CHAPTER ONE

INTRODUCTION

1.1 Background

Any crises in the crude oil market may cause crude oil price volatility, which has direct and indirect negative effects on the global economy and inflicts suffering on communities across the globe. The effects of crude oil volatility have no geographical boundary as there is no restriction to a specific country or region of the world. Even highly indebted poor countries, that do not contribute significantly to the world's Gross Domestic Product (GDP) (monetary value of product and services produced in a country), are highly affected by oil price volatility (Hamilton, 2011).

Globalization implies that interdependent products and their price will fluctuate when there is any change in the product price (Khashman and Nwulu, 2011). Crude oil is perhaps the commodity that exhibits such characteristics more than any other commodity. Distortion in oil markets has serious effects on other goods and services. Crude oil is the most significant commodity traded across the globe, although there is no published consensus on this. Almost all sectors of the world economy depend on crude oil. Therefore, any fluctuation in crude oil price will have a multiplier effect on the global economy (Khashman and Nwulu, 2011). Many products use oil derivatives (e.g., rubber, medicines) or depend on energy sources that use oil derivatives either for manufacturing or operations (e.g., motors).

Instability of crude oil price, partly triggered by political crises in oil producing countries, has a direct effect on almost all sectors in the global capital markets. Hence, other markets

are also affected if there is a crisis in the oil market (Rast, 2001). Crude oil production may be limited by unexpected events such as war or revolution, such as in the Middle East, which are generally viewed as exogenous (external factors) with respect to world macroeconomic conditions. Kilian (2009) stated that the effects of the decline in oil production due to unexpected events make crude oil a unique commodity. Regular short term crude oil price movements are caused by normal market forces such as US refinery capacity, Organization of Petroleum Exporting Countries (OPEC) crude oil production ceilings, and global demand and supply. On the other hand, volatility in the oil market is prompted by unexpected events such as war, revolution, earthquake, oil workers' strike, and hostage taking (Alizadeh and Mafinezhad, 2010).

Crude oil price were relatively stable at approximately US$11 per barrel for 25 years (the period of the 25 years was not provided). However, from February 1999 to September 2000 the price rose to a peak of approximately US$35 per barrel. In November 2001, the price of crude oil fell significantly, thereby slowing world economic activities. In industrialized countries, the fluctuation of crude oil price triggered the GDP to fall by 0.3% in 2001 and 2002 before it recovered. Domestic demand dropped by 0.4% due to the negative effects of reduced trade. The negative impact of crude oil price fluctuation on the United States and Europe was more devastating than that of average industrial countries. The negative impact of crude oil price volatility exerted on industrial countries was less compared to the average industrial countries because some of the big members of the group of industrial countries (UK and Canada) are net oil exporters (International Monetary Fund (IMF), 2000).

Inflation affected all countries in that period, particularly the United States and Europe, where real and nominal short term interest rates increased significantly. The increase in oil price also contributed to the decline in global growth and higher interest rates in advanced economies. Asia (China, India, Indonesia, Korea, Malaysia, Philippines, Pakistan, and Thailand) experienced the most devastating effects on economic development. Latin America (Argentina, Brazil, Chile, and Mexico), some developing European countries, and African countries (Poland, Russia, South Africa, and Turkey) were less affected by the oil shock. Despite the fact that indebted poor countries (Lao People's Dem.Rep, Sao Tome & Principe, Guyana, Mauritania, Mali, Ghana, Nicaragua, Sierra Leone, Senegal, Kenya, Ethiopia, Honduras, Madagascar, Moldova, Mongolia among others) generally characterized by transitional economies, contribute a small percentage to the world's GDP, volatile oil price also affected these indebted poor countries by negatively affecting their GDP (30 out of 40 heavily indebted poor countries) (IMF, 2000). A significant number of commonwealth countries such as Moldova, Mongolia, Kyrgyz Republic, Belarus etc. have very low per capita income, large current account deficits, high external debt, high levels of oil imports, and lower access to global capital markets (IMF, 2000). The recent rise in crude oil price impacted negatively on Korean economy as the country, mainly depends on crude oil importation from other countries. Furthermore, Korean won depreciated against US dollars, which created a burden on Korean oil importers (Yun, 2010).

Crude oil price were chosen as the subject of our research because of their global significance (He *et al*., 2012); the volatility of these price can lead into a worldwide economic recession (Jo, 2011; Hamilton, 1983), and the projection of crude oil price would be a significant contribution to overcoming its negative impact (Hamilton, 2011; Kulkar,

and Haidar, 2009). In the area of crude oil, projection remains one of the greatest challenges, and is an active research area over the years (He *et al*., 2009).

In the 1970s, when crude oil price started escalating to be higher than it was ever seen before, conventional econometric, statistical and mathematical models were the predominant methods that were used in the projection of crude oil price (Kaboudan, 2001). These methods can effectively solve only linear or near-linear problems and some complex nonlinear time-varying problems in a limited way, which cannot meet practical needs (Yu *et al*., 2008a) because these traditional methods are formulated based on classical mathematics, bivalent logic, and classical theory of additive measures which do not meet practical applications of crude oil price projection (Zalloi, 2009). These limitations have triggered growing interest in computational intelligence techniques due to their ability to handle complex problems more efficiently when compared to conventional methods (Yu *et al*., 2008b). Computational intelligence techniques, such as Artificial Neural Networks (ANN), Genetic Algorithms (GA), fuzzy logic, and, recently, hybrid intelligent systems (HIS), provide solutions to complex, nonlinear, and volatile crude oil price projection (Khan, 1998). As a result, several intelligent models have been proposed by various researchers to forecast the price of crude oil in order to counter the negative impact. However, the focus of most researchers has been on expert systems. Moreover, very few researchers have considered the effects of uncertainties in projecting crude oil price (Yu *et al*., 2005; Mehdi, 2009; Wang *et al*., 2004; Shouyang *et al*., 2005).

1.2 Neuro-Genetic model

1.2.1 Genetic Algorithm

Genetic algorithm (GA) was conceived by Holland (1975), as a biologically-inspired search method based on the principle of natural selection. In GA operations, an initial population with possible optimal solutions are generated. New populations are then created by selecting parents on the basis of their fitness, parents mingle (crossover) to create new offspring, and the strings that result from crossover are mutated in order to prevent a solution that is a local minimum. The new generated populations are used for further GA search steps, where the production of new generations (evolution) continues until a stoppage point is reached, by testing for some specified condition or set of conditions. If a stoppage condition is satisfied, the best chromosome within the current population is returned as the overall optimal solution (Haupt and Haupt, 2004; Amin, 2013). The GA has been successfully applied in oil demand (Assareh *et al*., 2010), pattern recognition (Demetgul *et al*., 2011), gas pressure (Ahmadi & Shadizadeh, 2013), classification (Amin, 2013), etc.

1.2.2 Artificial Neural Network

The ANNs are computer models constructed to mimic the functions of the human brain through parallel computation of multiple input vectors. ANNs comprise of neurons distributed in the input, hidden, and output layers. Neurons in the input layer supply inputs to neurons in the hidden layers. Signals to each hidden unit consist of the weighted sum of each input unit, and these are transformed to an output value by an activation function such as the sigmoid function. The computed output is weighted and passed forward to neurons in the subsequent layer, thereby creating a feed-forward path to the output layer. Weights

connecting two neurons in different layers are iteratively adjusted throughout the training process while inputs are fed to the network. Based on the pattern of connections between neurons in adjacent layers, the method of determining weights on the connections, and the node activation function (hidden and output layer nodes), the network is designed in a way that it can capture causal relationships between inputs and outputs in a dataset (Haykin, 2009). Some of the real world applications of the ANN include: fraud detection, financial management, risk assessment, production, etc. (Lisboa *et al.*, 2000), energy consumption (Ekonomou, 2010), wind speed (Azad *et al*., 2014), solar energy (Fadare, 2009), etc.

1.2.3 Neuro-Genetic

The pioneer work that combined ANN and GA (Neur-Genetic) was the research conducted by Montana and Davis (1989), the authors applied GA instead of back- propagation algorithms to evolve the ANN in finding the appropriate set of weights for a fixed set of ANN connections. This approach was used in order to deviate from problems associated with back propagation learning algorithms such as the probability of being stuck in local optima, long convergence time, etc. the Neuro-Genetic model can be used to determine ANN connection weights between layers, bias, selection of the most relevant inputs attributes to serve as inputs to the ANN, training of the ANN, instead of using the back propagation algorithms. The Neuro-Genetic have been applied in different domain, such as mathematics (Raja *et al*., 2013), medicine (Mantzaris *et al*., 2011), fruit classification (Fernandez-Lozano *et al*., 2013), miscible pressure (Ahmadi and Shadizadeh, 2013), gas (Wang *et al*., 2010), etc.

The difference between the ANN and Neuro-Genetic model are listed as follows:

i. ANN is an individual intelligent technique whereas Neuro-Genetic is a hybrid intelligent model that combined GA and ANN.

ii. The ANN has the possibility of being stuck in local minima whereas the Neuro-Genetic model can always escape from the local minima.

iii. Setting of ANN weights, neurons, bias, etc. is realized through trial and error when training ANN uses backpropagation algorithms whereas Neuro-Genetic model automatically searches for these parameters.

iv. Neuro-Genetic searches in a very large space to locate the best structure of the ANN with minimum error, whereas ANN searches for a limited ANN structure.

1.3 Problem Statement

Researchers and institutions perceive crude oil price volatility as a source of great concern, making the crude oil price projection a key issue, but a very difficult one to solve (Shouyang *et al*., 2005) because of the interactive effects of several factors, nonlinearity, and the dynamic nature of crude oil price in general (Bao *et al*., 2007).

Evidence indicated that crude oil price were significantly affected during the events of uncertainty, such as, during the first gulf war (August, 1990 – February 1991), the Venezuela unrest (December, 2002 – January, 2003) and the second gulf war (December, 2002 – May 2003), the Asian financial crises (July, 1997 – August, 1998), and the world financial recession (December 2007- Jun 2009). Static modeling (without retraining) as practiced by current researchers (Jammazi and Aloui, 2012; Pang *et al*., 2011; Shouyang *et al*., 2005) does not proffer practical solutions to the crude oil price projection problem. As

argued by Zhang and Zouh (2004) data models that are highly sensitive to changes in the data, require retraining as changes occur in the data.

There are very few researches in the literature that apply the expert system which incorporates the impact of uncertainties to project crude oil price while considering both regular factors and the impact of uncertainties (Yu *et al*., 2005; Mehdi, 2009; Wang *et al*., 2004; Shouyang *et al*., 2005). However, expert systems cannot handle the nonlinear, volatile, and dynamic nature of crude oil price. Expert systems require complete knowledge to perform well and new knowledge of the particular field has to be updated manually and incorporated in the system.

An uncertainty is expected to occur at any time, however, evaluating the impact of the uncertainty is extremely difficult as earlier stated, and even if a similar uncertain event occurs again at a different time, the impact this uncertainty event has on crude oil price could differ. Thus, there is a need for new methods that can handle the impact of uncertainties in different periods of time (Zhang *et al*., 2008) in the projection of crude oil price. Framing market information is a very difficult job because human experts do not completely (100%) understand the mechanisms that drive the market (Sotoudeh and Farshad *et al*., 2012). Impacts of uncertainty on a particular subject is not easy to collect, quantify, and approximate (Shouyang *et al*., 2005).

Despite the significance of attribute selection (Quek *et al*., 2008), very few studies have been found in the domain of crude oil price projection that involve attribute selection. The predominant methods used in this domain for selecting attributes are: principal component analysis (PCA), Correlation, regression, and correlation coefficient methods, whereas these

statistical methods assume normal distribution for input attributes (Su, & Wu, 2000) which makes them inappropriate for the projection of crude oil price due to its volatile nature (Zhang *et al*., 2009). Other studies use trial and error, and manual methods for selecting input attributes while trial and error methods are time consuming, laborious, and lacks justification. However, GA is viewed to be significantly better than statistical methods in selecting input attributes (Oreski *et al*., 2012). Some studies that compared one method against other chosen techniques have not proceeded further to check for statistical significant differences among the accuracies of the comparative techniques (Phichhang *et al*., 2011; Jammazi and Aloui, 2012).

In view of the limitations stated above and the capabilities of ANN and GA (refer to section 3.3 for capabilities of ANN and GA), we propose the Neuro-Genetic model as the most appropriate technique for the projection of crude oil price while considering the impact of uncertainty.

1.4 Aim of the Research

The aim of this research is to apply GA in order to optimize the parameters of ANNs and for the selection of attributes to build a Neuro–Genetic Model that can project crude oil price while considering both regular market factors and the impact of uncertainty.

1.5 Objectives

Objectives required in overcoming the limitations in the previous works are stated as follows:

i. To develop a Neuro – Genetic Model in order to project the price of crude oil.

ii. To project crude oil price while considering the impact of unexpected events using a Neuro – Genetic model.

iii. To compare the performance of the Neuro – Genetic model with support vector machine and neural network in projecting the price of crude oil.

1.6 Research Questions

To achieve the objectives of this research, the following questions require answers:

i. What are the problems of crude oil price projection models that have been proposed in the past?

ii. Which data standardization method is preferable in the domain of crude oil price projection?

iii. Can GA perform better than the statistical methods, trial and error, and manual methods in selecting input attributes?

iv. Can the proposed model capture the impact of unexpected events on crude oil price?

v. Are the crude oil price projected by the Neuro-Genetic model and the actual crude oil price equal?

vi. Can the Neuro – Genetic model performs better than the support vector machine and neural network?

vii. Is there a significant difference between the projection accuracy of the Neuro – Genetic model and the comparative methods?

1.7 Mapping the Objectives and Research Questions

The mapping between the objectives of the research and research questions are provided in Table 1. 1 to show how the research questions are connected with the objectives.

Table 1.1: Mappaing of Objectives and research questions

Objectives	Research questions
i. To develop a Neuro – Genetic Model in order to project the price of crude oil.	i. What are the problems of crude oil price projection models that have been proposed in the past? ii. Which data standardization method is preferable in the domain of crude oil price projection? iii. Can GA perform better than the statistical methods, trial and error, and manual methods in selecting input attributes?
ii. To project crude oil price while considering the impact of unexpected events using a Neuro – Genetic model.	iv. Can the proposed model capture the impact of unexpected events on crude oil price? v. Are the crude oil price projected by the Neuro-Genetic model and the actual crude oil price equal?
vi. To compare the performance of the Neuro – Genetic model with support vector machine and neural network in projecting the price of crude oil.	vi. Can the Neuro – Genetic model performs better than the support vector machine and neural network? vii. Is there a significant difference between the projection accuracy of the Neuro – Genetic model and the comparative methods?

1.8 Motivation

This research was triggered by the negative impact of crude oil volatility on communities, dependence of both private and public sectors on the projection of crude oil price for successful planning and formulation of international policy related to crude oil price as well as the influence of unexpected events on crude oil price. These factors prompted the

urgent need in order to have a relatively reliable crude oil price projection system that will offer government and commercial sectors across the globe a better understanding of the future crude oil price for strategic planning. Similarly, private sectors including oil investors can make better decisions based on the projected price, and plan ahead for proper business successes. National budget of oil producing countries such as Nigeria, Saudi Arabia, Venezuela, Kuwait among others heavily depend on crude oil price projection. Therefore, a relatively reliable projection system can assist these countries in making decisions on both national and international developmental planning. In addition, intergovernmental organization such as OPEC and Organization for Economic Co – operation and Development (OECD) require accurate knowledge of crude oil price projection for policy formulation.

1.9 Methodology

The research starts by conducting a literature review of the advances made in the projection of crude oil price, and unveils limitations in the literature. The limitations discovered in the previous works are the research problems of the present study. The problem statement was formulated based on the limitations. The research aims and objectives were derived from the problem definition. The crude oil price data requirements for the research are collected from EIAUSDE which is freely available. The data collected were cleaned and preprocessed.

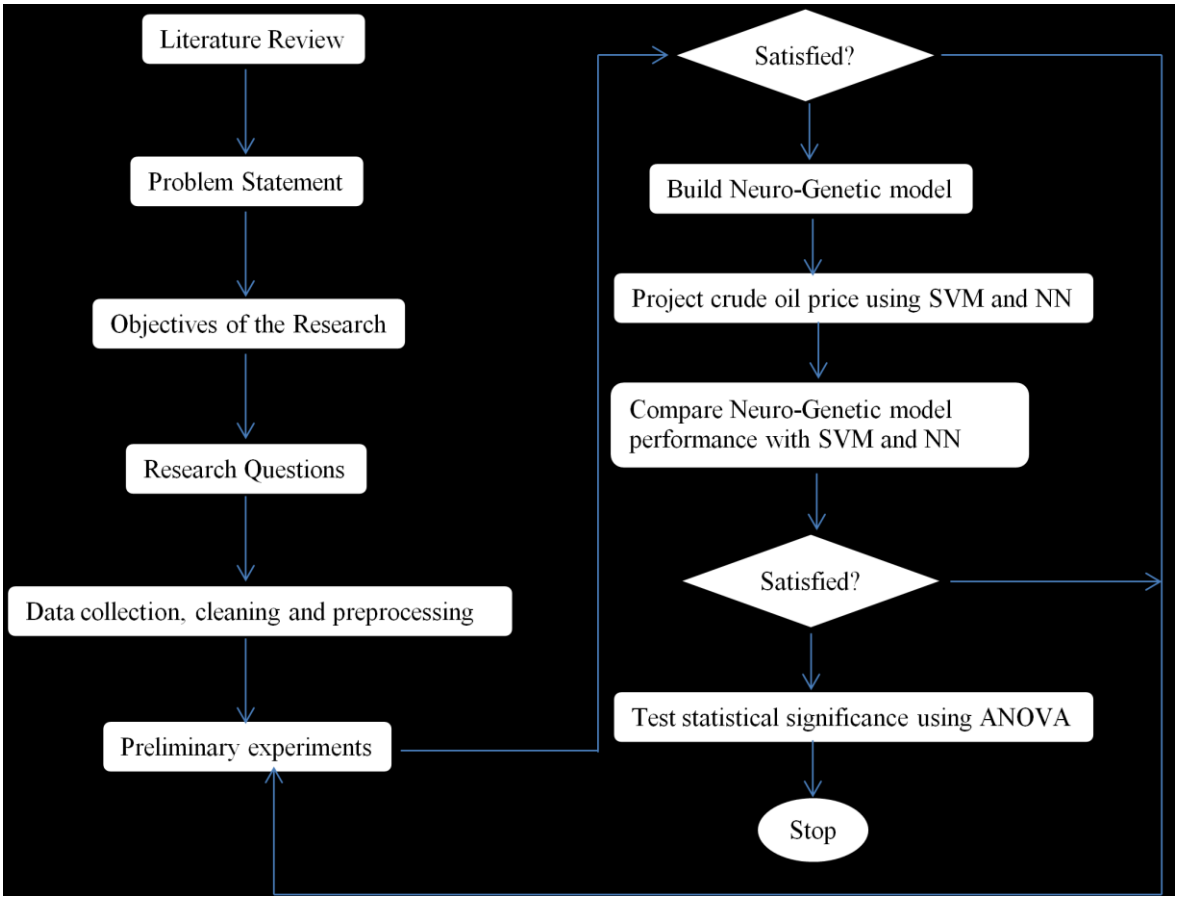


Figure 1.1: Proposed methodology of the research

Preliminary experiments were conducted with the data so as to determine the appropriateness of the data standardization method to be adopted for the study. The GA is used to select the input attributes and the optimization of ANN parameters for building the Neuro-Genetic model to project crude oil price while considering the impact of uncertainty. In this framework, the modeling will be continuous i.e retraining with relatively new data continuously so that new pattern in the new data distorted by uncertainty can be captured. The projection model can then be used for the projection of crude oil price using relatively new data with new information.

For the purpose of evaluation, a backpropagation ANN and a support vector machine (SVM) was used for the projection of crude oil price without retraining and with retraining. The results obtained were compared to that of Neuro-Genetic model. The accuracy and convergence speed of the proposed Neur-Genetic model was assessed using one-way analysis of variance (ANOVA) to explore the significance difference among the Neuro-Genetic model and the comparison methods. Also, the t-test was performed under the hypothesis that the crude oil price project by Neuro-Genetic model and the actual crude oil price are equal. The entire procedure of the proposed methodology is presented in Figure 1.1.

