions.

Update the Position and Velocity of Particles

Calculate the pbest & gbest Values

Is The Stopping Condition Satisfied? (Reaching
satisfaction proximity to Goal or Reaching
predetermined number of Iterations)

gbest = Best solution of Optimization

End
4.5: Defined Neighborhoods

PSO algorithm can have different versions of Local and Global based on radius of searching neighbors of particle. Global version will assume gbest for each particle. It is fast but, trapping in local optima is possible. Local version will assume lbest for each particle and use a neighborhood radius; it is slower comparing to global version but probability of getting trapped in local optima decreased. So it is better to use global version for starting the algorithm and gradually use the local one [82, 83].

Three well known topologies for neighborhood in local version defined as below:

- Fig 8: Circle

![Circle diagram]

- Fig 9: Wheel

![Wheel diagram]

- Fig 10: Star

![Star diagram]
4.6 PSO comparisons with GA

4.6.1 Similarities
Both methods are population based, stochastic and use fitness function for evaluating population. As both are metaheuristic methods, there is no guaranty for optimal solution. Both methods share information in population using chromosome in GA and gBest (lBest) in PSO.

4.6.2 Discriminations
There is no evolutionary operator (mutation, crossover) in PSO. Particle will converge to optimal solution using memory for best visited position. Also basic PSO algorithm is a continuous one comparing to basic GA which is for discrete optimization.

4.7 Pros and cons of using PSO

4.7.1 Pros
PSO implementation is easy and number of parameters to choose is low. Number of particles comparing to GA is not that much important, PSO with a few number of particles can be compared with GA with a big population size. Using global search and local search combined in PSO is one of the most important features. w factor make a good balance between global and local search.

4.7.2 Cons
It is possible for PSO to be converged in a short time and trap in local optima. Another problem is fast sharing information caused to similar particle which increase probability of getting trapped in local optima and more iteration of algorithm cannot solve this problem. For solving this problem hybrid PSO model and lBest defined. Also the performance of algorithm is dependant to problem domain.

4.8 PSO extensions

4.8.1 Hybrid algorithms
This method is a hybrid of particles movement in basic PSO and particle movement with Evolutionary Algorithm (EA) suing tournament algorithm for selecting particles. First evaluation function will calculate for each particle, then each value of particle will compare to other k particles and a rank number will assign to each particles. Then particles sort based on their rank value (notice that particles are not sort based on their pbest or lbest). When all particles sorted, position and velocity of half of particles will update using the better half set. Although, changes in hybrid PSO are small and it has overhead of calculations and ranking but can improve results.

4.8.2 Local Algorithm
Basic PSO also called gBest model algorithm. In local algorithm each particle have access to pBest and lBest which is best results of its neighbors comparing to global algorithm.
that particles have access to the global best. So in the local algorithm, gBest is substitute by lBest.

4.8.3 Attractive and Repulsive PSO (ARPSO) Algorithm

All particles update their position and velocity based on others in basic PSO. Particles attract each other in the basic method, here in ARPSO there is a repulsive phase where particle velocity will update based on relation below: [86, 93 and 103]

$$v(t + 1) = \omega \cdot v(t) - \phi_1(p(t) - \bar{p}(t)) - \phi_2(g(t) - \bar{g}(t))$$

Where in this phase an explosion happens and particles go far from gBest and pBest. In attract phase particles will be near each other and when diversity become less than \(d_{low}\), it switches to repulsive phase and again when diversity become more than \(d_{high}\), it switches to attract phase. So we have an algorithm in two different phases with using different diversity in each step. These phases can be identified using setDirection function as below: [81]

Function setDirection

if (dir > 0 & diversity < dLow) dir = -1;
if (dir < 0 & diversity > dHigh) dir = 1;

In above function, flying direction of particles can be identified by assigning dir variable. Diversity of particles will calculate using calculateDiversity function as below:

$$\text{diversity}(S) = \frac{1}{|S| \cdot |L|} \cdot \sum_{i=1}^{|S|} \sqrt{\sum_{j=1}^N (p_{ij} - \bar{p}_j)^2}.$$ 

Where S is Swarm, |S| is size of Swarm, |L| is length of diameter of search space, N is dimension size of problem, \(p_{ij}\) is jth value and ith particle and \(\bar{p}_j\) is jth value average of particle p. [93]

4.8.4 Fuzzy diversity-guided Attractive/Repulsive (FARPSO) Algorithm

In ARPSO, when diversity reaches less than \(d_{low}\), all particles start searching in search space of problem. To avoid this, a group of particles exploit environment and for defining their number, the diversity value is used. So Swarm will divide in two groups, one for searching in search space and the other for using current knowledge. When diversity of swarm is less than \(d_{low}\), exploit group size will be zero and when it is more than \(d_{high}\), the size of exploit group becomes the size of swarm as below: [95, 103]

$$|\text{exploit}| := \begin{cases} 0 & \text{for diversity } \leq d_{low} \\ \frac{|S|}{d_{high} - d_{low}} & \text{for } d_{high} \geq \text{diversity} \\ \frac{diversity - d_{low}}{d_{high} - d_{low}} & \text{otherwise} \end{cases}$$

This method use fuzzy diversity-guided attractive/repulsive PSO (FARPSO) and is similar to previous method (ARPSO).