1.10 Scope of the Research

The focus of this research is to build a hybrid projection model using ANNs and GA. Brent crude oil price is chosen as the international benchmark. Monthly historical data (1987 – 2011) are used in building the model. Data prior, during and after each of the following unexpected events: Iran revolution 1979, Iran – Iraq war 1980, Golf war 1991, Venezuela unrest and second gulf war 2003, Asian financial crises 1997, and financial recession 2007 are used for the experiments. The models adopted for the purpose of evaluating the performance of our proposal are backprogation ANN and Support Vector Machine (SVM).

1.11 Significance of the Study

1.11.1 To the Machine Learning Community

The research unveiled an alternative approach to the crude oil price projection, which has added to the approaches already discussed in the machine learning literature. Based on

empirical evidence by (Oreski et al., 2012) the GA proposed for feature selection is expected to significantly perform better than the statistical method, trial and error, and correlation in terms of attribute selection accuracy. The approach can improve computational efficiency over the approaches proposed in the machine learning literature for crude oil price projection. The research work contributes to the present effort being made in the projection of crude oil price by proposing a more realistic approach that may actually meet practical needs.

1.11.2 To the Energy and Economy

The research can advance the crude oil price projection accuracy with results that are statistically validated. Our proposed approach with little additional information will have the potential of real life applications than the existing models. The proposed approach may generate more revenue for investors in oil and gas sectors than the technical and fundamental analysis presently use by the investors.

1.12 Thesis Outline

Chapter One

This chapter provides the background of the research, including problem statement, objectives, aims, scope, significance of the studies, research questions, motivation, and brief explanation about the propose methodology.

Chapter Two

The background concept of the crude oil market is presented in the chapter to provide solid preliminary information to the readers. The devastating impact of crude oil price

fluctuation to the economic development and hardship typically triggered in society are described in the chapter. Previous attempts to proffer solutions based on machine learning are presented in chapter 3.

Chapter Three

In this chapter, the applications of artificial intelligence techniques such as ANN, Fuzzy Neural Network (FNN), Adaptive Neuro Fuzzy Inference Systems (ANFIS), GA, expert systems etc in the projection of crude oil price are discussed. The chapter is for previous attempts made by researchers based on machine learning to project crude oil price in order to deviate from its devastating effects. The accuracies obtained and critical analysis of the studies is also covered.

Chapter Four

The theoretical framework of the study that provide an overview of the ANN and GA to give the readers the basic operations of the techniques and how they operates to achieve their optimal goals. Diagrammatical and pictorial representation of the ANN and GA concepts are also presented.

Chapter Five

Data collection, cleaning and preprocessing are explained. Arguments among researchers on data standardization methods (normalized and original data) and the selection of input attributes by GA prompted preliminary experimental investigations in this study. The preliminary experiments and results are presented for preparing the stage for implementing the Neuro-Genetic model framework. Development of the Neuro-Genetic model for the

projection of crude oil price while considering uncertainty is presented, including the pseudo-code for the Neuro-Genetic model. The conceptual framework of this research was implemented in this chapter.

Chapter Six

The chapter discussed the results obtained from the experimental description presented in chapter 5. The results include: comparing performance between raw data and normalized data. The selection of inputs attributes using GA and state of the art methods. The Neuro-Genetic model for the projection of crude oil price while considering the impact of uncertainties. Lastly, comparison of the Neuro-Genetic model performance with the SVM and the ANN is made.

Chapter Seven

Chapter seven of the thesis discusses conclusions made from the empirical findings, and further research to be conducted in the future. Contributions made by the study are highlighted.

CHAPTER TWO

BACKGROUND OF THE CRUDE OIL PRICE MARKET

2.1 Introduction

This chapter discusses unexpected events related to crude oil price volatility and the impact of crude oil price fluctuation on developing and transitional economies. Crude oil market and international crude oil price benchmarks are also well discussed in the chapter. Furthermore, the negative effects of crude oil price volatility and regular factors affecting crude oil price are explained in great detail for the readers to appreciate the oil market scenario.

2.2 Unexpected Events

Unexpected events usually occur without any warning, such as war, revolution, financial crises, terrorist attacks, political conflicts, rumors, natural disasters, earthquakes, and extreme weather conditions. These types of unexpected events when related to crude oil, have significant effects on the price of crude oil and will contribute to oil price volatility. Table 2.1 reports the uncertainties that significantly affect crude oil price. Ji (2012) classified the major factors responsible for influencing crude oil price fluctuation into six categories as follows: Macroeconomics, speculation, stock market, supply and demand, exchange rate, and commodity market. Yi and Qin (2009) identified factors such as politics and military interference as the major contributors to the volatility of crude oil price. In a study conducted by Zhang *et al*. (2009), price of crude oil before, during and after the Gulf War of 1991 and the Iraq War of 2003 were analyzed using empirical mode decomposition based event analysis. Historical data of West Texas Intermediate (WTI) and Brent crude oil

price were collected from 30 March 1990 to 31 May 1991. The aim was to assess the impact of unexpected events on crude oil price. Empirical evidence from the study indicated that crude oil price hiked up in the face of unexpected events and dropped to normalcy afterwards. In a related study, Ortiz–Cruz *et al*. (2012) depicted daily WTI crude oil price on a graph for a period of 25 years. The WTI crude oil price were collected from the Energy Information Administration of the US Department of Energy (EIAUSDE). Unexpected events affecting crude oil price were graphed by Ortiz–Cruz *et al*. as shown in Figure 2.1. Zhang *et al*. (2008) used crude oil price of WTI covering the period between 1946 and 2006 which were graphically represented as shown in Figure 2. 2.

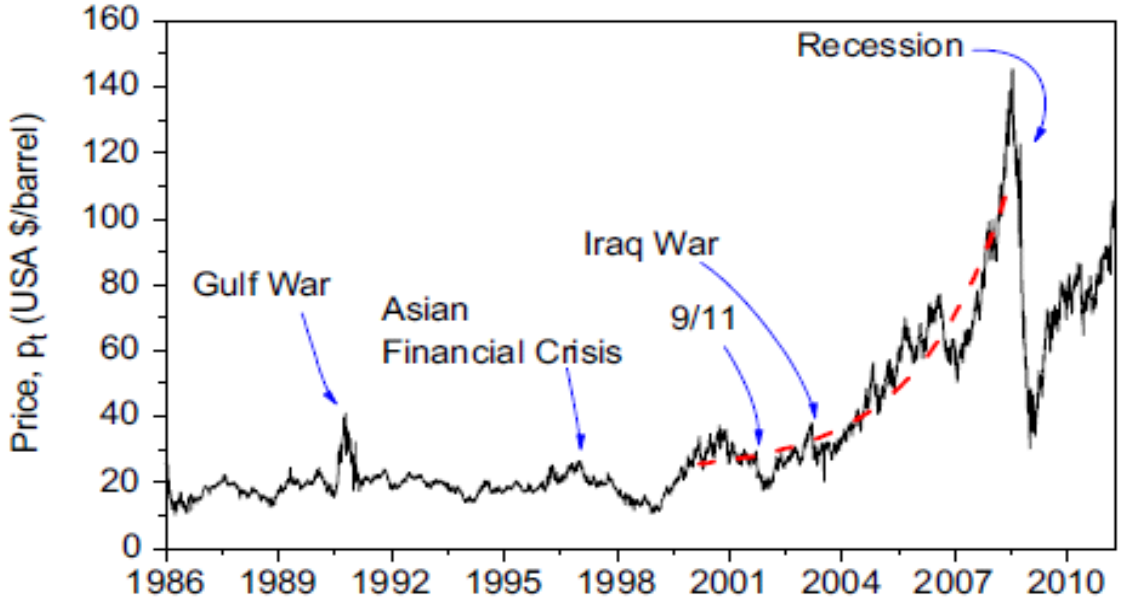


Figure 2. 1: WTI crude oil price over a period covering 1 January 1986 to 15 March 2011
Source: Ortiz – Cruz *et al*. (2012)

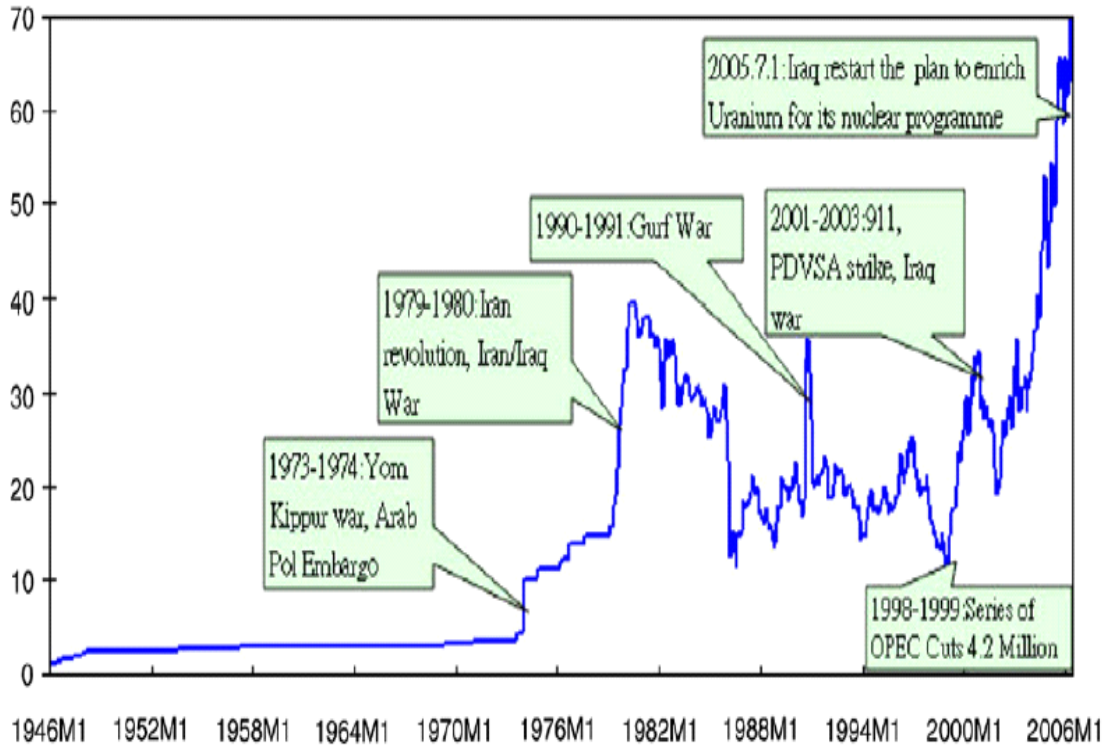


Figure 2.2: WTI crude oil price from January 1946 to May Source: Zhang *et al*. (2008)

Table 2.1: Unexpected events

Unexpected Event	Year event started	The period covered	Reference
Asian financial crises	1997	July, 1997 – August, 1998	Mitton (2002)
Gulf war	1990	August, 1990 – February 1991	Zhang *et al*. (2009)
Iraq war	2002	December, 2002 – May 2003	Zhang *et al.* (2009)
Venezuela unrest	2002	December, 2002 – January, 2003	Hamilton (2011)
World financial recession	2007	December 2007- Jun 2009	Ortiz – Cruz *et al.* (2012)

2.3 The Negative Effects of Crude Oil Price Volatility

In Hamilton (2011) the first shock on the crude oil market occurred from 1862 – 1864 as a result of the U.S civil war in which crude oil price was destabilized. In the period starting from 1952 – 1953 there was an interruption in oil supply in the world resulting from the factors which occurred within this period as presented in Table 2.2.

Table 2.2: Factors responsible for the Interruption of crude oil supply

Factors	Date
Boycott of Iran oil by the world in response to the nationalization of Iran oil industry and this succeeded in removing 19 million of Iran's crude supplies from the global oil market.	Jun, 1951
Korean war	January, 1950
Strike action of US refinery workers	April 1952
Suspension of all private flying in Canada	May, 1952
US and Britain cut down 30% supply of civil flight fuel	May, 1952

The Suez crises of 1956 - 1957: inversion of Egypt's Sinai by Israel forces in response to the Egypt nationalization of oil industry created a conflict that led to sinking down of 40 oil ships in the Mediterranean sea, blocking of the Suez canal that transported 1 – ½ million barrels of crude oil per day and prevented the shipment of oil. The pipeline that runs from Iraq to eastern Mediterranean was damaged and this created a production shortfall of 1.7 million barrels per day representing 10% of the total world supply. These unexpected events had serious economic consequences for European countries which have been depending on its 2/3 oil supply from the middle east. Some of the major consequences suffered by Europe and US are as follows Hamilton (2011):

i. Shortage of gasoline causing cut in motoring, reduction of work weeks and threats of job cuts in automobile industries.

ii. Block of rooms in hotels were closed down in order to save fuel.

iii. Driving on Sunday in Netherland, Switzerland and Belgium were banned and rationing was imposed in Britain, Denmark and France.

iv. Almost all automobile manufacturers in Britain had to reduce the rate of production.

v. Swedish leading manufacturer of Volvo had to cut down production by 30%.

vi. In Britain, 70% of automobile services were closed down.

vii. Ban on Sunday driving in Dutch, had cost hotels 85% of their usual business.

viii. U.S exportation of goods and services started declining down and this was the factor responsible for contributing to the 1957 US recession.

Attacks launched by Syria and Egypt against Israel in 1973 led Arab members of OPEC to cut oil supply to countries perceived as friends and allied with Israel, thereby reducing world oil production by 7.5%. In the same year, Persian Gulf countries double the oil price. The shortage of oil supply and increase in demand forces price of gasoline in urban cities to shoot up by 12% in 1973 and 50% in 1974. Rural areas suffer 24% and 84% increase in the cost of gasoline.

The Iranian revolution that took place between October 1978 – January 1979 reduced world oil supply by 7% and this event caused a shortage of gasoline inflicting pains on communities especially in US where long queues of vehicles were experienced. The Iran – Iraq war, which started in September, 1980, caused the shortage of oil production from

both countries to fall to almost 6% of the world oil supply. This affected oil price to double its price in 1981.

The invasion of Kuwait by Iraq in August 1990 pushed oil price to double its price within a few months and there was panic that the war may spread into Saudi Arabia. Both countries produced almost 9% of the total world oil at the time of the crisis, but gasoline queue was not experienced in the US as previously experienced during such kind of events (Hamilton, 2011).

Venezuelan unrest: the general strike action in Venezuela reduced oil production by 2.1 million barrels per day between December 2002 and January 2003.

The second Gulf war, US attack on Iraq (2003) eliminated 2.2 million barrels per day from April – July. The two events affected world crude oil supply a little bit, but not as significantly as the other unexpected events.

As a result of the Asian country's economic meltdown between 2007 – 2008 crude oil price fell down to less than $12 per barrel by December 1998, the lowest oil price since 1972 as pointed by (Hamilton, 2011).

Hike of crude oil price affects global economy in various aspects including (Hamilton, 2011):

i. Income of oil consumption will be transferred to oil producers. The tendency of losing income from oil consumers is generally higher than that of oil producers, in this case there will be a shortage of demand.

ii. Production cost in the economy will also increase as will the price of energy inputs thus affecting profit margins.

iii. Impact on price of goods and inflation. The impact depends on the extent to which monetary policy is tightened, the extent to which consumer try to adjust their actual income through an increase in wages and the degree to which producer strive for reinstatement of profit margins.

iv. Affected changes in economic activities, corporate revenue generation, inflation, equity, bond value and exchange rate of currency.

v. Based on period expected for the price increase to last, changes in the price create inducements for oil suppliers to raise production and investment and oil buyers to save.

2.4 Impact on Developing and Transitional Economies

The impact of crude oil price on individual developing countries is probably as large as that of industrial countries. However, the oil exporting countries that were hit by the 1997 – 1998 crude oil price declined, substantially benefitted from the price hike. These countries included Ecuador, Indonesia, Russia and Venezuela that were affected by the financial meltdown. On the other hand, oil importing countries significantly suffered an adverse effect, especially as dependency on crude oil has not declined to the same level to that of industrial countries (IMF, 2000). Significant relationship exists between grain price and crude oil price as described by (Chen *et al.,* 2010). Thus, the fluctuation in crude oil price can lead to fluctuation of the grain price. As such, when crude oil price increases, grain price is directly affected which contribute to the increase of grain price. This phenomenon can lead to shortage of food given that some countries cannot afford to

supply sufficient gains to the people, thereby, resulting in hunger and possibly death because food is required for survival. The increase in crude oil price directly affects the cost of crop production, transportation, price of fertilizer, and fuel as argued by Chen *et al.* (2010). For that reason, the formulation of the subsidy policy should consider these factors for inclusion in the policy issues. The hike in crude oil price can significantly contribute to unprecedented attention in the bio-fuel market, which in turn, create hunger in poor countries and the crude oil imports of the poor countries, can be affected negatively (Runge and Senauer 2007). Crude oil price, stock price, and exchange rates have a long run relation (Narayan and Narayan 2010). This means if the crude oil price increase, the stock market and the exchange rate will increase. Conversely, if the crude oil price decreases the stock market and the exchange rate will decrease.

2.5 Crude Oil Market

Settings of the market are essential in analyzing the activities of the market. The significant factors driving market activities include population and power of sellers/buyers, product type, goals among others (Gholamian, 2005). Before the 1970's, crude oil price were mostly fixed by key oil companies (Plourde and Watkins, 1998). Several alliances and organizations were set up based on crude oil. The most prominent among these organizations is OPEC, that, was set up with the objective of controlling crude oil price and world crude oil production. Decisions and announcements made by OPEC are closely monitored by government and the ordinary people (Khazem, 2007). However, the influence of OPEC in fixing crude oil price started diminishing in 1983, when market forces started taking over the influence, leaving a situation where crude oil price are determined by regular market forces (Plourde and Watkins, 1998). OPEC is a permanent,

intergovernmental organization set up in Baghdad in September 1960, consisting of five oil producing countries including: Iran, Iraq, Kuwait, Saudi Arabia and Venezuela as founding members now with headquarters in Vienna, Austria. OPEC currently constitutes the following countries as members: Algeria, Angola, Ecuador, Iran, Iraq, Kuwait, Libya, Nigeria, Qatar, Saudi Arabia and Venezuela (OPEC, 2012a). Another alliance comprising oil producing countries who are not members of OPEC and non oil producing countries, is the OECD, that was formed with the objective of fighting poverty through economic development and financial stability signed by 20 countries on 14 December, 1960. Presently, the organization constitutes 34 member countries, including: Australia, Canada, France, Germany, USA, Israel, Italy, Japan, Korea, Portugal, Spain, UK, among others. The US department of energy identified OECD countries, collective crude oil demand as one of the factors responsible for determining crude oil price (OPEC, 2012b).

2.5.1 International Crude Oil Price Benchmarks

Quality of crude oil is determined by viscosity (thickness or density) and the content of sulphur in the crude oil. Sulphur, is a natural element in crude oils and is not required by refiners because is expensive to be removed from the crude oil, which attract additional cost of refining the crude oil. Some crude oils possess high sulphur content (sour/heavy crudes) others have lower sulphur content (sweet crudes). Crude oil with lower density is referred to as light crude oil, which yielded more valuable refined products such as gasoline, among others, by a simple process of distillation. On the other hand, crude oil with high density requires coking and cracking (conversion of higher boiling fractions to lower boiling products) to be refined to the same level. The process is more complex than the simple distillation.

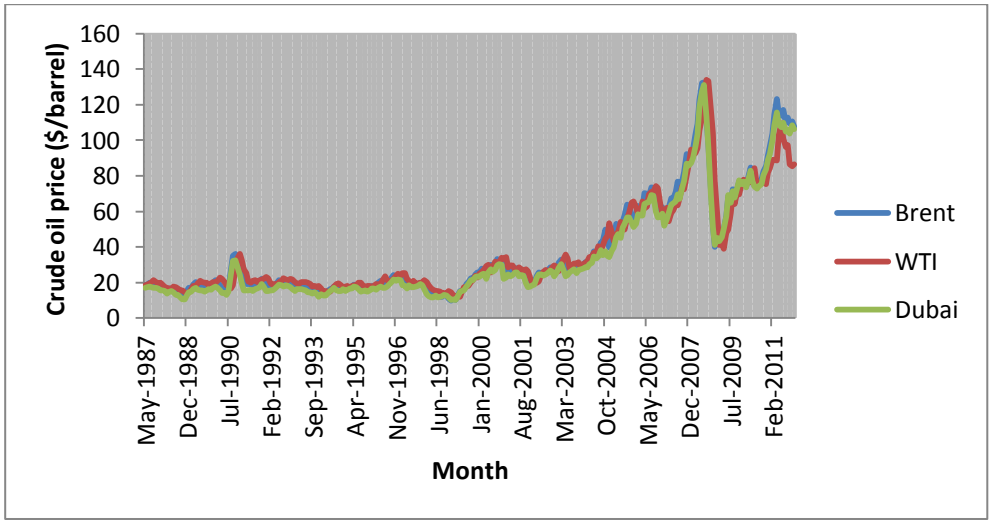


Figure 2.3: Plots of WTI, Brent and Dubai

Crude oils that are cheaper to refine and yield more valuable petroleum products are superior to crude oils that require tedious and expensive refining process to yield a lower fraction of the valuable petroleum products (light crude oils are preferred over sour/heavy crude oils) (Fattouh, 2010). The key players involved in the trading of crude oil around the world are as follows: oil producing countries, oil companies, refineries, oil importing countries and oil speculators. Moreover, shipment of crude oil takes a long period of time to be distributed in different locations across the globe. Therefore, price differ in different places in the world, depending on location (Yu *et al*., 2008c).

WTI, Brent and Dubai represent the major crude oil benchmarks for international oil pricing system at present. Please refer to Figure 2.3 for a sample data over a period of twelve (12) years. Availability of different varieties of crude oil makes buyers and sellers

price crude oil at a discounted rate with reference to certain benchmarks. Almost all crude oil traded outside the US and the far East refers to Brent as a benchmark while in the US, the WTI is referenced as the benchmark. In the Middle East, Dubai is the reference benchmarks for exporting crude oils to Asia. WTI trades light/sweet crude oils as well as Brent but WTI sulphur content (0.24%) is slightly less than the Brent sulphur content (0.37%) making it more superior to Brent. Dubai trades sour/heavy crude oil containing 2% of sulphur in the crudes (Fattouh, 2010). Brent combined crude oil from 15 separate crude oil fields located in the North sea (Charles and Darne, 2009). Fig. 2.3 is a plot of weekly crude oil price from WTI, Brent and Dubai international crude oil price benchmarks. Two – Third $(2/3)$ of the world crude oil exchange uses Brent crude oil market to specify price (Chavillon and Rifflart, 2009) while others use the WTI and Dubai. Several countries depend on the international crude oil price benchmarks for formulating development plans and estimation of expected government revenue. For example, in Oman, eighty six percent (86%) of the country's revenue generated comes from crude oil sales. Almost half of Kuwait's GDP is from crude oil sales and the crude oil sales account for approximately seventy percent (70%) of the government revenue generation. Approximately, ninety percent (90%) of Saudi Arabia's economy depends on crude oil sales. Oil and gas account for fifty two (52%) of Russia's national budget, and more than seventy percent (70%) of Russia's total export. In Iran fifty (50) to sixty percent (60%) of government revenue is derived from crude oil, whereas in Nigeria, ninety six percent (96%) of the revenue is from crude oil sales etc. (EIAUSDE, 2013).

2.5.2 Regular Attributes Affecting Crude Oil Price

In this study, we consider Brent crude oil as the international benchmark price because the majority of the world refers to the Brent for price formulation. The major factors that determine the fluctuation of the Brent crude oil price as revealed in the literature are presented in Table 2.3. The first column in Table 2.3 is the classes in which the main factors influencing the crude oil price are classified by researchers (Ji, 2012). The second column presents the main factors responsible for the fluctuation of the crude oil price with corresponding abbreviation and reference in the third and last column, respectively.

Table 2.3: Regular factors affecting crude oil price

Classification	Factors responsible for the price fluctuation	Abbreviation	Reference
Supply	OPEC crude oil production	OPECCP	Chevillon and Rifflart (2009); Alizadeh and Mafinezhad (2010)
Inventory	OECD crude oil ending stocks	OECDES	Abdullah and Zeng (2010)
Supply	World crude oil production	WCOP	Yu *et al.* (2005)
Supply	Non OPEC crude oil production	NOPECCP	Abdullah and Zeng (2010), US Energy Department website, Chevillon and Rifflart (2009)
Supply	US crude oil production	USCOP	Alizadeh and Mafinez (2010)
Demand	US crude oil imports	USCOI	Abdullah and Zeng (2010)
Demand	OECD crude oil consumption	OECDCOC	Alexandridis and Livani (2008); Abdullah and Zeng (2010), Chevillon and Rifflart (2009)
Inventory	US crude oil stocks at refineries	USCOSR	Alizadeh and Mafinez (2010)
Inventory	US gasoline ending stocks	USESTG	Alizadeh and Mafinez (2010)
Supply	US crude oil supplied	USCOS	EIAUSDE website

The attributes presented in Table 2.3 were found to influence the fluctuation of international crude oil price as reported by EIAUSDE (2013). Table 2.3 only presented the attributes and abbreviations for each of the attributes. The attributes used in our crude oil price projection model were selected using GA (refer to section 5.6 and 6.2.4 for details). However, the descriptions of how the attributes influence crude oil price are discussed as follows:

OPEC Crude Oil Production

OPEC members' oil production is strongly influenced by the organization, which controls the rate of oil production for each member country. History has shown that whenever OPECCP is reduced, the international oil price increases and if OPECCP increases, the international oil price reduce. This clearly indicates the influence of OPEC on international oil price, measured in thousand barrels per day (tb/d).

OECD Crude Oil Ending Stocks

OECDES is the OECD countries' inventories of crude oil reserved for use in the future. The stocks are normally reported on the last day of the period, either weekly or monthly depending on the country's priority, measured in million barrels (mb/d). Market analysts testified that whenever OECD inventories are released, the price of crude oil rise if the inventories of the OECD decline and vice versa.

World Crude Oil Production

WCOP is the volume of crude oil produced globally by non-OPEC, OPEC, OECD and non-OECD countries to various reservoirs for transfer through ships, pipelines and trucks to different locations in the world, measured in tb/d. The price of the crude oil typically

arise when the world production of the crude oil is reduced due to slow economic growth or occurrence of uncertain event related to crude oil production or transportation of the crude oil.

Non-OPEC crude oil production

NOPECCP are countries outside OPEC, currently accounting for 60% of the world's oil production. Major region of NOPECCP include North America, former Soviet Union regions and the North Sea. NOPECCP countries make independent decisions on their production, measured in tb/d. The NOPECCP production is significant to the world crude oil supply, the international crude oil price increases or decreases with respect to the decrease or increase of the production of crude oil by the NOPECCP.

US Crude Oil Imports

USCOI is the importation of crude oil and petroleum products from foreign countries, Puerto Rico, Virgin Islands and territories into the 50 states of the US and District of Columbia, measured in tb/d. The US imports crude oil from different part of the world including OPEC countries. Uncertainty in an OPEC country affects US crude oil import's. If the US economy is in recession, crude oil import's of will definitely reduce, thereby reducing demand, which makes the price of crude oil to fall. A boom in the US economy contributes to the rise in crude oil price since demand typically increases when the US economy is booming.

OECD Crude Oil Consumption

This refers to the total number of barrels consumed in OECD countries (53% of total world consumption). The OECD rate of crude oil consumption is greater than non-OECD

countries', but their consumption growth is slow. In each country's economy, structural conditions have a significant impact on the correlations among crude oil price, economic growth and crude oil consumption, measured in tb/d. As such, the reduction in OECD oil consumption directly affects the economic growth through which crude oil price will fall and vice versa.

US Crude Oil Stocks at Refineries

USCOSR is the US inventories of crude oil will reserve for use in the future. The stocks are normally reported on the last day of the period, either weekly or monthly depending on the stipulated date, measured in tb/d. Any time the US crude oil inventories in the refineries decrease, the price of crude oil typically increase and vice versa. Energy investors and analysts await the reports of the US crude oil inventories typically released by the EIAUSDE before taking a valid conclusion on possible price of crude oil.

US Gasoline Ending Stocks

USESTG is the US inventories of gasoline reserved for use in the future. The stocks are normally reported on the last day of the period (weekly or monthly) depending on the chosen last day, measured in tb/d. Gasoline affect the inventories of refineries because of its critical role in transportation. As a result, the gasoline inventories drive the crude oil demand, which increase or decrease crude oil price. The trend of gasoline is monitored by energy investors in order to project the possible direction of future crude oil price.

US Crude Oil Supplied

USCOS is the total amount of crude oil supplied by the US from field production, refineries production and imports, measured in tb/d. The production of crude oil in the US affects the price of crude oil. A decrease in US crude oil production increases the price of crude oil. Crude oil price increases as the US crude oil production decreases. On the other hand, the price of crude oil decreases in the event where the production capacity of the US increases.

2.6 Summary

Unexpected events are defined in relation to crude oil price fluctuation and volatility. Unexpected events that affect crude oil price are depicted in Figures 2.1 – 2.2 for easy identification of fluctuations caused by the events. It was revealed that the volatility of crude oil price has negative impacts on developed, developing and poorly indebted countries (IMF, 2000). In this chapter, crude oil market was discussed, including international crude oil price benchmarks such as Brent, WTI and Dubai. Based on the literature, it can be concluded that the Brent crude oil price is the most widely used benchmark for pricing crude oil. Regular market factors (OPECCP, OECDES, WCOP, NOPECCP, USCOP, USCOI, OECDCOC, USCOSR, USESTG, USCOS) influencing the crude oil price fluctuation are listed and described in detail in section 2.5.2. In addition to these regular factors, unexpected events such as gulf war, Venezuela unrest and second gulf war, Asian financial crises, and world financial recession significantly affect crude oil price volatility.

CHAPTER THREE

RECENT ADVANCES ON THE APPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNIQUES IN THE PROJECTION OF CRUDE OIL PRICE

3.1 Introduction

In this chapter, the methods employed to search the literature required for our study is briefly described. The HIS and Individual Intelligent System (IIS) is introduced and their application in the projection of crude oil price is discussed in detail including a critical analysis of the literature. The strengths and weaknesses of the previous works in this domain are highlighted. The literature is analyzed based on data collection frequency, weaknesses, strengths, and techniques chosen for comparison purposes.

3.2 Method Employed in the Literature Search

The literature review was carried out in two distinct phases in order to identify published articles on Artificial Intelligent Techniques (AIT), which were specifically applied in either hybrid or individual form to project crude oil price,. First, the body of existing literature was retrieved online through the ACM Digital Library, IEEEXplore, Science Direct, Scopus, Springer Link, Web of Science, Google Scholar, Direct Open Access Journals (DOAJ), Microsoft Academic Search, ProQuest and CiteSeerX. Subsequently, in the second search phase, every article initially retrieved was analyzed and reviewed to ensure it met the required criteria for selection before it was included in the literature review. The criteria require that each article had to contain an empirical description of applied AI techniques in hybrid or individual form to predict crude oil price.

3.3 Hybrid Intelligent Systems vs. Individual Intelligent Systems

All AIT have limitations (Editorial, 2009) despite their effectiveness in solving linear, nonlinear, and complex problems, such as the projection of price in financial time series. The AIT include, but are not limited to ANN, GA, SVM, expert systems, etc.

As an example, the most widely used ANN, namely, the back-propagation ANN, could become stuck in local minima and/or over-fit the training data. Unfortunately, there is no ideal framework for determining the optimal structure of a neural network or the selection of the initial training parameters. Researchers employ cumbersome trial-and-error techniques to determine the optimal ANN structure and selection of the parameters (Bahrammizaee, 2010).

GA, another AIT, despite its successes in various applications, have the following limitations, the GA fitness function is difficult to evaluate; the problem representation definition in the GA can be difficult; the occurrence of early convergence of GA is also susceptible to converge towards the local optima; no specific way to determine parameters, such as, the population size, mutation rate, crossover rate, and number of generations; difficulty in putting precise information in a problem; performs poorly in detecting local optima; does not have any real terminating criterion and difficulty in locating the precise global optimum (Sivanandam and Deepa, 2008).

An expert system only follows if/then rules which need to be modified manually, thus, the expert system requires regular updates, and the output cannot be improved from experience. Additionally, the expert systems lack the capability of identifying nonlinear relationships (Bahrammirzaee, 2010). Expert systems do not tolerate missing or erroneous values. They execute rules to perform actions, and the rules must be stored in a knowledge

base before they can be processed. The knowledge base is manually updated by a knowledge engineer, starting from the human expert knowledge, hence, when there is a change in the expertise, the system cannot effectively handle these changes. Increased experience does not improve the system's output, hence, the overall system performance (Niculescu, 2003).

Fuzzy systems, lack the capability of learning from the input data as it uses linguistic variables in the form of the human language to represent the input and output of the systems. Thus, incomplete or wrong rules are not well handled by fuzzy systems and tuning of the systems is not a direct task (Bunke and Kandel, 2000).

Some of the limitations of IIS are eliminated through hybridization to achieve a synergistic effectiveness in the design of an intelligent system (Editorial, 2009). The HISs are systems that fuse two or more AITs to solve complex and challenging problems. HISs are created to (Lertpalangsunti, 1997): enhance performance, provide opportunities for multiple applications of tasks, and improve the capability of handling multiple functions. However, IISs are standalone intelligent techniques that are applied for problem solving without combining with other intelligent techniques, and their performances are inferior to HISs in terms of accuracy (Bahrammirzaee *et al*., 2011).

3.4 Applications of the Artificial Intelligent Techniques in the Projection of Crude Oil Price

3.4.1 Conventional Methods

As stated earlier, crude oil price have mainly been projected using conventional methods. For example, Abramson and Finizza (1991) proposed a probabilistic model to project the price of crude oil. Barone-Adesi *et al*. (1998) projected crude oil price using a semi-

parametric technique. Morana (2001) proposed a semi-parametric statistical tool to project crude oil price using the Autoregressive Conditional Heteroskedasticity (GARCH) properties of crude oil as independent attributes. The Root Mean Square Error (RMSE) obtained by the proposed model was 1.394218 between the actual and projected crude oil price. Mean Squared Error (MSE) is the commonly used measure to assess prediction accuracy, RMSE is the square root of the MSE. The closer the value of RMSE or MSE is to zero (0), the better is the prediction accuracy. Zero (0) indicates a perfect prediction, which rarely occurs in practice. Ye *et al*. (2006) proposed a model with econometrics to project crude oil price based on OECD petroleum inventories. The crude oil price was projected with an accuracy of 1.7441 RMSE.

On the other hand, these traditional statistical and econometric methods cannot effectively solve nonlinear problems (Yu *et al*., 2009) (please refer to chapter 1 section 1.1 for details on limitations of the econometrics and statistical methods). In addition, documented evidence in Bahrammirzaee (2010) reveals that AITs are superior to these conventional methodologies in projecting nonlinear and volatile financial time series, but, there are very very few cases in which conventional methods perform better than the AITs.

However, these conventional methods are still relevant in the domain of crude oil price projection. For example, Shouyang *et al*. (2005) used the Autoregressive Integrated Moving Average (ARIMA) to model the linear constituents of the crude oil price time series. Liu *et al*. (2007) defined a Markov chain as input to a Fuzzy ANN (FNN), to improve the projection accuracy of crude oil price. He *et al*. (2009) used a moving average to eliminate noise in a crude oil price time series during data pre-processing. Phichhang *et al*. (2011) hybridized GARCH and ANNs to model the volatility of crude oil price. Apart

from the statistical and econometric methods, there are other techniques that are gradually gaining popularity in this domain, especially wavelet analysis, which have drawn unprecedented interest in their hybridization with AIT to improve the projection accuracy of crude oil price (2007; Jinliang *et al*., 2009; Pang *et al*., 2011; Jammazi and Aloui, 2012; Mingming and Jinliang, 2012; Bao *et al*., He *et al*., 2012;).

On the other hand, the wavelet has limitations, for example, the use of different wavelet families in the decomposition of non-stationary signals introduces estimation bias (estimation bias is the difference between the target values and the predicted values) based on the influence of individual wavelet families. As such, in view of the fact that the wavelet families can be 2, 3, 4, or more, the estimation bias for each of the family introduce in a final prediction value can give more error. This may have a disastrous effect on the final results (Abramovich *et al*., 2000). In discrete wavelet transform, translational invariance (translational invariance occurred when elements, object, etc. shift by the same amount to another position or a data transform into wavelet without any lost of contain) is lacking (Abramovich *et al*., 2000).

3.4.2 Neural Network

Kaboudan (2001) developed Genetic Programming (GP), naïve Random Walk (RW), and backpropagation ANN projection models using monthly closing price of crude oil covering a period from January 1993 to December 1998 and sourced from the EIAUSDE. The projection accuracies indicated the MSE for GP, ANN, and naïve RW were 0.24, 1.29, and 0.91 respectively. Conversely, the techniques in the study were not hybridized to improve effectiveness since hybrids are superior to IIS in performance when properly designed. Uncertain factors were also not included in the modeling.

Reudys (2005) proposed a backpropagation ANN to extract visual features from crude oil price. The data used to build the model dated from November 1993 to January 2005 and were collected from EIAUSDE. The data were for WTI, Natural Gas–Henry, Fuel Oil No. 2 (NY), Gasoline, Unld Reg. Non–Oxy (NY) and the American Stock Exchange. The results indicated that the highest upsurges and declines in crude oil price could be projected using ANN as shown in Figure 3.1. However, the researchers fail to include uncertain factors in the projection model. Furthermore, the backpropagation ANN used in building the model was not hybridized with another optimization algorithm to improve the robustness of the ANN.

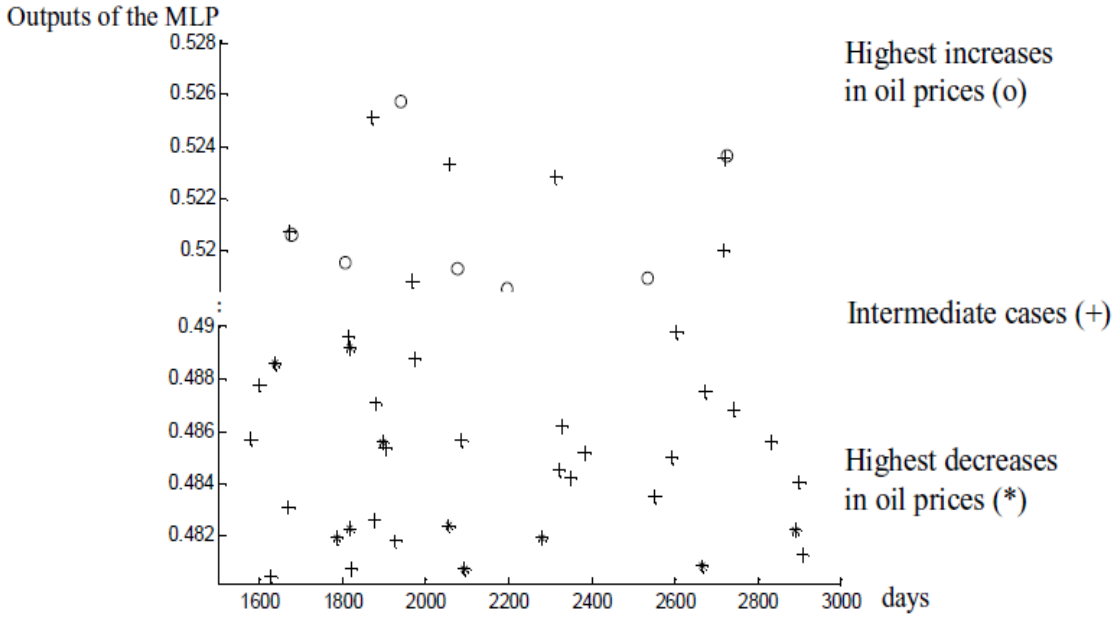


Figure 3.1: Visualization of the projection results by ANN. Source: Reudys (2005).

Moshiri and Foroutan (2006) obtained the data on crude oil price required to build a projection model from NYMEX covering a period from 1983 to 2003. ARIMA, GARCH

and backpropagation ANNs models were used to project crude oil price. The projection results showed that the models confirmed the effectiveness of the ANN models over ARIMA and GARCH in terms of projection accuracy, especially in a volatile time series. Nevertheless, the limitations highlighted by (Reudys, 2005) are also observed in the works by Moshiri and Foroutan (2006).