4.8.5 Fuzzy diversity-guided Random walk (FRANDPSO) Algorithm

Explosion behavior of above algorithm will send exploit group to the borders of search space, to improve this issue in ARPSO algorithm, Random walk cab apply. In FRANDPSO
algorithm, first a random number as scaling factor add to previous velocity. Then s initialize as a random number in \([1, v_{\text{max}}]\) and for each particle velocity will be checked to be valid based on relation below:

\[ \forall v_j : s \cdot v_j \leq v_{\text{max}} \]

Using this method, s will not be less than 1 unless maximum velocity is less than 1. The updating algorithm for velocities comes below: [81, 95]

**Algorithm** update velocity:

for each dimension \(j\) do  
\[ v_j := v_j \cdot U(0.8, 1.2) \]
\[ s := U(1, v_{\text{max}}) \]
for each dimension \(j\) do  
if \(v_{\text{max}} < |s \cdot v_j|\) then
\[ s := \frac{|s \cdot v_j|}{v_{\text{max}}} \]
for each dimension \(j\) do  
\[ v_j := s \cdot v_j \]

And algorithm for updating particle positions is:

**Algorithm** update position:

for each dimension \(j\) do  
\[ x_j := x_j + v_j \]
if \(x_j\) violates search space boundaries then  
\[ x_j := 2 \cdot (\text{violated boundary}) - x_j \]
\[ v_j := -v_j \]

In this algorithm, beside updating positions of basic PSO, when particle is on the borders of search space it will come back to search space \((v_j := -v_j)\) [98, 101, 102, 104 and 108]

**4.8.6 Attractive/Random-walk (ARANDPSO) Algorithm**

Above algorithm is not using fuzzy information, so using FARPSO another algorithm called ARPSO random attract-repulsive proposed. This algorithm uses different threshold levels, called attractive/random-walk PSO when random walk substitute repulsive phase. [95]
Chapter 5
Forecasting Methodology

As mentioned we use the PSO algorithm for solving this forecasting problem. In this problem we have linear and nonlinear estimates. Thus in a large dimension problem, we can't rely on simple LP or NLP Methods.

As mentioned before, Particle Swarm Optimization (PSO), as an optimization tool, provides a population-based search procedure in which individuals, called particles, change their positions (states) with time. In a PSO system, particles fly around in a multidimensional search space. During flight, each particle adjusts its position according to its own experience, and the experience of neighboring particles, making use of the best position encountered by itself and its neighbors.

Particle Swarm Optimization (PSO) shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation.

The main advantages of PSO are: it is very easy to implement and there are few parameters to adjust. In addition, PSO has been successfully applied in many different areas, such as artificial neural network training, fuzzy system control and mainly function optimization.

The language used to explain the PSO follows from the analogy of particles in a swarm. These key terms are:

1. Particle (individual, agent): each individual in the swarm;
2. position/location: a particle’s n-dimensional coordinates which represents a solution to the problem;
3. Swarm: the entire collection (population) of particles;
4. Fitness: the fitness function provides the interface between the physical problem and the optimization problem. The fitness function is a number representing the goodness of a given solution given by a position in solution space;
5. Generation: each iteration of optimization procedure using the PSO;
6. pbest (personal best): the position in parameter space of the best fitness returned for a specific particle;
7. gbest (global best): the position in parameter space of the best fitness returned for the entire swarm;
8. Vmax: the maximum velocity value allowed in a given direction.

Similar to the other population-based evolutionary algorithms, PSO is initialized with a population of random solutions (particles) using uniform distribution. However, each particle in PSO traces a trajectory in an n-dimensional search space, updating constantly a velocity vector based on best solutions found so far by that particle as well as others in the population (swarm).

Each particle keeps track of its coordinates in the problem space, which are associated with the best solution (fitness) it has achieved so far, pbest. Another “best” value tracked by the global version of the particle swarm optimizer is the overall best value, gbest, and its location, obtained so far by any particle in the population.
The Particle Swarm Optimization (PSO) algorithm, at each time step, changes the velocity of each particle moving towards its pbest and gbest locations. Velocity is weighted by random terms, with separate random numbers being generated for acceleration toward pbest and gbest locations, respectively. [96]

Each particle updates its velocity $V_t$ to catch up with the best particle $g$, as follows:

$$v_{t+1} = w \cdot v_t + c_1 \cdot rand() (pbest - present_t) + c_2 \cdot rand() (gbest - present_t)$$

$V []$ is the particle velocity, $present []$ is the current particle (solution). $Pbest []$ and $Gbest []$ are defined as stated before. $Rand ()$ is a random number between $(0,1)$. $c1$, $c2$ are two positive constants named learning factors. Usually $c1 = c2 = 2$.

As such, using the new velocity $V_t$, the particle’s updated.

$$present_{t+1} = present_t + v_{t+1}$$
Chapter 6
Modeling and Data Features Structure

In our Data, we have six types of features for 34 years of Iran and we want find a Linear or Non-linear model for our forecasting methodology, respect of below features to forecast the Gas Consumption of Iran.

1- Population
2- National Income (NI)
3- Gross Domestic Product (GDP)
4- Consumer Price Index (CPI)
5- Demand Of Previous Year
6- Temperature

For example, to model this problem, in the first attempt we use:

\[ F(\text{Population, GDP}) = A_1 + A_2 \times \text{Population} + A_3 \times \text{GDP} \]

We inject the Data (for example 25 years) to model and find \( A_1, A_2, \) and \( A_3 \).

In this thesis, MATLAB software is used for implementing the method. The main code is in Appendix A.

After Running the code and finding the solutions, we check the answers with 4 rests of our data to see the correctness of the solution. Then we change the main function and repeat these steps until we have a satisfactory solution to our problem.

While electricity demand has been the focus of the research community, studies on natural gas demand modeling and forecasting have not been reported in the scientific literature to the same extent. Furthermore most papers address the problem of very short-term forecasting (typically one-step-ahead), while medium and long-term energy demand forecasting is quite unusual. For example, Hippert et al [97] examined nearly 100 papers (published between 1991 and 1999) concerning only short-term electricity demand forecasting.

One of the main difficulties in accurate medium-term forecasting of demand is the determination of effective explanatory variables to use and the future prediction of them. Studies on energy consumption show that the energy consumption can be influenced mainly by calendar, climate-related and socioeconomic factors. Weather factors include temperature, relative humidity, wind speed, luminosity and rainfall (e.g Mirasgedis et al. [98]), being temperature the main driving weather factor for electricity demand. This non-linear influence is usually captured by using two temperature-derived variables: the heating degree days and the cooling degree days.

Total Natural Gas demand is also affected by socioeconomic factors such as the population growth, the gross domestic product (GDP), or the price.

Now we discuss important factor for proposed model in more detail:

1. Temperature
   One of the most important factors in gas consumption is temperature. Most of our heating instruments use natural gas as a source of energy, so we can assume
that decreasing the temperature leads to increasing gas consumption (especially in home use).

By severely decreasing the temperature, we saw extremely shortage of natural gas. We have so many methods for forecasting the temperature, so we can use average temperature of next years in our model. But we have two challenges:

a. *Type of average temperature:* there are several indexes of temperature such as Average temperature, average max temperature, average min temperature and etc.

   Among these factors it seems that average min temperature is the best factor. By going temperature less than zero, use of natural gas will increases severely.

b. *Average temperature of Cities:* there are several stations in Iran that we can use their measures. But because home natural gas consumption depends on population, we use average temperature of 10 most populated cities.

2. **Population**

   Clearly, increasing in population will lead to more energy consumption. Natural gas is an important energy factor, so we can assume more population leads to more consumption of natural gas.

3. **Gross Domestic Product (GDP)**

   One of the most important factors that help us to forecast industrial use of natural gas is Gross Domestic Product (GDP). Increasing this factor shows more production occurred, so this leads to more use of energy (also natural gas) by Factories.

4. **Consumer Price Index (CPI)**

   It seems that price of energy is very important in consumption. But because of subsidy in Iran, price is that low that we can ignore this factor.

**Proposed Models**

According to mentioned factors and importance of them, we propose four linear and nonlinear models:

1. \[ \alpha_0 + \alpha_1 \times \text{Pop} + \alpha_2 \times \text{GDP} + \alpha_3 \times \text{Temperature} \]

2. \[ \alpha_0 + \alpha_1 \times \text{Pop} + \alpha_2 \times \text{CPI} + \alpha_3 \times \text{GDP} + \alpha_4 \times \text{NI} + \alpha_5 \times \text{Temperature} \]

3. Demand of previous Year + \[ \alpha_1 \times \text{Pop} + \alpha_2 \times \text{CPI} + \alpha_3 \times \text{GDP} + \alpha_4 \times \text{NI} + \alpha_5 \times \text{Temperature} \]

4. Demand of previous Year + \[ \alpha_1 \times \text{Pop} + \alpha_2 \times \sqrt[3]{\text{CPI} \times \text{GDP} \times \text{NI}} + \alpha_3 \times \text{Temperature} \]
Data Collection
After setting our factors and models we collected necessary data. This data is collected from national sites in Iran such as: www.sci.org.ir, www.weather.ir [99, 100 and 101].

<table>
<thead>
<tr>
<th>YEAR</th>
<th>Average Temperature</th>
<th>Min Temperature</th>
<th>Population</th>
<th>NI</th>
<th>GDP</th>
<th>CPI</th>
<th>Demand of Previous Year</th>
<th>Yearly NG Demand*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973</td>
<td>8.64</td>
<td>30284.3</td>
<td>115368.8</td>
<td>162556.5</td>
<td>0.3</td>
<td>1970.4</td>
<td>2151.02</td>
<td></td>
</tr>
<tr>
<td>1974</td>
<td>8.91</td>
<td>31106.3</td>
<td>155649.4</td>
<td>174668.4</td>
<td>0.4</td>
<td>2151.02</td>
<td>2528.68</td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>9.20</td>
<td>31950.7</td>
<td>241369.7</td>
<td>196581</td>
<td>0.6</td>
<td>2528.68</td>
<td>2331.64</td>
<td></td>
</tr>
<tr>
<td>1976</td>
<td>9.50</td>
<td>32817.9</td>
<td>261044</td>
<td>206113.8</td>
<td>0.7</td>
<td>2331.64</td>
<td>2463</td>
<td></td>
</tr>
<tr>
<td>1977</td>
<td>8.88</td>
<td>33708.7</td>
<td>298227</td>
<td>242326</td>
<td>0.9</td>
<td>2463</td>
<td>2742.14</td>
<td></td>
</tr>
<tr>
<td>1978</td>
<td>8.58</td>
<td>35025.2</td>
<td>287920.3</td>
<td>236645.3</td>
<td>1.2</td>
<td>2742.14</td>
<td>2807.82</td>
<td></td>
</tr>
<tr>
<td>1979</td>
<td>8.38</td>
<td>36393.1</td>
<td>236256.5</td>
<td>219191</td>
<td>1.5</td>
<td>2807.82</td>
<td>2200.28</td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>8.80</td>
<td>37814.3</td>
<td>246977.3</td>
<td>209919.4</td>
<td>1.9</td>
<td>2200.28</td>
<td>2512.26</td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td>9.11</td>
<td>39291.1</td>
<td>181690.1</td>
<td>178149</td>
<td>2.4</td>
<td>2512.26</td>
<td>2118.18</td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td>8.05</td>
<td>40825.6</td>
<td>167943.3</td>
<td>170281.2</td>
<td>3</td>
<td>2118.18</td>
<td>2610.78</td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>8.16</td>
<td>42420</td>
<td>189465.9</td>
<td>191666.8</td>
<td>3.7</td>
<td>2610.78</td>
<td>3612.4</td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>8.20</td>
<td>44076.6</td>
<td>207798.7</td>
<td>212876.5</td>
<td>4.6</td>
<td>3612.4</td>
<td>4137.84</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>8.52</td>
<td>45720.7</td>
<td>193059.8</td>
<td>208515.9</td>
<td>5.6</td>
<td>4137.84</td>
<td>5123.04</td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>8.71</td>
<td>47541.4</td>
<td>190020.6</td>
<td>212686.3</td>
<td>6.9</td>
<td>5123.04</td>
<td>4975.26</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>9.34</td>
<td>49445</td>
<td>151510.1</td>
<td>193235.4</td>
<td>8.5</td>
<td>4975.26</td>
<td>4712.54</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>9.32</td>
<td>50650</td>
<td>178770</td>
<td>191312.4</td>
<td>10.4</td>
<td>4712.54</td>
<td>5402.18</td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>8.61</td>
<td>51890</td>
<td>129657.9</td>
<td>180822.5</td>
<td>12.7</td>
<td>5402.18</td>
<td>5582.8</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>9.13</td>
<td>53167</td>
<td>145080</td>
<td>191502.6</td>
<td>15.4</td>
<td>5582.8</td>
<td>8850.38</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>9.35</td>
<td>54483</td>
<td>175631.9</td>
<td>218538.7</td>
<td>18.6</td>
<td>8850.38</td>
<td>12298.58</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>8.13</td>
<td>55837</td>
<td>194495</td>
<td>245036.4</td>
<td>22.4</td>
<td>12298.58</td>
<td>16091.6</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>8.83</td>
<td>56963</td>
<td>199788.7</td>
<td>254822.5</td>
<td>27.9</td>
<td>16091.6</td>
<td>17930.64</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>9.39</td>
<td>58114</td>
<td>238445.5</td>
<td>258601.4</td>
<td>34.3</td>
<td>17930.64</td>
<td>20968.34</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>9.19</td>
<td>59290</td>
<td>228785.2</td>
<td>259876.3</td>
<td>46.3</td>
<td>20968.34</td>
<td>23825.42</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>9.39</td>
<td>59151</td>
<td>232124.5</td>
<td>267534.2</td>
<td>69.2</td>
<td>23825.42</td>
<td>28111.04</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>9.50</td>
<td>60055.5</td>
<td>246865.3</td>
<td>283806.6</td>
<td>85.2</td>
<td>28111.04</td>
<td>26813.86</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3: Data for 34 years of Iran Natural Gas (NG) including six features