Malik and Nasereddin (2006) used several models to project quarterly growth of GDP based on crude oil price. The models were simple RW, auto regressive, linear with lagged oil (lagged is a model for crude oil time series dataset through which the regression equation was applied for projecting crude oil price) and GDP, cascaded ANN with GDP only, cascaded ANN with GDP and lagged oil and ANN with GDP and lagged oil. Crude oil data for modeling were collected from the Bureau of Labor Statistics and GDP data were collected from the Bureau of Economic Analysis for the period from 1947 to 2004. All models were used for one-quarter-ahead-projection. The Mean Square Error (MSE) of Simple RW was 0.0000408 while Auto regressive was 0.0000238, Linear with lagged oil and GDP was 0.0000401, Conventional ANN with GDP and lagged oil was 0.0000330, Cascaded ANN with only GDP was 0.0000241, Cascaded ANN with GDP and lagged oil was 0.0000228. However, the model did not consider the impact of uncertain factors. In addition, the ANN lack convergence speed and suffered from the possibility of being trapped in local minima due to its iterative nature of finding the minimum error. Also, significant attributes that affect the crude oil price fluctuation such as OECD Crude Oil Consumption, OPEC Crude Oil Production, (Abdullah and Zeng, 2010) and WCOP (Shouyang *et al*., 2005) were not included in the study.

Shambora and Rossiter (2007) developed a crude oil trading system for generating buy and sell signals based on projections produced by ANNs incorporated into the system. Data used in this study consisted of crude oil futures contracts on NYMEX extracted from Datastream, covering a period from April 1991 to December 2003. Among the selected trading strategies where buy and hold, technical analysis, naïve RW and returns on treasury bills. The experimental results suggested the cumulative returns of ANN over six (6) years was 200%, for buy and hold, technical analysis, naïve RW and returns on treasury bills were 61%, 73%, 100%, and 22% respectively. However, the limitations noted in the work of Shambora and Rossiter (2007) are similar to that of Reudys (2005). In addition, the comparison between the ANN and the comparison methods (buy and hold, technical analysis, naïve RW and returns on treasury bills) cannot be considered as fair because documentary evidence in the literature overwhelmingly proved that the AIT performs better than the econometrics and statistical methods (Bahrammizaee, 2010; Zhang *et al*., 1998).

Lackes *et al*. (2007) aimed at projection of crude oil price in short, medium and long term periods using backpropagation ANNs and data were collected over a period from 1999 to 2006. The experimental result established that ANNs performed poorly in short term projection, but could project future price of crude oil in the middle and longer periods of time. It was observed in (Lackes *et al*., 2007) that the researchers ignored to include uncertain factors in the projection model. Furthermore, the backpropagation ANN used in building the model was not hybridized with another optimization algorithm to improve the robustness of the ANN.

Haidar *et al*. (2008) used an ANN model to project the next day and multiple of the day's ahead (two or more days into the future) price of crude oil using crude oil future data and market data. Simulated results favored multiple day's ahead projection over the next day. The backpropagation ANN used in the study was susceptible to being stuck in local minima and factors of uncertainty were not incorporated in the model despite evidence that shows the uncertain factors have an impact on crude oil price (Hamilton, 2010).

Malliaris and G. Malliaris (2008) experimented with projecting one-month-ahead spot price of crude oil (CL), heating oil (HO), gasoline (HU), natural gas (NG), and propane (PN) in view of the fact that their spot price (i.e the price of a commodity at a particular period of time) in the market were interrelated. The spot price data were extracted from Barchart (www.barchart.com) for the period of January 1994 to December 2002. A multi-linear regression, a backpropagation ANNs model and a simple model (unfortunately the author has not specified the simple model) was applied in each of the energy markets. One month ahead price were projected by each model for each market. In the CL market, which is our focus, the MSE of ANNs was 2.269, for a multi-linear regression, it was 6.653 and for the simple statistical model, it was 6.013. However, the shortcomings identified in (Haidar *et al*., 2008) were also noted in (Malliaris and G. Malliaris, 2008).

Quek *et al*. (2008) proposed an Elman neural network (ENN) to project price of gold, crude oil and currency. Sample data were collected from Bloomberg and Datastream databases covering a period from 2000 to 2002. Multilayer backpropagation ANN and Hop field network using Hebbian learning rule were also used to project the price of the mentioned commodities for the purpose of comparison. The projection accuracy of the ENN was 65%, for FFNN and Hop field network using Hebbian learning rule were 48%

and 30% respectively. Still, for the ENN to achieve optimal performance, a high number of hidden neurons were required in the hidden layer than necessarily needed for learning the problem which was considered as a limitation because it reduced the operational efficiency of the ENN (Quek *et al*., 2008) and further increased the structural complexity of the ENN. In addition, the limitations identified in (Haidar *et al*., 2008) were also noted in the study by Quek *et al*. (2008).

Yu *et al*. (2008c) proposed an artificial intelligent agent-based fuzzy ensemble projection model which integrated a support vector machine (SVM), Radial Basis Function ANN (RBFNN) and backpropagation ANN. The sample data for their study were collected for WTI and Brent crude oil spot price covering a period from January 2000 to December 2007. Each of the individual models were used for crude oil price projection, and the results showed that the Normalized Mean Square Error (NMSE) of ANN was 0.0154, SVM was 0.0083, RBFNN was 0.0077, the simple average was 0.0068, and the computational intelligent agent-based fuzzy ensemble projection was 0.0057. Uncertain factors were absent in this study and the techniques have the possibility of being stuck in local minima, slow in convergence speed, and determining of the optimal structure was difficult.

Kulkarni and Haidar (2009) collected WTI crude oil price and future contracts traded on the NYMEX from September 1996 to August 2007 sourced from the EIAUSDE. The two models of RNNs and multilayer ANNs (MLNN) were considered in the study in which the preliminary result showed that the MLNN has an MSE of 0.00038 and the Recurrent ANN (RNN) took eight (8) hours to converge far more than the MLNN as pointed out in the study. However, the convergence speeds of MLNN and MSE of RNN were not reported.

The three day moving average was used to remove noise in the data so as to improve the projection accuracy. Crude oil spot price were projected for three days ahead using the MLNN. The one day projection accuracy achieved was 78%, 60% for two days and 53% for three days. Uncertain events were not included in the research and limitations associated with the study conducted by Haidar *et al*. (2008) were observed in the study of Kulkarni and Haidar (2009) as well because the back propagation algorithm was used for the model optimization and was not hybridized to reduce the effects of these limitations.

Pan *et al*. (2009) developed a multilayer ANN with four different projection models including: from the spot price to spot price, from future price to spot price, from spot price and future price to spot price, from spot price and lead markets to spot price in order to project short term spot price of crude oil with different indicators. All the models were tested with out-of-sample data of crude oil price. The projected results obtained from time (t) +1 was 79.95% accurate, followed by 69.74% and 60.64% for t+2 and t+3 days respectively. The experimental data for crude oil price were collected for WTI and futures contract traded in New York Mercantile Exchange (NYMEX) and was sourced from the EIAUSDE covering the period from January 1996 to August 2007. Despite the accuracy obtained, the multilayer FFNN models were liable to the shortcomings of the backpropagation ANNs and the impact of uncertain factors were not incorporated in the modeling process.

Torben (2010) compared the projected performance of ARIMA, structural vector error correction and ANNs regression. The data used to build projection models were quarterly crude oil price for Freight on Board (FOB) covering a period from the first quarter of 1986 to the fourth quarter of 2009. All three models were applied to project crude oil price and

their performances compared. The results indicated that backpropagation ANN is better than structural vector error correction and ARIMA. The limitations indicated in Haidar *et al*. (2008) were also observed in Torben (2010).

Wang and Yang (2010) examined the probability of projecting crude oil, heating oil, gasoline and natural gas futures markets within a day by conducting an experiment with different models (ANNs, semi parametric function coefficient, nonparametric kernel regression and GARCH) with data collected for 30 minutes Intraday price and returns of the four energy futures contracts sourced from NYMEX. For each of the four individual future contracts, 15 to 20 year price of future contracts were analyzed, a period during which price were low and in steady decline (bear market), another period where price were high and on a steady increase (bull market) were identified. The result indicated that only heating oil and natural markets possessed the possibility of being projected within a day, especially under bull market conditions. Uncertain factors were not considered in the research despite having an impact on crude oil price. The comparison of ANN with statistical, and econometric models might not be considered as a fair comparison since ANN have the capability to approximate any nonlinear function, whereas econometric and statistical methods can only solve nonlinear problems in a limited way as previously discussed.

Pang *et al.* (2011) used wavelet ANN, linear relative inventory and nonlinear relative inventory models to project one, two and three months ahead price of crude oil based on data collected for the OECD inventories and WTI crude oil price. Both data were obtained from the EIAUSDE covering a period from January 1992 to August 2006. The subsequent analysis of experimental results revealed that the wavelet ANN model performed better

than the linear relative inventory and nonlinear relative inventory models in terms of projection accuracy. Unfortunately, the numerical values of the performance indicators were not reported in the study. The study has similar limitations noted in (Haidar *et al*., 2008).

Movagharnejad *et al*. (2011) used ANN to project price differences in crude oil price among five (5) selected countries (Iran, Saudi Arabia, Kuwait, Oman and the Emirates) in the Persian Gulf region. Data for the study were collected from the American Petrochemical Institute and the EIAUSDE for a period from January 2000 to April 2010. Results indicated that the ANNs project the price differences in crude oil price among the countries with an average error (error between the actual and projected price) of 8.82%. Nevertheless, the researchers ignored to include uncertain factors in the projection model. Furthermore, the backpropagation ANN used in building the model was not hybridized with another optimization algorithm to improve the robustness of the ANN.

Mehrara *et al*. (2011) proposed two models in their work for crude oil price projection, namely Group Method of Data Handling ANN (GMDHNN) and multi-layer backpropagation ANN (MLNN). Each projection model was built based on data covering a period from January 2002 to July 2009 collected from the EIAUSDE. Empirical results indicated that the accuracy of the models were eighty three percent (83%) and seventy three percent (73%) for GMDHNN and MLNN, respectively. Unfortunately, the uncertain factors were not incorporated in the model. The GMDHNN and MLNN were used as IIS whereas HIS was more effective than the IIS as shown in the literature (Bahrammirzaee, 2010).

Haidar and Woiff (2011) used ARIMA, GARCH and backpropagation ANNs models in their study. Daily crude oil/return spot price of WTI were collected for a period from January 1986 to March 2010 sourced from the EIAUSDE. A series of nonlinearity tests were performed on the collected data and crude oil time series was identified as non-linear, thus, they focused on backpropagation ANNs because ARIMA and GARCH were not suitable tools to handle the nonlinear problem. Crude oil price were projected using ARIMA and ANN models and the results indicated that the RMSE of ANN was 0.0066 and for ARIMA, it was 0.025. However, the ARIMA and GARCH models possessed the possibility of projecting crude oil price provided that the noise in the data was smoothened. The limitations identified in (Malliaris and G. Malliaris, 2008) were similarly observed in this study.

Sotoudeh and Fershad (2012) used the multilayer perceptron ANN with data of US gas price collected from the EIAUSDE for a period from 1949 to 2010. Projection results indicated that the MSE was 0.78. However, Sotoudeh and Fershad (2012) failed to incorporate the impact of uncertain factors in their model. In addition, the ANN lack convergence speed and suffered from the possibility of being trapped in local minima due to its iterative nature of finding the minimum error (Zweiri *et al*., 2003).

3.4.3 Wavelet Transforms and Artificial Neural Network

Wavelet transformation is gradually gaining popularity in this domain, which has drawn unprecedented interest in their hybridization with ANN to improve the forecasting accuracy of crude oil price. According to Jammazi and Aloui (2012), wavelet analysis is an advancement in the area of harmonic analysis. Wavelets have the capability of projecting

data into the time scale domain and conducting multi-scale analysis. Wavelets are also used to capture both smooth and low-frequency data and detailed and high-frequency data. Examples of studies where wavelet analysis was used as a hybrid of wavelet analysis and ANN in various ways are as follows. Bao *et al.* (2007) applied wavelets to decompose crude oil data before applying a Least-SVM to capture useful information in various partition segments of the dataset. Jinliang *et al.* (2009) used wavelets to decompose crude oil price data into approximate and random constituents. They then applied a ANN to forecast the approximate constituent, which represented the trend of the oil price. Pang *et al.* (2011) used wavelet analysis to decompose and incorporate several layers within the data. They then applied a ANN to reduce estimation bias introduced by the wavelet analysis to improve the crude oil price forecast accuracy. Pang *et al.* (2011) applied a wavelet ANN to forecast crude oil price. The ANN uses orthogonal wavelets as an activation function instead of a sigmoid activation because the orthogonal wavelet requires less iteration to converge. Jammazi and Aloui (2012) decomposed the non-stationary nature of crude oil data using wavelet analysis and then incorporated it into a ANN to forecast crude oil price. Mingming and Jinliang (2012) used wavelet analysis to capture multi-scale characteristics of crude oil price data, where RNN was used in different partitions of the dataset to forecast price. The numerous forecast results were recombined by a back-propagation ANN to produce an ensemble forecast. He *et al.* (2012) integrated wavelet analysis and a ANN to improve the reliability and accuracy of crude oil price forecasting during the modeling process. However, the wavelet approach has its own set of limitations. For example, in discrete wavelet transform, translational invariance is lacking. The use of different wavelet families in the decomposition of non-stationary signals

introduces estimation bias based on the influence of individual wavelet families, which may have a disastrous effect on the final results as earlier discussed (Abramovich *et al*., 2000).

3.4.4 Support Vector Machine

SVM was proposed by Xie *et al*. (2006) to project crude oil price using monthly crude oil spot price of WTI for a period from January 1970 to December 2003. ARIMA and back-propagation ANNs were also used to project crude oil price for the purpose of comparison. The performance indicator showed that the SVM RMSE was 1.8210 while the RMSE of ARIMA and backpropagation ANN were 1.9037, and 1.8534 respectively. However, the sensitivity of the SVM to free parameters (i.e the influence of kernel functions, support vectors, constant C, etc. on the performance of the SVM) was high, and the selection of arbitrary kernel functions in SVM was not probabilistic (Tipping, 2001). In addition, a high number of support vectors were used in an SVM because it determined the SVM accuracy, thus, the time computational complexity using SVM was high (Liyang, 2005). Furthermore, uncertain factors were not considered in the study.

Bao *et al*. (2007) sourced data for Brent and WTI crude oil as published by the EIAUSDE for a period from May 1987 to July 2007 and January 1991 to July 2007 respectively. Wavelet and Least-SVM (W-LSSVM) were utilized for the price projection of the crude oil price, for comparison and evaluation purposes, ARIMA, Least-SVM (LSSVM), SVM, Wavelet ARIMA (W-ARIMA) were also applied to project the crude oil price. Results indicated that the RMSE of ARIMA was 1.19, SVM was 1.17, W-ARIMA was 0.96, W-LSSVM was 0.98 and LSSVM was 1.21. However, the wavelet analysis may have

introduced bias into the projection results as stated earlier. The limitations observed in the study conducted by Xie *et al*. (2006) were similarly noted in the work by Bao *et al*. (2007). Qi and Zhang (2009) applied an SVM and cluster SVM for projecting crude oil price using 1000 records/cases. Both models were used to project crude oil price and it was concluded that the improved cluster SVM outperformed the conventional SVM but the numerical values of the accuracies for both the models were not reported in the study. The work by Bao *et al*. (2007) has the same limitations observed in (Qi and Zhang, 2009).

Khashman and Nwulu (2011a) designed an SVM model using weekly spot price of WTI crude oil sourced from the EIAUSDE covering the period from January 1986 to December 2009. The model was applied to project crude oil price and a rate of 81.3% accuracy was achieved. In another study, Khashman and Nwulu (2011b) conducted a comparative analysis of the SVM and back-propagation ANN for the projection of crude oil price. Both models were built based on WTI crude oil price data collected from the EIAUSDE covering the same period. The projection results indicate that SVM achieved 81.28%, whereas back-propagation ANNs 84.69%. However, the SVM and backpropagation ANN were applied as an IIS not as HIS. Furthermore, the limitations of SVM as discussed earlier, also applies here and the impact of uncertain factors was not considered in the models proposed by (Khashman and Nwulu, 2011a; Khashman and Nwulu, 2011b).

3.4.5 Fuzzy Systems

Zhang *et al*. (2010) introduced fuzzy time series into the projection of crude oil price. WTI spot price data were obtained from the EIAUSDE covering a period from January 1991 to December 2009. The results showed that the RMSE was 1.39 and concluded that the fuzzy

time series could be considered as a tool for short term projection. Nevertheless, uncertain factors were not considered in the modeling process.

3.4.6 Expert System

A few references were found in the literature (4 out of 59) that considered both demand/supply and unexpected events in building crude oil price projection models. These employed expert systems to handle the impact of unexpected events so that both regular market factors and uncertain events that are responsible for oil price volatility are considered during the modeling process (Wang *et al*., 2004; Yu *et al*., 2005; Shouyang *et al*., 2005; Mehdi, 2009). The listed references are discussed as follows:

Wang *et al*. (2004) proposed a hybrid model of ANNs, rule based expert system and web–based text mining called HIS. The HIS used historical data on monthly spot price of crude oil collected for WTI for a period from January 1970 to December 2002. Information on uncertain events affecting crude oil price were extracted through web–based text mining techniques. The proposed system operated by collecting information and comparing it to predefined patterns. If the information was based on irregular events affecting crude oil price, the rule-based expert system was executed to project crude oil spot price. Otherwise ANNs projection was executed using datasets of the crude oil. Simulation results showed that the HIS method has an RMSE of 1.916 and the ANN has an RMSE of 3.324. However, the proposed expert system was susceptible to the limitations which makes the system unsuitable for projection of crude oil price (refer to section 3.3 for details on the limitations of expert system in projections). Also, the study has not created a model that

allows retraining (refer to section 1.2 for justification on retraining). This makes the model unsuitable for modeling the impact of uncertainties that occur at different period.

Yu *et al*. (2005) proposed a knowledge-based expert system that hybridized text mining and rough sets (Rough-set-refined text mining). Text mining was responsible for searching the internet and the internal file system to collect both regular factors and uncertain events influencing crude oil price, creating a metadata repository and generating patterns and rules. Rough sets further refined the pattern and rules generated by the text mining to project crude oil price. ARIMA, Linear Regression Model (LRM), RW, and back-propagation ANNs were used to project crude oil price for the purpose of evaluating the effectiveness of the proposed technique. The data required for building the models were collected from the EIAUSDE covering the period from January 1970 to October 2004. The projection results show that the Backpropagation ANN had a hit ratio (correct percentage of projection) of 75.86%, ARIMA had 60.34%, LRM had 55.17%, RW had 51.72% and lastly the proposed Rough-set-refined text mining had 86.21%. However, the model proposed by Yu *et al*. (2005) was susceptible to the limitations which makes the system unsuitable for projection of crude oil price (refer to section 3.3 for details on the limitations of expert system in projections). In addition, retraining (refer to section 1.2 for justification on retraining) is lacking in the study.

Shouyang *et al*. (2005) proposed TEI@I nonlinear integration for projecting price of crude oil using both quantitative and qualitative data to build the projection model. This approach integrated several modules: a man–machine interface which provided an interaction point between the system and the user; web–based text mining to extract information on uncertain events influencing the fluctuation of crude oil price from the

internet and served as a rule-based expert system which received the information from the web–based text mining module to assesse its effect on crude oil price based on stored knowledge; the econometric module which applied ARIMA to model the linear component of crude oil price and backpropagation ANN which captured a non-linear pattern in the time series of crude oil price. The data used to build the ARIMA and backpropagation ANN model where monthly spot price of WTI covering the period from January 1970 to December 2003.The projection results produce by ARIMA and backpropagation ANN, the effects of irregular events as determined by web–based text mining and a rule-based expert system, were integrated to produce an integrated projection of crude oil price. The RMSE of ARIMA was 2.4868, while that of the ANN was 2.6436, Simple integration was 1.9665, and then propose TEI@I nonlinear integration was 0.5746. However, the expert system was susceptible to the limitations which makes the system unsuitable for projection of crude oil price (refer to section 3.3 for details on the limitations of expert system in projections). The proposed model does not allow retraining (refer to section 1.2 for justification on retraining) to capture the impact of uncertain factors in different period.