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>10.05</td>
<td>60936.5</td>
<td>244857.4</td>
<td>291768.7</td>
<td>100</td>
<td>26813.86</td>
<td>29211.18</td>
</tr>
<tr>
<td>1999</td>
<td>10.06</td>
<td>61830</td>
<td>234347.4</td>
<td>300139.6</td>
<td>118.1</td>
<td>29211.18</td>
<td>29178.34</td>
</tr>
<tr>
<td>2000</td>
<td>9.86</td>
<td>62736</td>
<td>259203.6</td>
<td>304941.2</td>
<td>141.8</td>
<td>29178.34</td>
<td>34498.42</td>
</tr>
<tr>
<td>2001</td>
<td>10.43</td>
<td>63663.9</td>
<td>271785.4</td>
<td>320068.9</td>
<td>159.7</td>
<td>34498.42</td>
<td>36682.28</td>
</tr>
<tr>
<td>2002</td>
<td>10.02</td>
<td>64528.2</td>
<td>282526.5</td>
<td>330623.6</td>
<td>177.9</td>
<td>36682.28</td>
<td>37897.36</td>
</tr>
<tr>
<td>2003</td>
<td>10.07</td>
<td>65540.2</td>
<td>317877.7</td>
<td>355554</td>
<td>206</td>
<td>37897.36</td>
<td>43233.86</td>
</tr>
<tr>
<td>2004</td>
<td>9.55</td>
<td>66991.6</td>
<td>341161</td>
<td>379838</td>
<td>238.2</td>
<td>43233.86</td>
<td>47306.02</td>
</tr>
<tr>
<td>2005</td>
<td>7.55</td>
<td>67477.5</td>
<td>373506</td>
<td>398235</td>
<td>274.5</td>
<td>47306.02</td>
<td>56139.98</td>
</tr>
<tr>
<td>2006</td>
<td>8.65</td>
<td>68467.4</td>
<td>357594.1</td>
<td>387392</td>
<td>307.6</td>
<td>56139.98</td>
<td>60392.76</td>
</tr>
</tbody>
</table>
Chapter 7
Running Modes and Results

We use this data in proposed models and run them by MATLAB software (code mentioned in Appendix A). The algorithm run 8 times for each model by different particle and iteration numbers. Also we use MAPE function (code mentioned in Appendix B) for fitness function as it proposed in mentioned literatures:

$$\frac{\sum_{i=1}^{TrainYear} (ActualUse_i - CalculatedData_i)}{ActualUse_i}$$

Our data is for 34 years from 1973 to 2006. For training part, data of 30 years is used. Also for testing prediction model, data of 4 years is used. Calculated data is summarized in following tables:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Particle Number</th>
<th>α0</th>
<th>α1</th>
<th>α2</th>
<th>α3</th>
<th>Gbest</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>500</td>
<td>-6.8950</td>
<td>-0.0061</td>
<td>0.0426</td>
<td>-478.8340</td>
<td>0.5783</td>
</tr>
<tr>
<td>10000</td>
<td>500</td>
<td>-108.3063</td>
<td>0.1537</td>
<td>-0.0112</td>
<td>-95.3451</td>
<td>0.4441</td>
</tr>
<tr>
<td>5000</td>
<td>1000</td>
<td>17.3428</td>
<td>0.4545</td>
<td>-0.0794</td>
<td>47.6435</td>
<td>0.7737</td>
</tr>
<tr>
<td>10000</td>
<td>1000</td>
<td>0.0768</td>
<td>0.1115</td>
<td>-0.0102</td>
<td>-5.0890</td>
<td>0.5300</td>
</tr>
<tr>
<td>5000</td>
<td>1500</td>
<td>9.6384</td>
<td>0.1295</td>
<td>0.0079</td>
<td>-363.2421</td>
<td>0.4043</td>
</tr>
<tr>
<td>10000</td>
<td>1500</td>
<td>-7970.5</td>
<td>0.2414</td>
<td>0.0091</td>
<td>86.2578</td>
<td>0.4212</td>
</tr>
<tr>
<td>5000</td>
<td>2000</td>
<td>-7.7092</td>
<td>0.1448</td>
<td>-0.0100</td>
<td>-35.9945</td>
<td>0.4124</td>
</tr>
<tr>
<td>10000</td>
<td>2000</td>
<td>11.8126</td>
<td>0.1090</td>
<td>-0.0124</td>
<td>4.7826</td>
<td>0.6053</td>
</tr>
</tbody>
</table>

Table 4: Model 1 Results

Model 1: $α0 + α1*Pop + α2*GDP + α3*Temperature$
<table>
<thead>
<tr>
<th>Iteration</th>
<th>Particle Number</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
<th>$\alpha_5$</th>
<th>Gbest</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>500</td>
<td>-10.0928</td>
<td>0.3217</td>
<td>-12.2257</td>
<td>0.0585</td>
<td>-0.0034</td>
<td>-362.8567</td>
<td>3.3536</td>
</tr>
<tr>
<td>10000</td>
<td>500</td>
<td>163.4144</td>
<td>-0.4917</td>
<td>39.8980</td>
<td>0.4587</td>
<td>-0.4366</td>
<td>932.0010</td>
<td>3.2257</td>
</tr>
<tr>
<td>5000</td>
<td>1000</td>
<td>7.3529</td>
<td>-0.2813</td>
<td>-54.6686</td>
<td>0.1978</td>
<td>-0.1579</td>
<td>-22.6196</td>
<td>1.9684</td>
</tr>
<tr>
<td>10000</td>
<td>1000</td>
<td>30.5722</td>
<td>1.4174</td>
<td>229.5585</td>
<td>-0.4349</td>
<td>0.0782</td>
<td>1943.9</td>
<td>1.7945</td>
</tr>
<tr>
<td>5000</td>
<td>1500</td>
<td>8.0635</td>
<td>0.1914</td>
<td>143.9911</td>
<td>-0.0211</td>
<td>0.0114</td>
<td>8.2013</td>
<td>0.6454</td>
</tr>
<tr>
<td>10000</td>
<td>1500</td>
<td>4.5909</td>
<td>-0.2155</td>
<td>2.6805</td>
<td>0.1762</td>
<td>-0.1012</td>
<td>57.3520</td>
<td>0.8208</td>
</tr>
<tr>
<td>5000</td>
<td>2000</td>
<td>175.8150</td>
<td>0.2582</td>
<td>17.5303</td>
<td>-0.0182</td>
<td>-0.0154</td>
<td>58.4994</td>
<td>0.5390</td>
</tr>
<tr>
<td>10000</td>
<td>2000</td>
<td>8.6960</td>
<td>-0.6316</td>
<td>7.4754</td>
<td>0.3679</td>
<td>-0.2389</td>
<td>-11.6640</td>
<td>1.3973</td>
</tr>
</tbody>
</table>

Table 5: Model 2 Results

Model 2: $\alpha_0 + \alpha_1 \cdot \text{Pop} + \alpha_2 \cdot \text{CPI} + \alpha_3 \cdot \text{GDP} + \alpha_4 \cdot \text{NI} + \alpha_5 \cdot \text{Temperature}$

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Particle Number</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
<th>$\alpha_5$</th>
<th>best</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>500</td>
<td>0.0253</td>
<td>6.1186</td>
<td>0.0026</td>
<td>-0.0013</td>
<td>-101.7891</td>
<td>0.1034</td>
</tr>
<tr>
<td>10000</td>
<td>500</td>
<td>0.0223</td>
<td>11.3883</td>
<td>0.0044</td>
<td>-0.0020</td>
<td>-113.0266</td>
<td>0.1033</td>
</tr>
<tr>
<td>5000</td>
<td>1000</td>
<td>0.0335</td>
<td>3.7234</td>
<td>0.0082</td>
<td>-0.0042</td>
<td>-198.0347</td>
<td>0.1045</td>
</tr>
<tr>
<td>10000</td>
<td>1000</td>
<td>0.0251</td>
<td>7.7092</td>
<td>0.0032</td>
<td>-0.0015</td>
<td>-107.3546</td>
<td>0.1030</td>
</tr>
<tr>
<td>5000</td>
<td>1500</td>
<td>0.0236</td>
<td>11.6457</td>
<td>0.0023</td>
<td>-0.0014</td>
<td>-85.9110</td>
<td>0.1032</td>
</tr>
<tr>
<td>10000</td>
<td>1500</td>
<td>0.0237</td>
<td>7.3269</td>
<td>0.0014</td>
<td>-0.0008</td>
<td>-78.0451</td>
<td>0.1032</td>
</tr>
<tr>
<td>5000</td>
<td>2000</td>
<td>0.0263</td>
<td>12.5147</td>
<td>-0.0022</td>
<td>-0.0006</td>
<td>-18.7125</td>
<td>0.1048</td>
</tr>
<tr>
<td>10000</td>
<td>2000</td>
<td>0.0227</td>
<td>9.8789</td>
<td>0.0028</td>
<td>-0.0014</td>
<td>-91.9696</td>
<td>0.1032</td>
</tr>
</tbody>
</table>

Table 6: Model 3 Results

Model 3: Demand of previous Year + $\alpha_1 \cdot \text{Pop} + \alpha_2 \cdot \text{CPI} + \alpha_3 \cdot \text{GDP} + \alpha_4 \cdot \text{NI} + \alpha_5 \cdot \text{Temperature}$