Mehdi (2009) proposed a FNN model and gathered crude oil price on a weekly basis from January 1989 to January 2009 collected from the Mediterranean Sidi Kerir Iranian light spot price FOB. The proposed projection model in the research was able to project Iranian light spot price with an accuracy of over ninety five percent (95.23%) directional statistics whereas the simple integration technique (summation of the results from each IIS) achieved over eighty nine percent (89.33%). Uncertain events influencing price of crude oil were translated into knowledge-based expert system elements from which were derived

the fuzzy rules to integrate expert judgment in the projection of crude oil price and thus complement the proposed model.

However, the expert system applied in the studies described above (Wang *et al*., 2004; Yu *et al*., 2005; Shouyang *et al*., 2005; Mehdi, 2009) are susceptible to the limitations which makes the system unsuitable for projection of crude oil price (refer to section 3.3 for details on the limitations of expert system in projections). In addition, the researches reviewed in this chapter have not created models that allow retraining (refer to section 1.2 for justification on retraining). In contrast, in view of the limitations of expert system and the capabilities of ANN, ANN might be the most appropriate technique for handling the impact of unexpected events on crude oil price. The ANN and GA are considered the most reliable and promising AIT (Woll *et al*., 1997), and ANN are referenced as the most powerful techniques ever established (Ma and Wu, 2010). Five techniques, namely, ANN, GAs, statistical inference, rule induction, and data visualization, were compared using the following criteria: optimization capability, computation complexity, flexibility, interpretability, scalability, ease of problem encoding, autonomy, and accessibility. The criteria were measured based on a five-point scale (very high, high, medium, low, and very low). ANN and GA were found to be the most suitable techniques for extracting knowledge from historical data (Zhang and Zou, 2004). An opportunity for ANN parameters optimization is provided through GA by utilizing their strengths and eliminating their limitations (Shapiro, 2002). Experimental evidence suggests that the optimization of ANN by GAs converges to an optimum solution (Huang *et al.*, 2009) in less computational time complexity than the backpropagation NN (Abbas *et al.*, 2003). Thus, optimizing ANN parameters using GAs is ideal because the limitations attributed to

ANN design will be eliminated and the combined approach will thereby become more effective than using by ANN alone.

3.4.7 Hybrid Intelligent Systems

In an earlier study conducted by Rast (2001), crude oil price were projected with FNN using crude oil time series extracted from NYMEX. Simulation results indicated that over ninety percent (90.4%) projection accuracy was achieved. However, the performance of the FNN suffers as the number of inputs and fuzzy rules increase, also, a systematic framework for determining optimal fuzzy rules is lacking (Malek *et al*., 2012). Impact of uncertain factors were not considered in the study, which can limit its practical applications in the real world.

Gholaman *et al*. (2005) designed a hybrid intelligent system composed of fuzzy rule-base and ANN. The data ranging from June 1998 to November 2000 and July 1988 to December 2000 were collected for experimental simulations from Sahand Naftiran (oil and gas Investment Company). The analysis of the experimental results showed that the projection of crude oil price was possible by using the hybrid intelligent system. However, the limitations that undermined the robustness of the study conducted by Rast (2001) were also noted in (Gholaman *et al*., 2005).

Fernandez (2006) adopted and applied the three models ARIMA-ANN, ARIMA-SVM (ARIMA-SVM) and SVM-ANN to project crude oil and natural gas spot price using data for the period of 1994 to 2005 sourced from DataStream. The MSE achieved was -8.80 for ARIMA-ANN, -11.82 for ARIMA-SVM, and 1.97 for SVM-ANN. However, the study

did not include uncertain factors and the limitations with respect to SVM and ANN discussed earlier undermine the performance of the SVM-ANN.

Lai *et al*. (2007) proposed WDNEVaR to project values at risk in crude oil market price. Data on WTI crude oil price used in this research were collected from Global Financial Data ranging from April 1983 to June 2006. The simulated result indicated that the MSE of WDNEVAR was 0.0069, for ARMA–GARCH, it was 0.0076, and for WDVAR, it was 0.0072. It was concluded that the proposed model was able to measure the risk value in crude oil market price. Conversely, the study suffered from shortcomings of wavelet transformation as earlier pointed out and uncertain factors were not incorporated.

Reza and Ahmadi (2007) applied the GA to optimize the architectural parameters of ANNs. Data on crude oil price were collected for a period covering the years 1983 to 2006 extracted from EIAUSDE to create a hybrid system for the projection of monthly crude oil price. Results suggested that the proposed model had an accuracy of 83%, whereas the technique chosen for comparison, such as, Short Term Energy Outlook (STEO) had an accuracy of 73%, GP had 75%, a hybrid of ANN, expert system, and web mining achieved 81% and lastly, the ANN had an accuracy of 57%. In the study, ANNs was trained to optimize weights using Levenberg-Marquardt backpropagation gradient descent algorithm to build a crude oil projection model. The study used the GA for selecting neurons in the hidden layer, the activation function, the number of layers and the connections between neurons in the ANN layers. Then, the Levenberg – Marquardt backpropagation (LMBP) algorithm was used to train the network and optimize the ANN weights and bias. However, the LMBP lacked convergence speed and suffered from the possibility of being trapped in local minima due to its iterative nature of finding the minimum error (Zweiri *et al*., 2003).

In addition, uncertain events that affect crude oil price were also not considered in the development of the crude oil price projection model.

Liu *et al*. (2007) proposed a FNN where, a RBFNN, a Markov chain-based semi-parametric model and a wavelet analysis projected results served as input to the FNN and the target output constituted the actual crude oil price. Using sample data of crude oil price for a period from May 1987 to August 2006 acquired from EIAUSDA, a projection model was built for the projection of crude oil price. The projection results based on the hybrid model was found to have an MSE of 1.0298, for RBFNN, it was 2.81591, for Markov Chain based Semi parametric model, it was 2.70602, for Wavelet analysis based projection model, it was 4.52168. The model proposed based on the FNNs was trained using backpropagation algorithm, but, the FNN was liable to limitations discussed earlier (Malek *et al*., 2012). Furthermore, uncertain factors were not considered in the proposed model.

Xu *et al*. (2007) proposed a rough sets and wavelet ANN (RSWNN) hybrid model for analyzing the factors that affect crude oil price and project future price. Their approach consisted of text mining, rough sets and wavelet ANNs. Text mining was used to retrieve the necessary data on the factors affecting crude oil price from the EIAUDE, Reuters and Brent crude oil. Rough set was used to further refine the documents and extract the main factors affecting crude oil price, while, wavelet ANNs was used to classify the factors based on the patterns that affect crude oil price. Data of these main factors were collected from 1970 to 2005 sourced from Brent crude oil. The simulated results showed the model projected crude oil price with an MSE of 2.94. However, the use of wavelet analysis could introduce bias in the estimation results as earlier discussed, and the model was trained with

backpropagation training algorithm which is susceptible to the limitations of the gradient descent algorithms as discussed earlier. Uncertain factors were also not considered in the model despite the evidence presented by Hamilton (2011) that they significantly affect crude oil price.

In the work of Fan *et al*. (2008) Pattern Modeling and Recognition System (PMRS), ENN and Generalize Pattern Matching based on Genetic Algorithms (GPMGA) models were used for creating a multi-step projection tool applicable to crude oil price. Crude oil data were sourced from the EIAUSDE. The projection results indicated that GPMGA achieved an RMSE of 1.09 while the PMRS and ENN obtained RMSEs of 1.57 and 1.09, respectively. However, only demand and supply factors were considered in the study without considering the impact of uncertainties in the modeling. Attribute selections were not discussed in the study. Since the ENN is a class of RNN, the network structure of the ENN is similar to that of the RNN. As the RNN becomes more complex, which further complicates the selection of optimal parameters; the computation of the error gradient in an RNN also becomes complicated because more attractors are present in the state space (Blanco *et al*., 2001).

Alaxandridis and Livanis (2008) performed a projection experiment using data collected from the EIAUSDA for WTI crude oil price for a period from January 1986 to October 2007 and applied the wavelet ANNs to project the crude oil price of the next one (1), three (3) and six (6) months. For comparison and evaluation purpose, the Hybrid of backpropagation ANN, expert system, and web mining (NESWM), backpropagation ANN, and STEO were applied to project crude oil price. The MSE recorded were 5.61, for NESWM, 9.25 for STEO, 3.71 for backpropagation ANN, and 3.39 for the wavelet ANN.

The wavelet and the ANN applied in this study were not immune to their limitations highlighted earlier in our discussion. In addition, the researchers have not considered the impact of uncertainties.

Zimberg (2008) began his research work by conducting a series of experiments with ENN, FFNN and the ANFIS based on WTI and Brent crude oil price covering the period from1991 to 2003. After preliminary experiments, the ANFIS was selected as the best performing model having the minimum MSE but unfortunately the numerical values of the preliminary results were not reported. The selected model and econometric model were used to project crude oil price in which simulated results showed that the RMSE of ANFIS was 0.8 in Brent crude oil price benchmark while the RMSE was 0.55 in WTI crude oil price benchmark, but the RMSE of the econometric model was not reported. The limitations noted in (Rast, 2001) were also noted in (Zimberg, 2008).

Yu *et al*. (2008a) gathered data for crude oil price (WTI and Brent spot price) of the EIAUSDE covering a period from January 2000 to March 2008. The generalized intelligent–agent–based fuzzy group model incorporated back-propagation ANNs, SVM regression and RBFNN. The IIS and HIS models were used to project crude oil price. The generalized intelligent–agent–based fuzzy group, SVM regression, BPNN, and ARIMA achieved MSE of 0.6458, 0.8778, 1.6443, and 1.2376 respectively. However, the limitations observed in Xie *et al*. (2006) were also noted in Yu *et al*. (2008a).

Yu *et al*. (2008b) adopted an empirical mode decomposition-based ANNs as proposed earlier in 1998 by Huang *et al*. and integrated it with ensemble learning to form the Empirical Mode Decomposition-Based ANN Ensemble Learning Pattern (EMD–FNN–

ALNN). Data for crude oil price was acquired from the EIAUSDA for a period of 1986 to 2006. The dataset was used to model EMD–FNN–ALNN, and its effectiveness was evaluated with data from Brent and WTI crude oil for one-step-ahead projection. ARIMA, Empirical Mode Decomposition–Feed Forward ANN–Adaptive Linear ANNs (EMD–ARIMA–ALNN), Empirical Mode Decomposition- Autoregressive Integrates Moving Average (EMD–ARIMA–Averaging), Single-Feed-Forward ANNs (Single FNN) and Single Autoregressive Integrates Moving Average (Single ARIMA) models were also used for the projection of crude oil price for the purpose of performance evaluation. The RMSE of EMD–FNN–ALNN was found to be 0.225, EMD–FNN–Averaging was 0.457, EMD–ARIMA–ALNN was 0.872, EMD–ARIMA–Averaging was 1.392, Single FNN was 0.743, and lastly the RMSE of Single ARIMA was 1.768. However, the researcher failed to incorporate the effects of uncertain factors in the modeling procedure and the ANNs was susceptible to the limitations (Zweiri *et al*., 2003) similar to the statistical methods (Yu *et al*., 2009; Zalloi, 2009) as discussed earlier.

He *et al*. (2009) proposed a Wave Decomposition Network Value at Risk (WDNEVaR) model to estimate crude oil market values. Data in the form of daily closing price of crude oil were collected over a period from April 1983 to Jun 2006, May 1987 to Jun 2006 and January 1997 to January 2005 sourced from EIAUSDE and Dubai respectively. ARMA–GARCH, Wavelet Decomposition Value at Risk (WDVaR) and WDNEVaR models were used to project oil price in all three markets. The findings demonstrated that the MSE of WDNEVaR model (hybrid of wavelet analysis and ANNs) was 0.0059, for WDVaR, it was 0.0060, and for ARMA – GARCH, it was 0.0057. However, the use of wavelet analysis could introduce bias in the estimation results as earlier discussed, and the model

was trained with backpropagation training algorithm which is susceptible to the limitations of the gradient descent algorithms as discussed earlier. Uncertain factors were also not considered in the model despite the evidence presented by Hamilton (2011) that they significantly affect crude oil price.

Ghaffari and Zare (2009) collected data of crude oil price from 2004 to 2007 and adopted an ANFIS to project the price. Simulation results indicated a projection accuracy of 68.18%. However, the performance of the ANFIS suffers as the number of inputs and fuzzy rules increase, also, a systematic framework for determining optimal fuzzy rules is lacking (Malek *et al*., 2012). Furthermore, the impact of uncertain factors was not considered in the study, which can limit its practical applications in the real world.

Qunli *et al*. (2009) proposed a hybrid of the wavelet transforms and RBFNN (WRBFNN) in their study. Data were collected from UK Brent crude oil spot price FOB covering a period from January 1997 to October 2008. The wavelet transform was applied to improve the projection accuracy and WRBFNN was used to project crude oil price. Unfortunately, the specific numerical performance result of the model was not reported, but it was claimed in the study that the WRBFNN achieved improved performance. The impact of uncertain events on crude oil price was not considered in the model and this study is also subjected to the limitations discussed in the work of Xu *et al*. (2007).

Ma (2009) collected data covering New York Harbor residual oil price from November to April 2007. The data were used to develop two ANNs models including symbol evolutionary immune clustering ANNs and RBFNN to project crude oil price. An evolutionary immune clustering algorithm was utilized to optimize the centers of an

RBFNN to build a crude oil projection model to avoid trial and error, and to avoid searching within a limited space (a space that could not lead to the optimal model). However, the width, hidden-layer neurons, and connection weights were also significant in the RBFNN design and must be optimized with the immune clustering algorithms because they lack the ideal framework for choosing the best values. The comparative performance showed that the symbol evolutionary immune clustering ANNs performed better than the RBFNN. Unfortunately, the values of the performance indicators were not reported and the impact of uncertain factors were also not considered in the research.

Abdullahi and Zeng (2010) proposed an Artificial ANN–Quantitative (ANN-Q) model. Quantitative data were derived from online news and monthly crude oil price were obtained from EIAUSDE for a period from January 1984 to February 2009. Two other models were also employed to evaluate the performance of the proposed model, namely Mining + Econometrics + Intelligence (intelligent algorithms) + Integration (TEI@I) nonlinear integration model (TEI@I nonlinear integration) and the Empirical Mode Decomposition FFNN Adaptive Linear ANNs (EMD–FNN–ALNN) model. All three models were used to project crude oil price and their performances were compared. The results from the performance analysis showed that the RMSE of EMD–FNN–ALNN was 0.2730, for TEI@I) nonlinear integration was 1.0549 and for the ANN - Q was 2.2690. However, the model proposed in the study did not improve the accuracy of crude oil price projection. The shortcomings noted in the research conducted by Haidar *et al.* (2008) were also presented in this research, in addition, uncertain factors were not incorporated.

Phichhang and Wang (2010) proposed a hybrid model of ANN and GARCH to project the crude oil price volatility in Chinese and US crude oil markets. US spot price FOB and

Chinese Daqin spot price FOB were collected from the EIAUSDE, the sample data covered the period from 1997 to 2010. The comparative analysis of the results indicates that the MSE of ANN-GARCH in US and China markets are 0.46 and 0.53 respectively, whereas for the US and China markets the MSE for GARCH are 0.75 and 0.76 respectively. However, the training of the ANN suffers from the same limitations pointed out by (Shambora and Rossiter, 2007).

Panella *et al*. (2011) considered two different models for the projection of the dynamics in crude oil, natural gas and electricity price in European and US markets. Sample data were collected from Europe (Brent crude oil) and the EIAUSDE (WTI) covering a period from 2001 to 2010. The models were RBFNN, and Adaptive network-based fuzzy inference system (ANFIS). Experimental results suggested NMSE for ANFIS and RBFNN were 0.057 and 0.814 respectively. However, the performance of the ANFIS suffers as the number of inputs and fuzzy rules increase, also, a systematic framework for determining optimal fuzzy rules is lacking (Malek *et al*., 2012). Impact of uncertain factors were not considered in the study, which can limit its practical applications in the real world.

Mingming and Jinliang (2012) built a multiple-wavelet RNNs model incorporating wavelet and RNN. The wavelet was used to capture patterns in the crude oil price dataset, and RNN was used to project crude oil price at each scale. The data covered the period from 1946 to 2010 acquired from EIAUSDA. The data about Brent and WTI crude oil, representing European and US oil markets were used. The study showed that the multi-wavelet RNNs model achieve 4.06 Mean Percentage Errors. It was concluded that the model possessed the ability to project the subsequent year's world crude oil price, but lacks the ability to project long term crude oil price and uncertain factors were not included in the study. The

wavelet analysis suffered from the limitations pointed out earlier (Abramovich *et al*., 2000) and the RNNs robustness were undermined by the limitations reported by (Blanco *et al*., 2001).

Crude oil price data were obtained from the EIAUSDA for a period from 1985 to 2007 by Azadeh *et al*. (2012) who proposed ANNs–fuzzy regression algorithms to estimate long term crude oil spot price (price of a crude oil price at a particular time). Fuzzy regression (FR) and five different ANNs models including Bayesian Regulation (BR), Gradient Descent BEP (GDX), Levenberg–Marquardt (LM), Batch weight/bias learning rules (B), BFGS Quasi-Newton (BFG) were used to project the price. The Mean Absolute Percentage Error (MAPE) record in the experiments was 0.192000 for ANN-LM while for ANN-BFG, it was found to be 0.213250, for ANN-B, it was 0.224375, for ANN-BR, it was 0.267500, for ANN-GDX, it was 0.280125 and for FR, it was 0.402875. The impact of uncertainties on crude oil price was not incorporated in the model proposed by Azadeh *et al*. (2012) and the Neuro-fuzzy regression was associated with limitations that reduce its effectiveness as explained in the works by (Malek *et al*., 2012).

Jammazi and Aloui (2012) built a hybrid intelligent model called Harr a Trous Wavelet multilayer back- propagation ANNs (HTW-MBPNN). Monthly crude oil price were collected from January 1988 to March 2010, sourced from the EIAUSDE. The experiment was aimed at projecting of oil price for the next 19 months (June 2011 to December 2012). The MSE result of HTW–MBPNN was found to be 3.59066 and for back-propagation ANNs the MSE was 6.192867. However, the shortcomings pointed out in the work of Xu *et al*. (2007) also applies in the work of Jammazi and Aloui (2012).

Wang *et al*. (2012) embedded a jump stochastic time effective function into ANNs in order to improve its projection accuracy. Data used in the study were extracted from five different sources, namely Shanghai Composite Index (SHCI), Shenzhen Compositional Index (SZCI), Shenzhen Petrochemical Index (SZPI), Daqing Oil Field (Daqling), Shengli Oil Field (Shengli), and the Stock Price of China's Largest Oil Company: China Petroleum & Chemical Corporation (SINOPEC). The proposed model was applied to project fluctuation of crude oil price. The Mean Absolute Error (MAE) achieved by the proposed model for SHCI, SZCI, SZPI, Daqling, Shengli, and SINOPEC were 116.5679, 433.4098, 37.8812, 2.8028, 3.6815, and 0.4957 respectively. The result analysis showed that the smaller the fluctuation in price was, the more accurate the projection and vice versa. The limitations of this study are similar to those found in the study conducted by Shambora and Rossiter (2007). Again, uncertain factors were not incorporated in the research as well.

3.5 Hybrid Intelligent System Vs. Individual Intelligent System, Attribute Selection, and Model Evaluation in Crude Oil Price Projection

In Table 3.1, a comparison of the characteristic of HIS and that of the IIS are listed. As can be observed in Table 3.1, the HIS outperforms IIS in all of the studies. Thus, HISs demonstrates superiority over IIS in terms of crude oil price projection. The comparison methods are summarized in Table 3.2, strengths and weaknesses of the studies are presented in Tables 3.1 and 3.3 respectively, while the proposed methods are in section 3.4.