---

Dalarna University  
Röda vägen 3S-781 88  
Borlänge Sweden  
Tel: +46(0)23 7780000  
Fax: +46(0)23 778080  
http://www.du.se
<table>
<thead>
<tr>
<th>Iteration</th>
<th>Particle Number</th>
<th>α₁</th>
<th>α₂</th>
<th>α₃</th>
<th>Gbest</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>10</td>
<td>-0.0094</td>
<td>0.1476</td>
<td>-0.8186</td>
<td>0.1092</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>0.0201</td>
<td>0.0623</td>
<td>-81.6813</td>
<td>0.1060</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>-0.0096</td>
<td>0.1671</td>
<td>-10.5017</td>
<td>0.1088</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>0.0070</td>
<td>0.0959</td>
<td>-43.9315</td>
<td>0.1061</td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>0.0090</td>
<td>0.1049</td>
<td>-48.8668</td>
<td>0.1084</td>
</tr>
<tr>
<td>200</td>
<td>30</td>
<td>0.0195</td>
<td>0.0460</td>
<td>-69.7949</td>
<td>0.1059</td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>0.0075</td>
<td>0.0872</td>
<td>-39.9476</td>
<td>0.1065</td>
</tr>
<tr>
<td>200</td>
<td>40</td>
<td>0.0143</td>
<td>0.0692</td>
<td>-61.2409</td>
<td>0.1058</td>
</tr>
</tbody>
</table>

Table 7: Model 4 Results

Model 4: Demand of previous Year + α₁*Pop + α₂*√CPI * GDP * NI + α₃*Temperature
Chapter 8
Data Analysis and discussion

Here is the analysis of obtained data after running the algorithm several times with different parameters for each model.

Model 1: Fitness Function \((a_0 + a_1*\text{Pop} + a_2*\text{GDP} + a_3*\text{Temperature})\)

In this model that we used most important factors, we didn't have good result. The main idea for explaining this large error is because our model was trained with data that doubled or maximum tripled over 30 years, natural gas consumption is multiplied by 20. So our model can’t predict satisfactorily.

The following formula is average of coefficient that is calculated 8 times for each model:

\[ -14.0195 + 0.167288*\text{Pop} - 0.00795 * \text{GDP} + -104.978 * \text{Temperature} \]

So in model 1 average of coefficients are:
- \(a_0 = -14.0195\)
- \(a_1 = 0.167288\)
- \(a_2 = -0.00795\)
- \(a_3 = -104.978\)

By using these average coefficients in test years, we have these predications and error percent:

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Use Of Natural Gas</th>
<th>Estimated Use Of Natural Gas</th>
<th>Distance Between Data</th>
<th>Error Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>43233.86</td>
<td>7066.287</td>
<td>36167.57</td>
<td>83.65567</td>
</tr>
<tr>
<td>2004</td>
<td>47306.02</td>
<td>7170.619</td>
<td>40135.4</td>
<td>84.84206</td>
</tr>
<tr>
<td>2005</td>
<td>56139.98</td>
<td>7315.604</td>
<td>48824.38</td>
<td>86.96899</td>
</tr>
<tr>
<td>2006</td>
<td>60392.76</td>
<td>7451.929</td>
<td>52940.83</td>
<td>87.66089</td>
</tr>
</tbody>
</table>

Table 8: Model 1 Analysis

Average Error Percent is \(\%\) 85.7819025 that shows our model can’t predict satisfactorily.
As we can see, Model 1 which is a linear model can’t predict satisfactorily with a considerable underestimation issue.
Model 2: Fitness Function \( (a_0 + a_1 \cdot \text{Pop} + a_2 \cdot \text{CPI} + a_3 \cdot \text{GDP} + a_4 \cdot \text{NI} + a_5 \cdot \text{Temperature}) \)

In this model that we use all of our factors except demand of previous year, we didn’t have good results too. It is because of same reason in previous model. But we have slight improvement. The main idea for explaining this improvement is using Consumer Price Index (CPI) parameter that multiplied by 3000 during these 30 years.

The following formula is average of coefficient that is calculated 8 times for each model:

\[
48.55151 + 0.071075 \cdot \text{Pop} + 46.77994 \cdot \text{CPI} + 0.098113 \cdot \text{GDP} + -0.10798 \cdot \text{NI} + 325.3517 \cdot \text{Temperature}
\]

So in model 2 averages of coefficients are:
- \( a_0 = 48.55151 \)
- \( a_1 = 0.071075 \)
- \( a_2 = 46.77994 \)
- \( a_3 = 0.098113 \)
- \( a_4 = -0.10798 \)
- \( a_5 = 325.3517 \)

By using these coefficients in test years, we have these predictions and error percent:

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Use Of Natural Gas</th>
<th>Estimated Use Of Natural Gas</th>
<th>Distance Between Data</th>
<th>Error Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>43233.86</td>
<td>18179.82</td>
<td>25054.04</td>
<td>57.95005</td>
</tr>
<tr>
<td>2004</td>
<td>47306.02</td>
<td>19488.55</td>
<td>27817.47</td>
<td>58.80323</td>
</tr>
<tr>
<td>2005</td>
<td>56139.98</td>
<td>20382.87</td>
<td>35757.11</td>
<td>63.69278</td>
</tr>
<tr>
<td>2006</td>
<td>60392.76</td>
<td>21513.85</td>
<td>38878.91</td>
<td>64.37677</td>
</tr>
</tbody>
</table>

Table 9: Model 2 Analysis

Average Error Percent is % 61.2057075 that shows our model can’t predict satisfactorily.
As we can see, Model 2 which is a linear model can’t predict satisfactorily with a considerable underestimation issue.
Model 3: Fitness Function (Demand of previous Year + $a_1$*Pop + $a_2$*CPI + $a_3$*GDP + $a_4$*NI + $a_5$*Temperature)

In this model we used all of our factors and we saw very good results. It shows that using of previous year demand improves ability of our model to forecast natural gas consumption strongly.

The following formula is average of coefficient that is calculated 8 times for each model:

\[
\text{Demand of previous Year} + 0.025313 \times \text{Pop} + 8.788213 \times \text{CPI} + 0.002838 \times \text{GDP} - 0.00165 \times \text{NI} - 99.3554 \times \text{Temperature}
\]

So in model 3 averages of coefficients are:
- $a_1 = 0.025313$
- $a_2 = 8.788213$
- $a_3 = 0.002838$
- $a_4 = -0.00165$
- $a_5 = -99.3554$

By using these coefficients in test years, we have these predictions and error percent:

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Use Of Natural Gas</th>
<th>Estimated Use Of Natural Gas</th>
<th>Distance Between Data</th>
<th>Error Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>43233.86</td>
<td>40850.81</td>
<td>2383.054</td>
<td>5.512008</td>
</tr>
<tr>
<td>2004</td>
<td>47306.02</td>
<td>46589.19</td>
<td>716.8288</td>
<td>1.515301</td>
</tr>
<tr>
<td>2005</td>
<td>56139.98</td>
<td>51190.22</td>
<td>4949.765</td>
<td>8.816827</td>
</tr>
<tr>
<td>2006</td>
<td>60392.76</td>
<td>60226.31</td>
<td>166.4464</td>
<td>0.275606</td>
</tr>
</tbody>
</table>

Table 10: Model 3 Analysis

Average Error Percent is % 4.0299355 that shows our model can predict satisfactorily.
As we can see, Model 3 which is also a linear model can predict satisfactorily with a few Average Error Percent of % 4.0299355.
Model 4: Fitness Function (Demand of previous Year + $a_1 \cdot \text{Pop} + a_2 \cdot (\sqrt[3]{\text{CPI} \cdot \text{GDP} \cdot \text{NI}}) + a_3 \cdot \text{Temperature})$

In the last model - like model 3 - we used all of our factors but we used social factors (CPI – GDP – NI) as Geometric average in our model. Our results are similar to third model that shows nonlinear modeling is good as linear modeling. Also we mention that running our forth model was so time consuming that we had to reduce our particle number and iterations. 

The following formula is average of coefficient that is calculated 8 times for each model:

$$\text{Demand of previous Year} + 0.0073 \cdot \text{Pop} + 0.097525 \cdot (\sqrt[3]{\text{CPI} \cdot \text{GDP} \cdot \text{NI}}) + -44.5979 \cdot \text{Temperature}$$

So in model 3 averages of coefficients are:

- $a_1 = 0.0073$
- $a_2 = 0.097525$
- $a_3 = -44.5979$

By using these coefficients in test years, we have these predictions and error percent:

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Use Of Natural Gas</th>
<th>Estimated Use Of Natural Gas</th>
<th>Distance Between Data</th>
<th>Error Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>43233.86</td>
<td>40439.96</td>
<td>2793.901</td>
<td>6.462298</td>
</tr>
<tr>
<td>2004</td>
<td>47306.02</td>
<td>46055.35</td>
<td>1250.672</td>
<td>2.643791</td>
</tr>
<tr>
<td>2005</td>
<td>56139.98</td>
<td>50486.99</td>
<td>5652.992</td>
<td>10.06946</td>
</tr>
<tr>
<td>2006</td>
<td>60392.76</td>
<td>59322.05</td>
<td>1070.707</td>
<td>1.772907</td>
</tr>
</tbody>
</table>

Table 11: Model 4 Analysis

Average Error Percent is % 5.237114 that shows our model can predict satisfactorily.
As we can see, Model 3 which is also a linear model can predict satisfactorily with a few Average Error Percent of % 5.237114.

Fig 17: Model 4 Results (Non-Linear Model)

Fig 18: Model 4 Results (Underestimation, Average Error Percent is % 5.237114)
Conclusion

All the aspects of development in agriculture, power generation, social welfare and industry in Iran are crucially related to the energy and its revenue. Forecasting end-use natural gas consumption is an important factor for efficient system operation and a basis for planning decisions. Both modeling and long-term forecasting covered for end-use natural gas consumption in Iran.

Particle swarm optimization (PSO) used to forecast long run natural gas consumption in Iran. In some unusual situations, such as abnormal temperature changes, the forecasting error is high. Although this error might seem high, one does not need to be deeply concerned about the overall results since these unusual situations could be filtered out to yield more reliable predictions. Gas consumption data in Iran for the previous 30 years is used to predict the consumption for four years. Four linear and nonlinear models proposed for this purpose. Six factors such as Gross Domestic Product (GDP), Population, National Income (NI), Temperature, Consumer Price Index (CPI) and yearly Natural Gas (NG) demand investigated. Models 1, 2 and 3 are linear models where model 4 is a non-linear model, the first and second model can’t predict satisfactorily with high Average Error Percent by order % 85.7819025 and % 61.2057075. Model 3 and 4 can predict satisfactorily with a low Average Error Percent by order % 4.0299355 and % 5.237114. Although model 3 is a linear model and model 4 is a non-linear model, but both can predict satisfactorily. Model 4 uses just three coefficients where model 3 uses five coefficients. Best prediction for model 3 is with 1000 particles in 10,000 iterations where best prediction for model 4 is with 40 particles in just 200 iterations.

Future Works

All four models for prediction suffer from underestimation (although it is with a very low Average Error Percent of % 4.0299355 or % 5.237114); in future we continue to build other robust models (by using more experts’ opinion) for better estimation of long-term natural gas usage in Iran. It is also possible to add more features to our prediction model, for example the effect of other political issues (such as sanction), Interest Rate, Rate of Inflation, National Power Index (NPI), relative humidity, wind speed, luminosity, rainfall, etc. Also we can use some other metaheuristic models such as ACO, GA. Also using different forecasting methods such as ANN [Back propagation, Recurrent], Time series [ARIMA, AR, MA, MAAR, ARMA], Regression, to apply them as a forecasting module. Another approach besides above is to use Particle Swarm Intelligence (PSO) as a tool for other optimization problems like long-term electricity usage estimation.