Table 3.1: Summary of published papers by strengths

References	Strength
Wang *et al*. (2004), Shouyang *et al*. (2005), Yu *et al*. (2005), Malik and Nasereddin (2006), Fernandez (2006), Bao *et al*. (2007), Liu *et al*. (2007), Fan *et al*. (2008), Abdel-Aal (2008), Zimberg (2008), Yu *et al*. (2008a), Yu *et al*. (2008b), Yu *et al*. (2008c), Ma (2009), Panella *et al*. (2011), Mehrara *et al*. (2011), Jammazi and Aloui (2012).	The authors demonstrated that HISs performs better than IISs.
Kaboudan (2001), Wang *et al*. (2004), Aladwani and Iledare (2005), Yu *et al*. (2005), Shouyang *et al*. (2005), Malik and Nasereddin (2006), Moshiri and Foroutan (2006), Xie *et al*. (2006), Liu *et al*. (2007), Bao *et al*. (2007), Khazem (2007), Shambora and Rossiter (2007), Lai *et al* (2007), Fan *et al*. (2008), Quek *et al*. (2008), Malliaris and Malliaris (2008), Yu *et al*. (2008a), Yu *et al*. (2008c), Kulkar and Haidar (2009), He *et al*. (2009), Qi and Zhang (2009), Ma (2009), Qunli *et al*. (2009), Wang and Yang (2010), Torban (2010), Phichhang and Wang (2011), Pang *et al*. (2011), Panella *et al*. (2011), Chen and Qu (2011), Mehrara *et al*. (2011), Haidar and Wolff (2011), Khashman and Nwulu (2011a), Jammazi and Aloui (2012).	The authors established that the techniques proposed (refer to section 3.4) in their study, perform better than the methods (refer to Table 3.2) chosen for comparison purposes.
Gholamian *et al*.(2005), Yu *et al*. (2005), Rast (1997), Reza and Ahmadi (2007), Xu *et al*. (2007), Alexandridis and Livanis (2008), Mingming *et al*. (2009), Jinliang *et al*. (2009), Mehdi (2009), Ghaffari and Zare (2009), Mingming and Jinliang (2012), Wang *et al*. (2012).	A HIS is used for projection of crude oil price. It was found that HIS is better than the IIS (refer to section 3.4.7)
Kaboudan (2001), Shouyang *et al*. (2005), Raudys (2005), Yu *et al*. (2005), Khazem (2007), Xu *et al*. (2007), Reza and Ahmadi (2007), Quek *et al*. (2008), Malliaris and Malliaris (2008), Alexandridis and Livanis (2008), Alizadeh and Mafinezhad (2010), Abdullah and Zeng (2010), Pang *et al*. (2011).	Performed attributes selection

Each study establishes that all of the techniques proposed by the researchers perform better than the methods (see Table 3.2) that were chosen for comparison purposes. There is only one exception in a study conducted by Abdullah and Zeng (2010), which found that the performance of the proposed model was inferior to the comparative methods.

Table 3.2: Overview of published papers by methods

Reference	Method/s compared with for evaluation purposes
Wang and Yang (2010)	Econometrics model
Shouyang *et al*. (2005), Xie *et al*. (2006)	ANN and ARIMA
Kulkar and Haidar (2009)	SVM
Kaboudan (2001), Quek *et al*. (2008), Malik and Nasereddin (2006), Mehrara *et al*. (2011), Khashman and Nwulu (2011b)	ANN
Ma (2009)	RBFNN
Bao *et al*. (2007)	Wavelet transform and L-SVMleast-squares support vector machine
Fan *et al*. (2008)	Pattern modeling and recognition system and RNN
Moshiri and Foroutan (2006), Haidar and Wolff (2011)	ARIMA and GARCH
Torban (2010)	ARIMA and Structure Vector Error
Malliaris and Malliaris (2008)	MLR and simple model
Yu *et al*. (2008a)	RBFNN, SVM, and ANN
He *et al*. (2009)	Wavelet decomposition value at risk, ARIMA, and GARCH
Jammazi and Aloui (2012)	Back-propagation ANN, STEO, and WTI future projection price
Kulkar and Haidar (2009)	RNN
Chen and Qu (2011)	Polybasic linear regression
Qunli *et al*. (2009)	Wavelet transform
Pang *et al.* (2011)	Linear relative inventory and nonlinear relative inventory
Shambora and Rossiter (2007)	Buy and hold, technical analysis, and naïve RW
Phichhang and Wang (2011)	GARCH
Yu *et al*. (2005)	Linear regression model, ARIMA, and back-propagation ANN
Liu *et al*. (2007)	RBFNN, Markov chain, and wavelet analysis
Khazem (2007), Aladwani and Iledare (2005)	Regression analysis
Abdullah and Zeng (2010)	Hybrid of text mining, an econometrics model, and a back-propagation ANN
Panella *et al*. (2011)	RBFNN
Lai *et al*. (2007)	Autoregressive Moving Average and GARCH

Table 3.2 continou

Azadeh *et al*. (2012)	BR, GDX, LM, B, and BFG
Fernandez (2006)	ARIMA-ANN and ARIMA-SVM
Alaxandridis and Livanis (2008)	NESWM, STEO, and backpropagation ANN

Some studies applied HISs to oil price projection; this approach was more effective, but do not compare the effectiveness of the HIS to other methods for evaluation, as indicated in Table 3.3. Tables 3.1 and 3.3, show that very few of the studies (12 out of 59) in the literature actually performed attribute selection. Attribute selection increases the projection accuracy and minimizes the computational cost and complexity, as argued in (Peter *et al*., 2001). The researchers that compared their proposal with other chosen techniques did not proceed further to check for statistical significant differences between the projection accuracies of the compared techniques and between the projected and observed values. Furthermore, other studies (refer to Table 3.3 for the studies) did not compare their results with other techniques for benchmarking purposes and performance evaluation. In previous years, the attention of the machine learning and data mining research community has been drawn to the need for validating their results statistically (Demˇsar, 2006). This trend can be attributed to the growing interest in the research area, the development of real-life applications of these research approaches, and the performance comparisons made between existing, modified, and newly developed algorithms (Demˇsar, 2006). Only Khazem (2007) investigated the relationship among independent attributes. All of the other studies did not investigate the extent of the relationship among the independent attributes, as stated in Hair *et al*. (2010) that successful projection requires a set of independent attributes to form a positive correlation coefficient relationship. To investigate the relationship between attributes, a correlation coefficient analysis is required. The correlation coefficient analysis is used to describe the strength and direction of the relationship among attributes (Pallant,

2010). The correlation coefficient, which is a value that indicates the degree of a relationship, is measured using Equation 3.1 as follows:

$$r = \frac{N\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[N\sum x^2 - \left(\sum x\right)^2\right]\left[N\sum y^2 - \left(\sum y\right)^2\right]}}, \tag{3.1}$$

where

N is the number of attributes

$\sum x^2$ is the sum of the product of attributes

$\sum x$ and $\sum y$ are sums of attributes

$\sum x^2$ is the sum of the squared x attribute

$\sum y^2$ is the sum of the squared y attribute

r = correlation coefficient

The output is the size of the value of correlation coefficient, which ranges from −1.00 to 1.00. This value indicates the strength of the relationship between the attributes. A correlation coefficient of 0 indicates no relationship, a correlation of 1.0 indicates a perfect positive correlation, and a value of -1.0 indicates a perfect negative correlation. Different authors provide various interpretations, however, Cohen (1977) suggested the following guidelines: Small correlation coefficient = 0.10 to 0.29, Medium correlation coefficient = 0. 30 to 0.49, Large correlation coefficient = 0. 50 to1.0. The positive correlation coefficient relationship indicates that increases or decreases of the inputs attributes also apply to the output attribute (e.g. crude oil price).

Table 3.3: Summary of published papers by weaknesses

References	Weakness
Rast (2001), Pan *et al*. (2009), Raudys (2005), Gholamian *et al*.(2005), Reza and Ahmadi (2007), Xu *et al*. (2007), Haidar *et al*. (2008), Abdel-Aal (2008), Alexandridis and Livanis (2008), Mehdi (2009), Ghaffari and Zare (2009), Mingming *et al*. (2009), Lacks *et al*. (2009), Jinliang *et al*. (2009), Alizadeh and Mafinezhad (2010), Zhang *et al*. (2010), Panella *et al*. (2011), Khashman and Nwulu (2011a), Movagharnejad *et al*. (2011), Mingming and Jinliang (2012), Wang *et al*. (2012), Sotoudeh and Farshad (2012).	The effectiveness of the proposed method (refer to section 3.4) was not compared to the performance of another method, for evaluation.
Kaboudan (2001), Aladwani and Iledare (2005), Raudys (2005), Xie *et al*. (2006), Moshiri and Foroutan (2006), Shambora and Rossiter (2007), Khazem (2007), Haidar *et al*. (2008), Malliaris and Malliaris (2008), Quek *et al*. (2008), Kulkar and Haidar (2009), Qi and Zhang (2009), Pan *et al*. (2009), Ma (2009), Lacks *et al*. (2009), Torban (2010), Movagharnejad *et al*. (2011), Alizadeh and Mafinezhad (2010), Chen and Qu (2011), Qunli *et al*. (2009), Zhang *et al*. (2010), Wang and Yang (2010), Haidar and Wolff (2011), Khashman and Nwulu (2011a), Khashman and Nwulu (2011b), Sotoudeh and Farshad (2012).	IISs are used to built a projection model, whereas HISs are more effective.
Rast (2001), Wang *et al*. (2004), Aladwani and Iledare (2005), Shouyang *et al*. (2005), Gholamian *et al*. (2005), Malik and Nasereddin (2006), Fernandez (2006), Moshiri and Foroutan (2006), Xie *et al*. (2006), Shambora and Rossiter (2007), Liu *et al*. (2007), Bao *et al*. (2007), Lai *et al*. (2007), Fan *et al*. (2008), Zimberg (2008), Abdel-Aal (2008), Haidar *et al*. (2008), Yu *et al*. (2008a), Yu *et al*. (2008b), Yu *et al*. (2008c), Kulkar and Haidar (2009), He *et al*. (2009), Qi and Zhang (2009), Qunli *et al*. (2009), Jinliang *et al*. (2009), Ghaffari and Zare (2009), Mingming *et al*. (2009), Pan *et al*. (2009), Ma (2009), Lacks *et al*. (2009), Mehdi (2009), Torban (2010), Wang and Yang (2010), Zhang *et al*. (2010), Movagharnejad *et al*. (2011), Chen and Qu (2011), Phichhang and Wang (2011), Panella *et al*. (2011), Mehrara *et al*. (2011), Haidar and Wolff (2011), Khashman and Nwulu (2011a), Khashman and Nwulu (2011b), Sotoudeh and Farshad (2012), Jammazi and Aloui (2012), Mingming and Jinliang (2012), Wang *et al*. (2012).	Attribute selection is not present in these studies.

Comprehensive and substantial effort are typically made for data cleansing, and pre-processing which deserves more attention in the domain of crude oil price projection. It is estimated that approximately 80% of the data mining process is devoted to data cleansing and pre-processing. Quality data mining results are obtained from quality data; similarly,

poor-quality data typically yield poor results regardless of the intelligence of the algorithms being applied to the task (e.g., projection) (Zhang *et al*., 2003). The model must be simplified because it is not feasible to consider all of the attributes that are involved in the actual problem. Only the attributes that have a major impact on the projection should be considered for the task (Zimberg, 2008). Quek *et al*., (2008) argued that the inclusion of irrelevant inputs decreases the projection accuracy and the generalization ability, and increases the computational complexity. Dehuri and Cho (2010) pointed that the quality results of data mining algorithms are obtained from quality data; likewise, poor quality data typically yield poor results regardless of how intelligent the algorithms are (e.g. projection). Despite the significance of attribute selection, few studies (12 out of 59) have been found in the domain of crude oil price projection. For example, (Alexandridis and Livanis, 2008; Pang *et al*., 2011) used the correlation coefficient analysis to select the optimum input attributes. Similarly, multiple linear regression has also been applied to select significant input attributes and used to built an ANNs model for the projection of crude oil price (Khazem, 2007; Malliaris and S.G. Malliaris, 2008). In a separate study, Amin-Naseri and Gharacheh (2007) applied partial autocorrelation function as well as PCA for selection of input features for building an ANNs model for the projection of crude oil price. However, these statistical methods assume a normal distribution of input attribute data (Su and Wu, 2000) which makes them unsuitable for nonlinear and volatile crude oil price data. In addition, PCA is an unsupervised learning technique which does not correlate target inputs and outputs (Grimm and Yarnold, 2002). On the other hand, GAs are statistically better than these statistical methods in terms of the accuracy of input attribute selection (Oreski *et al*., 2012). GA is able to handle a very large search space to obtain

optimal solutions. Genetically optimized ANNs can effectively avoid the statistical methods' assumption of normal distribution (Su and Wu, 2000). For model simplification and improved efficiency (Alizadeh and Mafinezhad, 2010; Kaboudan, 2001) a trial and error technique, which is, time consuming, and laborious was used to select the relevant attributes as inputs to the ANNs model for the projection of crude oil price (Alizadeh and Mafinezhad, 2010; Kaboudan, 2001)

3.6 Frequency of Data Collection

The frequency of the data collection for the studies reviewed in this research is summarized in Table 3.4. As indicated in Table 3.4, data were collected on several different frequencies depending on the research objective. A large number of researchers collected data on a daily and monthly frequency during the modeling process. Very few extracted their data on a weekly, quarterly, or annual basis, while others do not disclose their data collection frequency. Weekends and unexpected events cause the oil market to be halted, which creates inconsistencies and missing points in the daily data. In its place, weekly or monthly data should be used to avoid the missing points (Shouyang *et al*., 2005; Xie *et al*., 2006). The use of monthly data restricts the projection horizon to monthly intervals and restricts the amount of training and testing data significantly (Pan *et al*., 2009). Additionally, quarterly and annual data avoid missing points in the data, reduce training and testing data more significantly, and restrict the projection horizon to quarterly and annually, respectively.

Table 3. 4 Overview of published papers by the frequency of the data collection

Reference	Frequency
Rast (2001), Raudys (2005), Gholamian *et al*. (2005), Moshiri and Foroutan (2006), Fernandez (2006), Khazem (2007), Shambora and Rossiter (2007), Lai *et al*. (2007), Yu *et al*. (2008b), Fan *et al*. (2008), Malliaris and Malliaris (2008), Quek *et al*. (2008), Haidar *et al*. (2008), Zimberg (2008), Yu *et al*.(2008a), Yu *et al*. (2008c), Lacks *et al*. (2009), Ma (2009), Mingming *et al*. (2009), Qunli *et al*. (2009), Kulkar and Haidar (2009), He *et al*. (2009), Qi and Zhang (2009), Ghaffari and Zare (2009), Pan *et al*. (2009), Wang and Yang (2010), Panella *et al*. (2011), Haidar and Wolff (2011), Mehrara *et al*. (2011), Mingming and Jinliang (2012), Wang *et al*. (2012).	Daily
Kulkar and Haidar (2009), Phichhang and Wang (2011), Khashman and Nwulu (2011a), Khashman and Nwulu (2011b)	Weekly
Kaboudan (2001), Wang *et al*. (2004), Aladwani and Iledare (2005), Yu *et al*. (2005), Shouyang *et al*. (2005), Xie *et al*. (2006), Bao *et al*. (2007), Reza and Ahmadi (2007), Abdel-Aal RE (2008), Alexandridis and Livanis (2008), Zhang *et al*. (2010), Alizadeh and Mafinezhad (2010), Abdullah and Zeng (2010), Movagharnejad *et al*. (2011), Pang *et al*. (2011), Jammazi and Aloui (2012).	Monthly
Malik and Nasereddin (2006), Torban (2010)	Quarterly
Xu *et al*. (2007), Azadeh *et al*. (2012), Sotoudeh and Farshad (2012)	Annually
Liu *et al*. (2007), Jinliang *et al*. (2009), Chen and Qu (2011)	Not disclosed

Such attributes as oil supply, demand, inventory (Haidar *et al*., 2008; Kulkar and Haidar, 2009), and GDP are not available on a daily frequency, which further complicates the projection of crude oil price (Kulkar and Haidar, 2009). Inventory data of OECDs are available only on a monthly interval (Pang *et al*., 2011). To reconcile the contradictions on the issue of the frequency of the data collection, the suggested criteria for selecting relevant independent attributes for crude oil price projection can be considered when collecting data; the criteria includes: data availability (Khazem, 2007; Yu *et al*., 2005), positive correlation coefficient between attributes (Yu *et al*., 2005), retrieval of the data on a timely basis, consistent availability of the data, and justification of the attribute influence on crude oil price (Khazem, 2007).

3.7 Data Normalization

In the real world, there is a strong chance that large databases may be inconsistent, have outliers, missing values, typing mistakes, differences in measurement units, etc. These anomalies trigger the need for data cleansing and pre – processing which constitute an important aspect of any intelligent model construction process, so that the anomalies found in the data can be cleaned up (Witten *et al*., 2011). Several branches of computer science, including pattern recognition, machine learning, data mining, web intelligence and information retrieval, data need to be standardized by preprocessing the original (raw) data. Such a practice is expected to yield quality data (Zhang *et al*., 2003). Data transformation (normalization) is part of the data standardization process, which prevents the saturation of neurons (Kaynar *et al*., 2011), improves the projection accuracy (Kafaee and Saramad, 2009; Abdel-Aal, 2008; Xu *et al*., 2007), improves the computation time and minimizes the error (Quek *et al*., 2008; Movagharnejad *et al*., 2001) and avoids numerical overflows, while weight interpretations are preserved (Tan *et al*., 2012). However, some researchers have questioned the efficacy of data normalization, arguing that the use of raw data is preferred to prevent the destruction of the original patterns in the historical data (Jammazi and Aloui, 2012), which is typically used due to its simplicity (Xie *et al*., 2006). It was argued in (Peter *et al*., 2001), that the use of the linear activation function in the output layer of an ANN invalidates data normalization and, thus, renders the exercise meaningless. ANN automatically adjust its weights adaptively; thus, data normalization is not necessary (Zhang *et al*., 1998).

3.8 The Trend of the Researches in the Domain of Crude Oil Price Projection

Figure 3.2 illustrates that the largest number of published papers in crude oil price projection using AIT was recorded in 2008, with 11 publications, whereas the smallest number was recorded in 2004, with only one publication. This trend typically exhibits an increasing number of publications; of the 59 available for this review, 45 of them were published between 2008 to 2012.

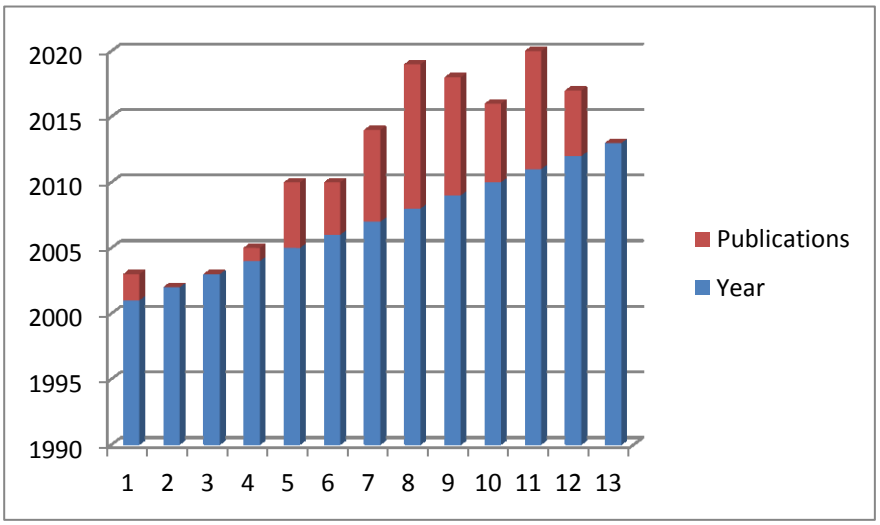


Figure 3.2: Publications on the topic of crude oil price projection

Publications in 2002, 2003, and the second quarter of 2013 are scarce. According to He *et al*. (2009), crude oil price projection did not receive adequate attention when compared to the price of other assets such as stock price despite its global importance.

3.9 Justification for Alternative Approach to Crude Oil Price Projection

An uncertainty is expected to occur (refer to section 2.2 for details on uncertainty), evaluating the impact of uncertainty is extremely difficult as earlier stated, and even if a

similar uncertain event occurs at a different time, the impact on crude oil price could differ. Thus, some new methods should be proposed to handle the impact of these uncertainties in different periods of time (Zhang *et al*., 2008). This statement and the limitations attributed to various oil price projection methods such as statistical, ANN, etc. motivates the researcher to propose an alternative model framework based on the Neuro-Genetic model that can learn new patterns from new datasets which may have been distorted by uncertainties (see Table 2.1). Completely replacing the initial dataset with a new dataset reduces the training data which retards projection accuracy (Behzadian *et al*., 2009) and large sample of data is required for building a robust model (Zhang and Zhou, 2004). The size of the new subset of dataset added into the initial data set has to be equal to the size of old subset that is removed. Therefore, the Neuro-Genetic model is to be periodically retrained with volatile data, such as data from the uncertainties listed in Table 2.1. Retraining would allow the Neuro-Genetic model to learn and capture new patterns based on the volatile data to predict crude oil price while considering the impact of the uncertainties on the crude oil price at different periods of time.