Furthermore, from the operational point of view, it is important to know not only the future total demand but also the daily spatial distribution of it in order to check future pressures, temperatures and gas flows through the pipeline. Due to this last operational requirement, it is interesting to forecast the demand at the metering points with a daily resolution.
References


[100] Statistical Center of Iran: http://www.amar.org.ir

[101] Iran Meteorological Organization: http://www.irimo.ir


Appendix A
Particle Swarm Intelligence (PSO) MATLAB Implementation

%Parameters of the PSO
MaxIterationsNo = 5000;
ParticleNum = 1000;
Dim = 4;
% if we Use Minimum or Maximum Distance with Goal we can Use:
%MinDelta = -0.0001;
%MaxDelta = 0.0001;
c1 = 2;
c2 = 2;

TemperatureData = xlsread('NaturalGasData.xls', 'Yearly NG', 'C4:C37');
PopulationData = xlsread('NaturalGasData.xls', 'Yearly NG', 'D4:D37');
NIData = xlsread('NaturalGasData.xls', 'Yearly NG', 'E4:E37');
GDPData = xlsread('NaturalGasData.xls', 'Yearly NG', 'F4:F37');
CPIData = xlsread('NaturalGasData.xls', 'Yearly NG', 'G4:G37');
DPreviousYearData = xlsread('NaturalGasData.xls', 'Yearly NG', 'H4:H37');
YearlyNGDemandData = xlsread('NaturalGasData.xls', 'Yearly NG', 'I4:I37');

TrainYear = 30;
TestYear = 4;

%Initialize the Position and Velocity of Particles
for Cnt1 = 1:ParticleNum
    for Cnt2 = 1:Dim
        ParticlePos(Cnt1,Cnt2) = rand*10;
P1articleV(Cnt1,Cnt2) = rand;
        PBest(Cnt1,Cnt2) = ParticlePos(Cnt1,Cnt2);
    end
end

%Initialize PBestFitness Array
%Because Our goal is to minimize, we use a large number for initializing
for Cnt = 1:ParticleNum
    PBestFitness(Cnt) = 1000000;
end
GBestValue = 1000000;

%Main PSO Routine
for Cnt = 1:MaxIterationsNo
    %find the fitness of each particle
    %The Fitness Function Depend On Dimension of Problem (Currently 4)
    for Cnt1 = 1:ParticleNum
        CurrentFitness(Cnt1)=0;
        for Cnt2 = 1:TrainYear
            A0 = ParticlePos(Cnt1,1);
            A1 = ParticlePos(Cnt1,2);
            A2 = ParticlePos(Cnt1,3);
            A3 = ParticlePos(Cnt1,4);
            %other calculations...
        end
        %update PBestFitness
        if CurrentFitness(Cnt1) < PBestFitness(Cnt1)
            PBestFitness(Cnt1) = CurrentFitness(Cnt1);
        end
        %update GBestValue
        if CurrentFitness(Cnt1) < GBestValue
            GBestValue = CurrentFitness(Cnt1);
        end
    end
end
% A4 = ParticlePos(Cnt1,5);
% A5 = ParticlePos(Cnt1,6);
% Different fitness functions for Models 1 - 4
MyEquation = A0 + ...
PopulationData(Cnt2) * A1 + ...
GDPData(Cnt2) * A2 + ...
TemperatureData(Cnt2) * A3;

% MyEquation = A0 + ...
% PopulationData(Cnt2) * A1 + ...
% CPIData(Cnt2) * A2 + ...
% GDPData(Cnt2) * A3 + ...
% NIData(Cnt2) * A4 + ...
% TemperatureData(Cnt2) * A5;

% MyEquation = DPreviousYearData(Cnt2) + ...
PopulationData(Cnt2) * A0 + ...
CPIData(Cnt2) * A1 + ...
GDPData(Cnt2) * A2 + ...
NIData(Cnt2) * A3 + ...
TemperatureData(Cnt2) * A4;

% MyEquation = DPreviousYearData(Cnt2) + ...
PopulationData(Cnt2) * A0 + ...
CPIData(Cnt2) * nthroot(CPIData(Cnt2) * GDPData(Cnt2) * NIData(Cnt2),3) * A1 + ...
TemperatureData(Cnt2) * A2;

CurrentFitness(Cnt1) = CurrentFitness(Cnt1) + ...
abs((YearlyNGDemandData(Cnt2) - MyEquation)) / ...
YearlyNGDemandData(Cnt2);
end
CurrentFitness(Cnt1) = CurrentFitness(Cnt1) / TrainYear;
end

% Find PBest for each particle
for Cnt1 = 1:ParticleNum
  if CurrentFitness(Cnt1) < PBestFitness(Cnt1)
    PBestFitness(Cnt1) = CurrentFitness(Cnt1);
    for Cnt2 = 1:Dim
      PBest(Cnt1,Cnt2) = ParticlePos(Cnt1,Cnt2);
    end
  end
end

% Find the global best among all of the particles
if min(CurrentFitness) < GBestValue
  [GBestValue, GBestIndex] = min(CurrentFitness);
  for Cnt2 = 1:Dim
    GBest(Cnt2) = ParticlePos(GBestIndex,Cnt2);
  end
end
% GBest contains the position of global best

% Update the position and velocity of Particles
for Cnt1 = 1:ParticleNum
  for Cnt2 = 1:Dim
PCurrent(Cnt2) = ParticlePos(Cnt1,Cnt2);
end

for Cnt2 = 1:Dim
    ParticleV(Cnt2) = 0.01*ParticleV(Cnt2) + c1*rand*(PBest(Cnt2)-PCurrent(Cnt2)) + c2*rand*(GBest(Cnt2)-PCurrent(Cnt2));
    ParticlePos(Cnt1,Cnt2) = PCurrent(Cnt2) +ParticleV(Cnt2);
end
end
GBest
GBestValue

Appendix B
Mean Absolute Percentage Error (MAPE) function Excel Implementation

mean absolute percentage error (MAPE)
Function MAPE(Original As Range, Revised As Range)
    Dim Divisor As Long
    Dim x As Long
    Dim Change As Double
    Dim TotalChange As Double
    Divisor = Original.Rows.Count
    Change = 0
    TotalChange = 0
    For x = 1 To Divisor
        Change = Abs(((Revised.Cells(x, 1) / Original.Cells(x, 1)) - 1) * 100)
        TotalChange = TotalChange + Change
    Next x
    MAPE = TotalChange / Divisor
End Function

The end.