3.10 Summary

This chapter presents a review of the research conducted on the application of AIT in solving problems of crude oil price projection. Experimental data were mostly extracted from the EIAUSDE, and the most patronized oil markets are the WTI and Brent. Daily and monthly forecasting models received much attention from researchers. The HIS demonstrates superiority over IISs, as all of the studies that compared the two methods provided empirical evidence that the HISs are preferable in the projection of crude oil price because the HIS improves the strength of the IIS and eliminate the weaknesses of the IISs.

In the review of the various methods undertaken by several researchers (refer to section 3.4) for the projection of crude oil price, almost all the studies argued that their proposed method is better than the method/s chosen to evaluate the effectiveness of their proposed method. The integration of wavelet analysis and AIT is attracting unprecedented interest in crude oil price projection. In the area of crude oil price projection, statistical evaluation of results is limited. Thus a real life application of a projection model to be developed and their projection results need to be assessed. Despite their limitations, ANN parameters optimized with GA should further be researched to explore the full potential of the techniques in forecasting crude oil price because of their superiority over other AIT in solving complex, and dynamic problems. It is evident that this particular area of research is rapidly developing because 44 out of 59 research papers have been published within the last five years. In view of the limitations of expert system and the capabilities of ANN, a Neuro-Genetic model might be the most appropriate technique for handling the impact of unexpected events on crude oil price. The researchers in the domain of crude oil price projection depend heavily on back-propagation ANN for the projection of crude oil price. Some researchers have applied statistical methods, manual, and trial and error techniques for the selection of input attributes. Only one published document (Khazem, 2007) has been founded on the investigation of the correlation coefficient relationship among independent attributes and dependent attributes. The researchers that compared their proposed method with other chosen techniques did not proceed further and check for statistical significant differences between the projection accuracies of the compared techniques and between the projected and observed values. Some researchers did not compare their results with other techniques for performance evaluation and benchmarking

purposes. In view of the arguments in the literature about data normalization, an initial experiment in the domain of application has to be performed in order to ascertain the data standardization method to be used. Therefore, the performance of the model using normalized versus raw data needs to be compared in order to choose the best option.

CHAPTER FOUR

THEORETICAL FRAMEWORK OF THE RESEARCH

4.1 Introduction

Artificial Intelligence (AI) is the effort to introduce intelligence in machines. This effort started from the early days of the computer age (the specific date was not mentioned). Alan Turin, John Von Neumann, Norbert Wiener among others, were pioneers in the field of computer science fortified with a dream of building intelligence (self- reproduction, ability to learn and have regulation of their settings) into computer programs. Apart from electronics, early computer scientists were also involved in biology and psychology, in which, the natural system was their guide towards incorporating intelligence into computer programs. This is why applications of computers were not only restricted to missile trajectory computation and deciphering military code, but also extended to the representation of the biological brain, imitating human learning paradigms and mimicking biological evolution. The biological, computational algorithms have dwindled away over a period of years, but it was resurrected in the 1980's with full force by the computing research community and this led to the development of ANN, machine learning, and evolutionary computation like GA (Mitchell, 1999). This chapter briefly describes the basic theory and operations of ANN and GA.

4.2 Artificial Neural Networks

The original aim of the creation of the ANN was to mathematically represent the processing of information in biological systems (McCulloch and Pitts 1943). The ANN is a system that processes information similar to the human brain and constitutes a general

mathematical representation of human reasoning. These networks are built on the following assumptions (Fausett, 1994):

i. Information is processed by neurons
ii. Signals are communicated between neurons through established links
iii. Every connection between neurons is associated with weight; transmitted signals between neurons are multiplied by the weight
iv. Every network neuron applies an activation function to its input signals so as to regulate its output signal

The ANN structure comprises neurons that are distributed in the input, hidden, and output layers as depicted in Figure 4.1. Neurons in the input layer supply inputs to neurons in the hidden layers and, subsequently, to neurons in the output layer (Haykin, 2005). Input neurons in the input layer correspond to the independent variables in the problem definition (Abdi *et al*., 2012), and the dependent variable corresponds to the output node (Peter, 2001). There can be more than one hidden layer; however, theoretical works, such as (Zhang *et al*., 1998), argued that one hidden layer is sufficient to approximate any complex non-linear function.

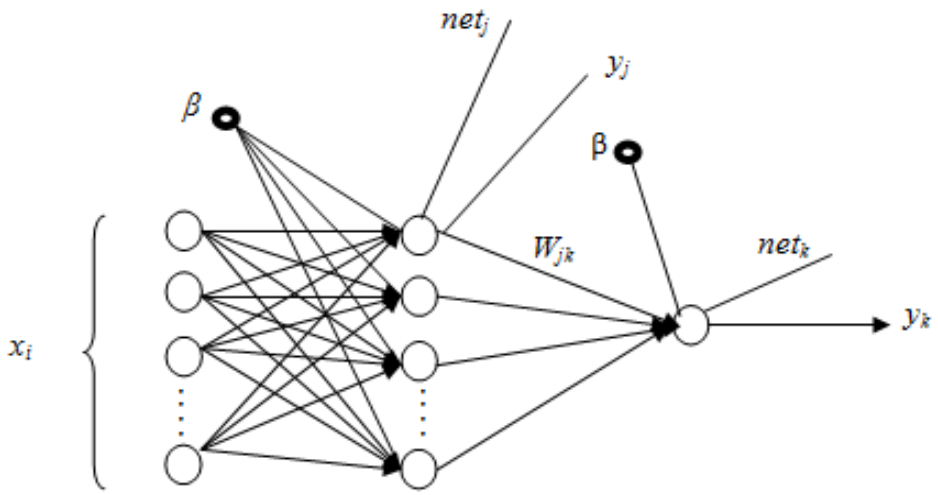


Figure 4.1: Typical ANN structure

The ANN is an algorithm for processing information in parallel and can model complex and nonlinear associations using input – output training from experimental datasets that are extracted from the application domain. The intrinsic capabilities of the ANN enable the algorithm to provide a nonlinear mapping of input and output vectors. The robustness and reliability of the ANN model depend on the suitability of the data preprocessing technique that is used, the design of the correct architectural configurations, and the choice of appropriate network training parameters (Hornick *et al*., 1989). A systematic framework for determining the appropriate number of hidden layer neurons does not exist; the best number of hidden layer neurons depends on the problem that is being modeled (Cheng and Li, 2008). Furthermore, an ideal framework for selecting the activation function, learning algorithms and other initial parameters is difficult to find in the literature (Azar, 2012). These optimum parameter values are usually obtained through the commonly used trial-and-error method (Pan and Wang, 2004). The structure of the ANN receives information from the input neurons, which flows to the output neurons in a forward direction. The neurons in the network operate by making a computation from their weighted inputs, and

the results are propagated through a nonlinear transfer function to perform a nonlinear mapping. The network is trained to optimize its weights and biases until the values predicted by the network at the output layer neurons are as close as possible to the target outputs. The error between the predicted and target values is measured using the MSE, the NMSE, and the MAE, among others, depending on the intended problem to be solved (Sedki *et al*., 2009). However, the MSE is chosen as an error measure because it is more appropriate than other performance metrics such as MAE, NMSE, etc. for measuring the performance of several algorithms on the same datasets as argued in (Peter *et al*., 2001). Iterative non–linear optimization algorithms are responsible for minimizing the error in most common ANN architectures. The accuracy of the global error function depends on the ANN weights. The best performing learning algorithms converge to the minimum global error function with a minimal number of iterations in less computation time (Pacifici *et al*., 2008). According to Sedki *et al*. (2009) the ANN detects patterns in the experimental data in such a way that it can make predictions with reasonable accuracy on the test dataset. However, precaution must be taken to avoid over-training the ANN, which could overfit the training data and degrade the prediction accuracy (Beale *et al*., 2013). The ANNs hidden layer can use the sigmoid activation function, and if another non-linear activation function were used, the output would be restricted to only a limited range of values, thus, cannot take large values (Beale et al., 2013) which can be disastrous to the raw data that was measured in millions (as would be in the case of attributes affecting crude oil price fluctuation such as, WCOP, and OPECCOP). According to the suggestion given in (Beale *et al*., 2013) the dataset can be partitioned into 70%:15%:15% for training, validation and testing in modeling ANNs.

4.2.1 Mathematical Model of the Artificial Neural Network

From Figure 4.1, input to the *jth* hidden layer node can be obtained based on Eqn. (4.1)

$$net_j = \sum_{i=1}^{N_i} w_{ji} x_i + \beta_j \quad 1 \le j \le N_q \tag{4.1}$$

Where net_j, w_{ji}, β_j are weighted sum of the inputs $x_1, x_2, x_3 ... x_q$, connection weight between the input x_i and the hidden node *j*, and bias respectively. The output of the *jth* hidden node is expressed as:

$$y_j = f(net_j) \tag{4.2}$$

$$f(net_j) = \frac{\ell^{net_j} - \ell^{-net_j}}{\ell^{net_j} + \ell^{-net_j}} \tag{4.3}$$

The β_j contribute to the right or left shift of the activation function depending on the value (positive or negative) taken by the β_j. The input to the *kth* output node is expressed as:

$$net_k = \sum_{j=1}^{N_q} y_j w_{jk} \tag{4.4}$$

The output of the ANN in Figure 4.1 can be presented as expressed in Eqn. (4.5)

$$y_k = f(net_k) \tag{4.5}$$

The overall performance of the ANN depends on the error between the trained ANN model output and the actual value. The lower the error the better is the accuracy, and vice versa.

The backpropagation is a gradient method for minimizing error cost function. In backpropation, the synaptic weights W are updated as follows:

$$W_{k+1} = W_k + \Delta W_k \tag{4.6}$$

Where k is the iteration in discrete time and ΔW_k is the current weight adaptation given by

$$\Delta W_k = -\eta \frac{\partial_{ek}}{\partial W_k} \tag{4.7}$$

Where η is the learning rate (typically between 0 and 1), and $\frac{\partial_{ek}}{\partial W_k}$ gradient of the error cost function to be minimized (Haykin, 1999).

4.3 Genetic Algorithm

Evolution is a procedure in which organism evolves over generations through genetic operators (Yacci, 2009). Evolutionary computation is learning algorithms that repeatedly search for an optimum solution to a problem by mimicking the natural biological processes (Searson, 2005). It accepts a problem as input and produces a set of solutions as output, the three major components of such computations are: initialization, processing of generations and post – processing (Barton, 2009; Foster, 2002). The idea of GA (formerly genetic plans) was conceived by Holland (1975) in pursuit of optimum solution to problems based on the principles of natural selection and natural genetics (Sakawa, 2002; Coley, 1999; Goldberg, 1999). They hinge on Darwin's theory as an inspirational guide and carefully learned the principle of evolution and applied the tacit knowledge acquired to develop algorithms based on selection in biological genetic systems (Hamdan, 2008; Van, Jain & Johnson, 1998). The concept of GA was derived from evolutionary biology and survival of

the fittest (Capraro *et al*., 2008). There is no consensus among the evolutionary computation community on the definition of GA (Mitchell, 1999). According to Bartschi (1996) GA is an algorithm that evolves the initial population through selection, crossover, and mutation. Stochastic and meta-heuristic searches were evolved from GA which allow the GA to operate on any data type according to the operators of such data. GA maintained a fixed number of solutions through the repeated procedure at every generation. At each new population, individual fitness is evaluated and the next population of solutions is created according to the fitness value of each chromosome (Grefenstte, 1986).

4.4 Genetic Algorithms Terminologies

4.4.1 Fitness

Naturally, survival of the fittest is the principles obeyed by nature, such that individuals with superior fitness possess a better chance to reproduce than individuals with an inferior fitness rate. This technique is employed by GA in which a high value of fitness indicates high superiority and a low value suggest inferior fitness. GA uses the fitness function to determine the best solution of the target problem to be resolved. The fitness value is an indicator in GA which is used to select individual in the present population to produce the next generation (Sakawa, 2002).

4.4.2 Gene, Chromosome, Allele, Phenotype, Genotype and Breeding

The building blocks of GA are called, genes (bit strings of arbitrary length). A sequence of genes is called a chromosome. Possible solutions to a problem may be described by genes without really being the answer to the problem. The smallest unit in the chromosome is referred to as an allele, represented by a single symbol or binary bit. A phenotype gives the

external description of an individual while a genotype is a deposited information in a chromosome (Sivanandam & Deepa, 2008; Rothlauf, 2002). The concepts of these terminologies are shown in Figure 4.2 adopted from (Sivanandam & Deepa, 2008).

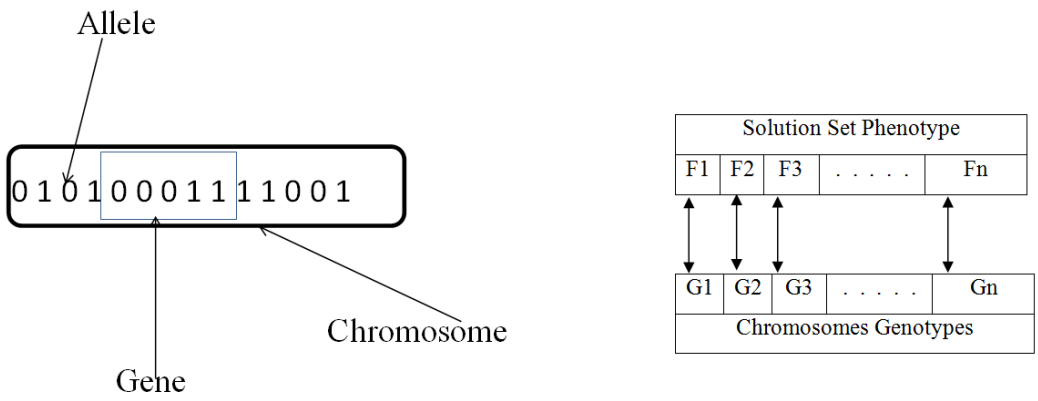


Figure 4.2: Representation of Allele, Gene, Chromosome, Genotype and Phenotype

4.4.3 Encoding and Decoding

Various representations are applied to encode most optimization problems in the form of binary or continuous string (Rothlauf, 2002). Both binary and continuous GA use genetic recombination and natural selection in modeling, but binary GA works with the binary string to minimize the cost function while continuous GA works with continuous variables to minimize the cost function (Huapt & Ellen, 2004). The problem to be optimized is converted into a form suitable for handling by GA. Each genetic individual is encoded by assigning it a value to represent the optimization problem variable as shown in Figure 4.3. One of the characteristics of GA is to use binary strings for representation of genetic individuals in a population, despite the fact that binary bits are used for representation, it

does not mean a binary system is applied for decoding the binary string of individuals automatically (Riechmann, 2001).

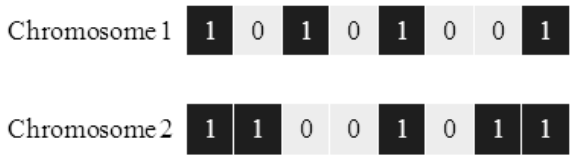


Figure 4.3: Binary Encoding. Source: Malhotra *et al*. (2011)

4.4.4 Crossover (Single Point)

This is a random point locus in an encoded bit string. In crossover, bits of string of the chromosomes are exchanged at random sites along the length of the chromosomes to have two new offspring. The offspring is formed by combining fragments of the parents bit strings (Mitchell, 1999; Hideyuki, 1997) as presented in Figure 4.4. The crossover probability indicates how often the crossover will be performed. A 100% probability, indicates that all offspring is made by crossover, while 0%, indicates that the chromosome of the present offspring will be the exact replica of the old generation. The crossover is used in GA operation for the production of better chromosomes containing the best part of the old chromosomes. The survival of some segment of the older population into the next generation is allowed. There are other crossover algorithms including: two point, multi - point, uniform, three parents and crossover with reduced surrogate among others. However, single point crossover is preferred because of its superiority over the others, as listed earlier, in terms of not destroying the building blocks while additional points reduce the GA performance. The exchange of genes between the parents to produce offspring is referred to as TailSwap, please refer to (Sivanandam & Deepa, 2008) for details.

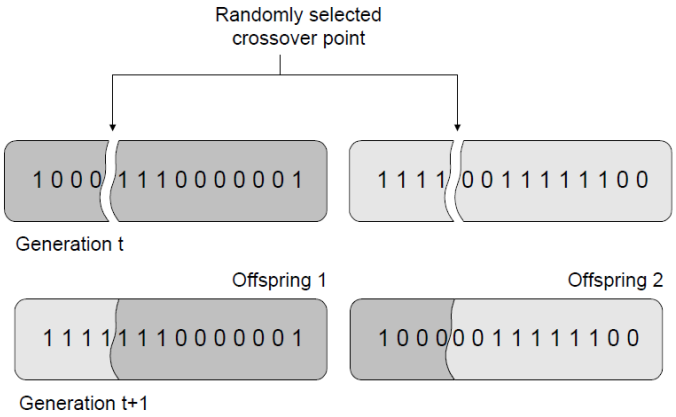


Figure 4.4: Crossover (single point) Source: Searson (2005)

4.4.5 Mutation

This is the creation of the offspring from a single parent by inverting one or more randomly selected bits of chromosomes of the parents as showed in Figure 4.5. The mutation can be performed on any bit using a small probability fraction, e.g. 0.001 (Mitchell, 1999). Strings that are developed from crossover are mutated to avoid the possibility of being stuck in local minima. Genetic materials that are lost in the process of crossover and the distortion of genetic information are fixed by mutation. The mutation probability indicates how often parts of the chromosomes will be mutated. If the mutation probability is 100%, then the whole chromosome is mutated, if it is 0% then the mutation is changed. The decision to mutate sections of the chromosomes depend on the mutation probability, if mutation is not applied the offspring are generated immediately from crossover without any part of the chromosomes being tempered (Sivanandam & Deepa, 2008). Other types of mutation exist, such as Gaussian mutation, uniform mutation, etc.

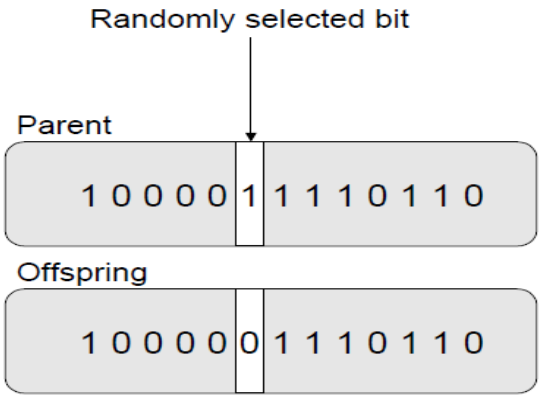


Figure 4.5: Mutation (single point) Source: Searson (2005)

During GA operation, Gaussian mutation is responsible for distinguishing phenotype from genotype, so that the binary representations of the problems can be used effectively. The Gaussian mutation is simulated in the GA by adding the GA gene (Hinterding, 1995). In uniform mutation, the value of the chosen gene is replaced with the uniform random value chosen by the user between upper and lower values, e.g. between 3 to 6 (Hasan *et al.*, 2011).

4.4.6 Search Space

Most solutions to a problem work to reach a specific target objective, with searching being the major component, ranging from maze running through allocation of resources, up to complex scheming in both government and private sectors, including research (Holland,1992). Specifically, the best solution in a pool of alternative solutions is obtained by searching the location of the best solution. When solving a problem, we are typically looking for some solutions, which will be the best among others. The space of all the feasible solution is called the search space, each location in a search space represents a possible solution to a problem. Thus, the fitness value of each possible solution is marked

based on the classification problem. The best solution among possible solutions is represented by a single location in the search space; GA is used to search and locate the best solution (global optimum or maximum) in the search space (Sivanandam & Deepa, 2008; Ghai, 2006).

4.4.7 Population

Individuals gathered in a group are called a population as shown in Table 4.1. Each individual in the population is evaluated to measure its fitness value. Individuals and certain information about the search space are defined by phenotype parameters. Initial population and population size are the two major features of population in GA. The complexity of the problem to be solved determines the population size which is typically generated randomly, referred to as the initialization of the population (Sivanandam & Deepa, 2008).

Table 4.1: Population

	Individuals	Encoded
Population	Chromosome 1	1 1 1 0 0 0 1 0
	Chromosome 2	0 1 1 1 1 0 1 1
	Chromosome 3	1 0 1 0 1 0 1 0
	Chromosome 4	1 1 0 0 1 1 0 0

Source :Sivanandam & Deepa (2008)

4.4.8 Reproduction

Based on the fitness values, two mates are selected from the population, in which, the mating pool only accommodates chromosomes with higher fitness value while chromosomes with lower fitness values are denied access. At the first instance, the values

of the fitness are mapped to positive values which are upturned for the purpose of minimization, thereby, assigning the lowest value with the highest probability (Harpham *et al*., 2004).

4.4.9 Selection

In this stage, the best 50% of genetic individuals with high value of the fitness function in a population are selected for reproduction, while those with poor value are rejected, thereby, dropping the genetic variety of the population. The distinction is not made between good and very good among the 50% surviving genetic individuals (Coley, 1999). The reason for the selection procedure in GA is to give prominence to individuals with higher fitness value so that their offspring will have a high fitness. The parents involve in the reproduction of the offsprings are chromosomes selected from the initial population (Sivanandam & Deepa, 2008). The basic selection process in GA is illustrated in Figure 4.6 adopted from (Sivanandam & Deepa, 2008), Rank Selection (Figure 4.7) and Roulette Wheel (Figure 4.8).

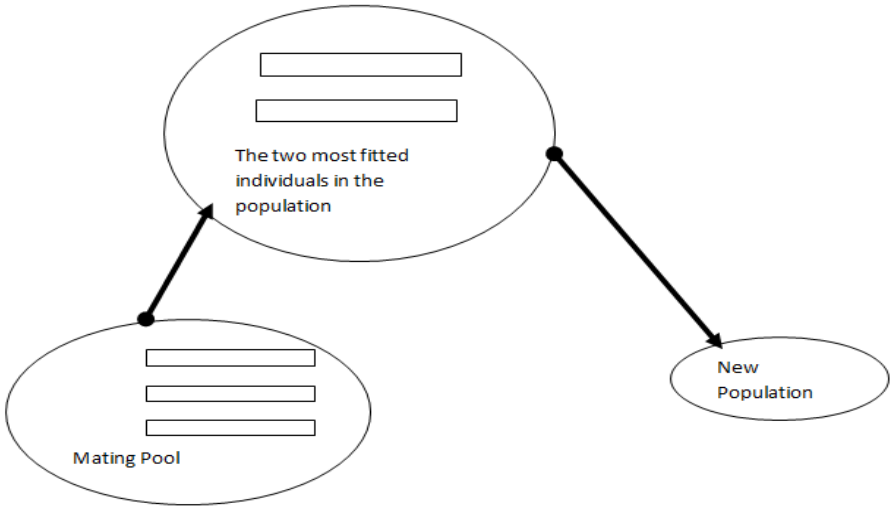


Figure 4.6: Basic selection process

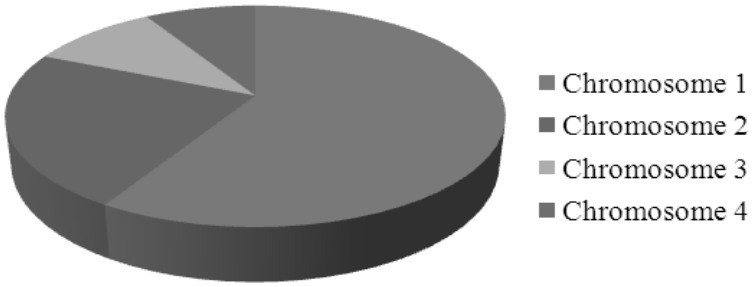


Figure 4.7: Rank Selection Malhotra *et al*. (2011)

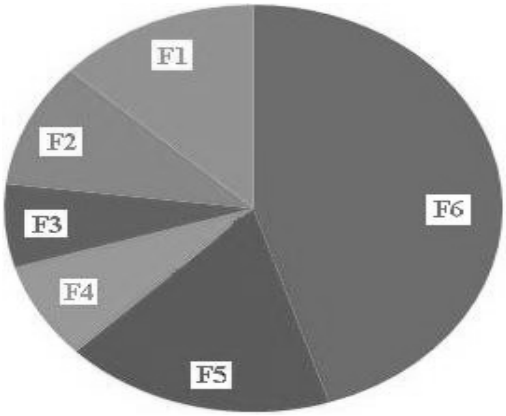


Figure 4.8: Roulette Wheel. Source: Malhotra *et al*. (2011)

4.4.10 Elitism

Elitism is the process of selecting the best individuals or selecting individuals with a bias towards the better ones. Elitism allows solutions to get better all the time and the population will converge quickly. At the stage of elitism, no genetic individual has the assurance of being selected automatically including the most fitted individual. This seems contrary to productivity, but is good for some type of problem solving because it retards the algorithms speed as more room is given to the GA to explore the search space to converge. In elitism 0 or 1 can be used to indicate the application of elitism. For example, 1 represents elitism whereas 0 specified no elitism (Coley, 1999).

4.4.11Convergence and Search Termination

Convergence is the point of acceptable solution to the problem, in other words, it is the stage in which the performance has stopped improving. Sivanandam & Deepa (2008) pointed varieties of conditions for terminating the GA process as follows: maximum generations, Elapsed time, unchanged Fitness Value, stalled generations and stalled time limit**.**

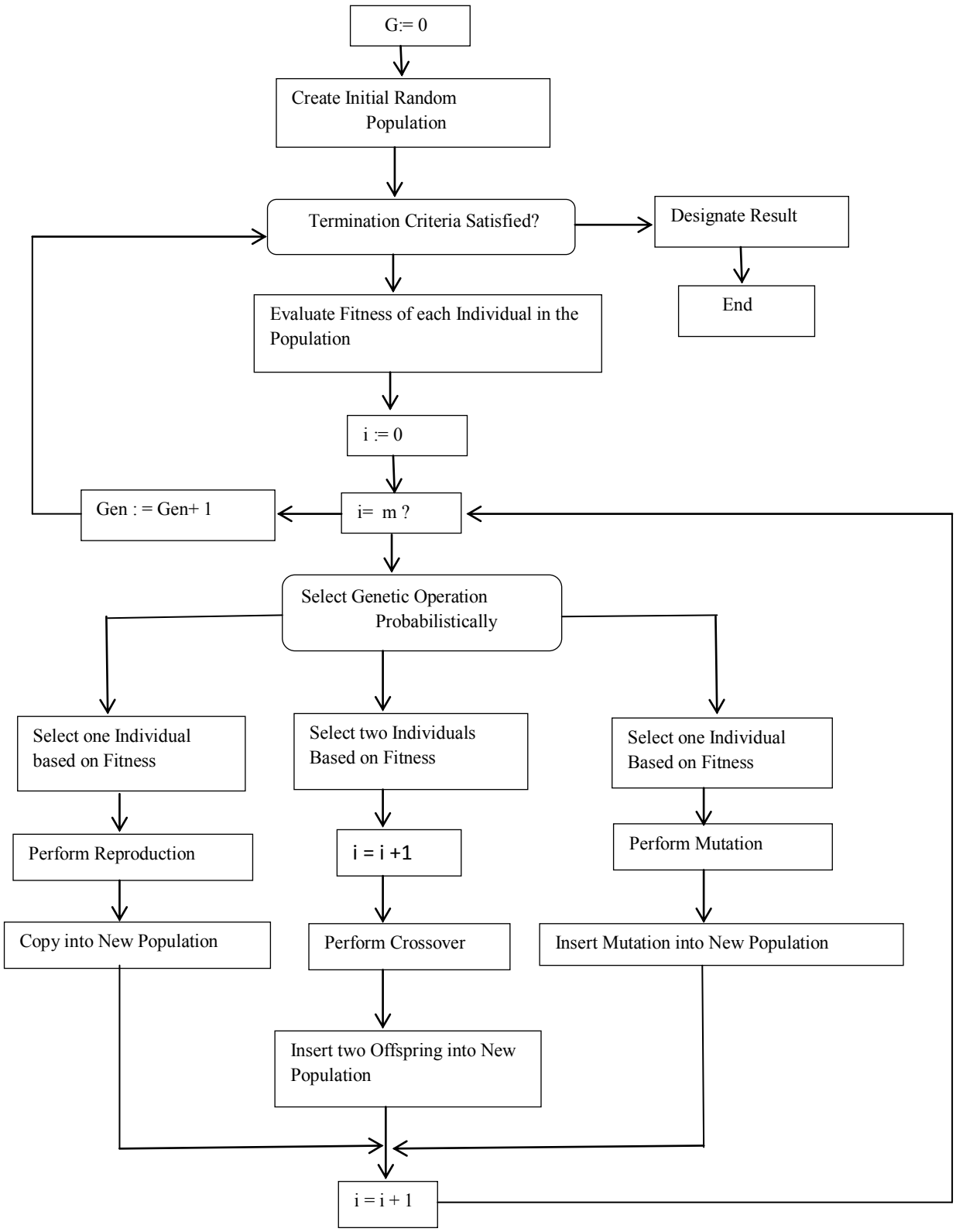


Figure 4.9: Genetic algorithm flow chart. Source: Koza (1998)

The GA circle is summarized as illustrated in Figure 4.10 and the flow chart showing the stages of implementation is shown in Figure 4.9.

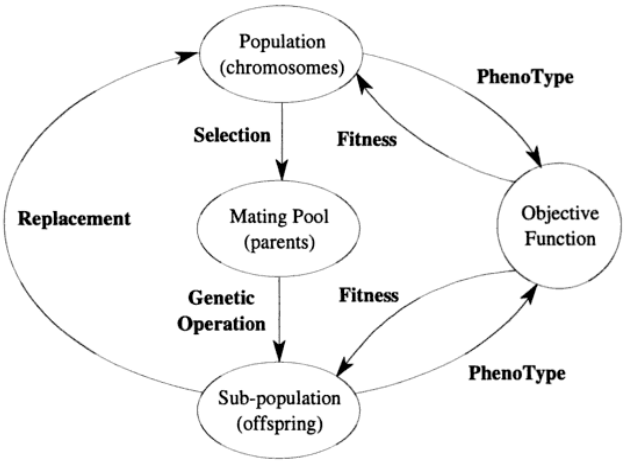


Figure 4.10: Genetic algorithm cycle. Source: Man *et al*. (1999).

4.5 Strengths of Genetic Algorithm

GA have the capability to accommodate different types of problems either continuous or discrete. This makes the GA very flexible in dealing with a variety of optimization problems, such as, prediction, classification, clustering, etc. In addition, GA performs very well using experimental data or analytical functions (this means prediction or classification data as well as benchmark functions such as, Rosenbrock's, De Jong, Rastrigin**,** Schwefel, etc.) (Haupt and Ellen, 2004). As such, the GA attracted attention of the research community in the applications of the GA to different domains (Nagata and Hoong, 2003).

GAs does not require the derivation of information before solving the problem: GAs typically holds a single solution to a problem at a particular time, and search to find out the next direction of movement based on the gradient function of the latest solution. When the GA decided the distance to move, then new solution is selected.

In a large sample of cost surface the GA operates simultaneously (that means the GA do not operate sequentially) in searching for optimum solutions. The GA does not search sequentially for solutions which typically take a long period of time. The GA simultaneously searches for the best solution to a problem. For example, five ANN with different configurations such as hidden nodes, weights, bias, etc. can be optimized simultaneously by GA not one ANN at a time. This characteristic of the GA enhances the GA convergence speed.

Local minima can effectively be avoided by the GA and to provide several alternative solutions to a problem. Local minima are a poor solution that pretends to be the best, through which algorithms can be deceived from reaching the optimal solution, but GA has the capability to avoid the local minima. This is one of the major attractive characteristics of the GA. In GA, variables are encoded to provide way for easy optimization, and probabilistic transformation rather than deterministic (Haupt and Ellen, 2004).

4.6 Neuro-Genetic model

In a Neuro-Genetic model, GA optimizes the weights, bias and hidden layer nodes of the ANN for the projection of crude oil price, by minimizing Eqn. 4.8 (the error between projected and actual price) and to maximize Eqn. 4.9, the percentage of accuracy (*ac*).

$$fitness = \frac{1}{\partial}\sum_{i=1}^{\partial}(\vartheta_i - \varphi_i)^2 \tag{4.8}$$

$$ac = \frac{1 - abs\ \ (fitness)}{\varepsilon - \phi} \times 100 \tag{4.9}$$

Where , ϑ_i, and φ_i are total points of data, crude oil price project by the Neuro-Genetic model, and actual crude oil price respectively. In Eqn. (4.8), ε, and ϕ are the upper and lower limits of the normalized data respectively, as recommended by Clementine Manual (2007).

4.7 Summary

The ANN basic theoretical background and operations, including mathematical representation and diagram are presented in this chapter. The GA theory is also discussed, including the major GA operators: crossover, mutation and population. The basic concepts of ANN and GA intended for application to develop the Neuro-Genetic model for the projection of crude oil price are briefly introduced so that any reader can understand how these techniques operate and how they can attain their main goals.

CHAPTER FIVE

DATASET AND EXPERIMENTAL SETUP

5.1 Introduction

Data collected for this research is real world data, and thus, needs cleaning and pre-processing in order to improve the performance of the model. Data preparation activities including data collection, data cleaning analysis, attribute selection, attribute subset selection, normalization and the correlation coefficient analysis of the attributes are discussed in this chapter. The main goal of the chapter is to generate quality data by recovering missing values, purification of the datasets, settling of discrepancies such as missing values, input attributes subset selection, etc. Subsequently, the chapter discusses the proposed approach and experimental setup based on the cleansed dataset.

5.2 Data Collection

Monthly Brent crude oil in USD/barrel (42 US gallons), widely used as a benchmark for formulating oil price (refer to chapter two, section 2.5), was chosen as the international crude oil price. The Brent crude oil price was accessed from the EIAUSDE official website (http:// www.eia.doe.gov) over the period from May 1987 through December 2011. These data are freely available from the official website of the organization. The EIAUSDE gathers, analyzes, and disseminates impartial information about energy to assist policy makers in making good policies in the energy sector. The data of Brent, OPECCP, OECDES, WCOP, NOPECCP, USCOP, USCOI, OECDCOC, USCOSR, USESTG, and USCOS (refer to chapter 2, section 2.5.2 for full descriptions including their respective units) were collected from various repositories of the websites (http:// www.eia.doe.gov) to create a central database. The first column of Table 5.1 shows the attributes of the dataset

whereas the second column provided the links in the website through which the data were retrieved.

Table 5.1: Sources of the data from different repositories at http:// www.eia.doe.gov

Attribute	Data source (links)
OPECCP	http://www.**eia.gov**/cfapps/ipdbproject/iedindex3.cfm?tid=50&pid=53 &aid=1&cid=CG9,&syid=1994&eyid=2012&freq=M&unit=TBPD
OECDES	http://www.**eia.gov**/cfapps/ipdbproject/iedindex3.cfm?tid=50&pid=5& aid=5&cid=CG5,&syid=1986&eyid=2012&freq=M&unit=MBBL
WCOP	http://www.**eia.gov**/cfapps/ipdbproject/iedindex3.cfm?tid=50&pid=57 &aid=1&cid=ww,&syid=1994&eyid=2014&freq=M&unit=TBPD
NOPECCP	http://www.**eia.gov**/cfapps/ipdbproject/iedindex3.cfm?tid=50&pid=57 &aid=1&cid=&syid=1999&eyid=2014&freq=M&unit=TBPD
USCOP	http://www.**eia.gov**/dnav/pet/pet_crd_crpdn_adc_mbbl_m.htm
USCOI	http://www.**eia.gov**/dnav/pet/pet_move_wkly_dc_NUS-Z00_mbblpd_w.htm
OECDCOC	http://www.**eia.gov**/cfapps/ipdbproject/iedindex3.cfm?tid=50&pid=54 &aid=2&cid=CG5,&syid=1995&eyid=2014&freq=M&unit=TBPD
USCOSR	http://www.**eia.gov**/dnav/pet/pet_stoc_wstk_dcu_nus_w.htm
USESTG	http://www.**eia.gov**/dnav/pet/pet_stoc_wstk_dcu_nus_w.htm
USCOS	http://www.**eia.gov**/dnav/pet/pet_cons_psup_dc_nus_mbbl_m.htm
Brent	http://www.**eia.gov**/dnav/pet/pet_pri_spt_s1_d.htm

5.3 Data Analysis and Cleansing

As was previously mentioned, the data collected for this research are real-world data. Therefore, cleaning and preprocessing must be performed on the collected data to improve the quality of the data, and thus improve the performance of the Neuro-Genetic model. Although the data was collected from a credible source, it was still thoroughly analyzed for possible missing values, outliers, and other inconsistencies. The majority of the data in the dataset was found to be complete. Only data for US domestic crude oil supply and OPEC crude oil production were found to have missing values, but the missing data are only a small fraction (refer to Appendix A). This is not surprising, as stated by Gheyas and Smith (2010) it is not possible to have real world data without any missing values. Figure 5.1 provides a graphical representation of the time series for illustrating the attributes of the

dataset for easy identification of potential discontinuities, outliers, noise and missing values. Figure 5.2 is the attribute representation formulated based on the input and output attributes (refer to section 2.5.2). Imputation technique was used to fill in the missing values by adding zeros to all existing empty spaces because ANN is robust in handling incomplete or noisy data of which both are properties of time series projection (Versace *et al*., 2004).

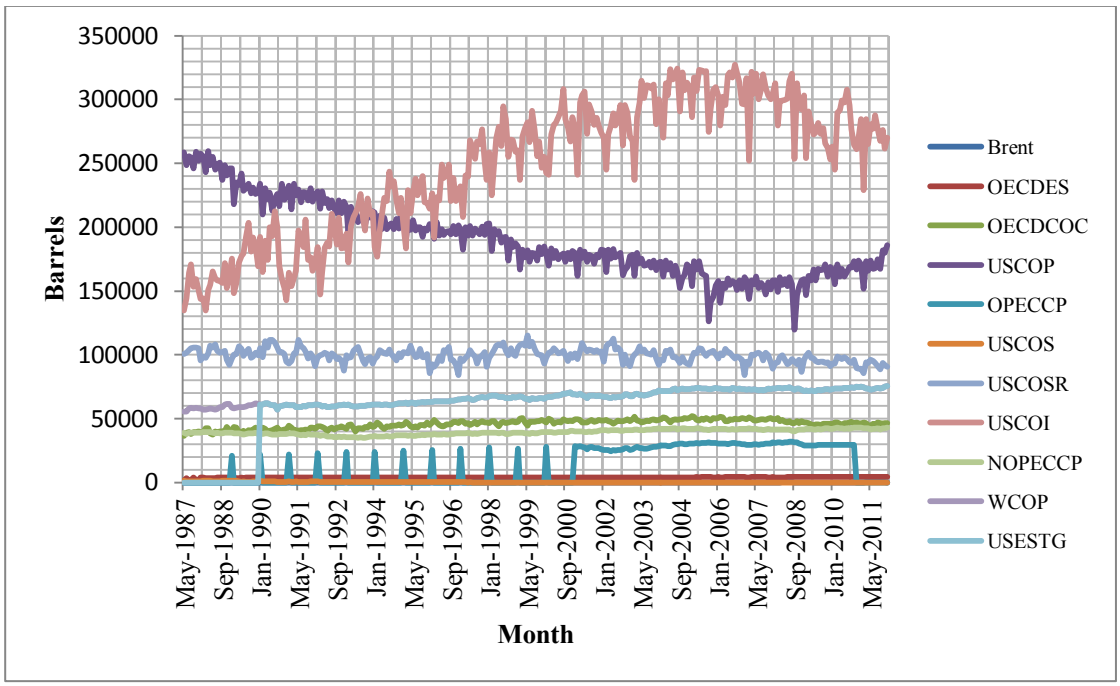


Figure 5.1: Graphical representation of the dataset

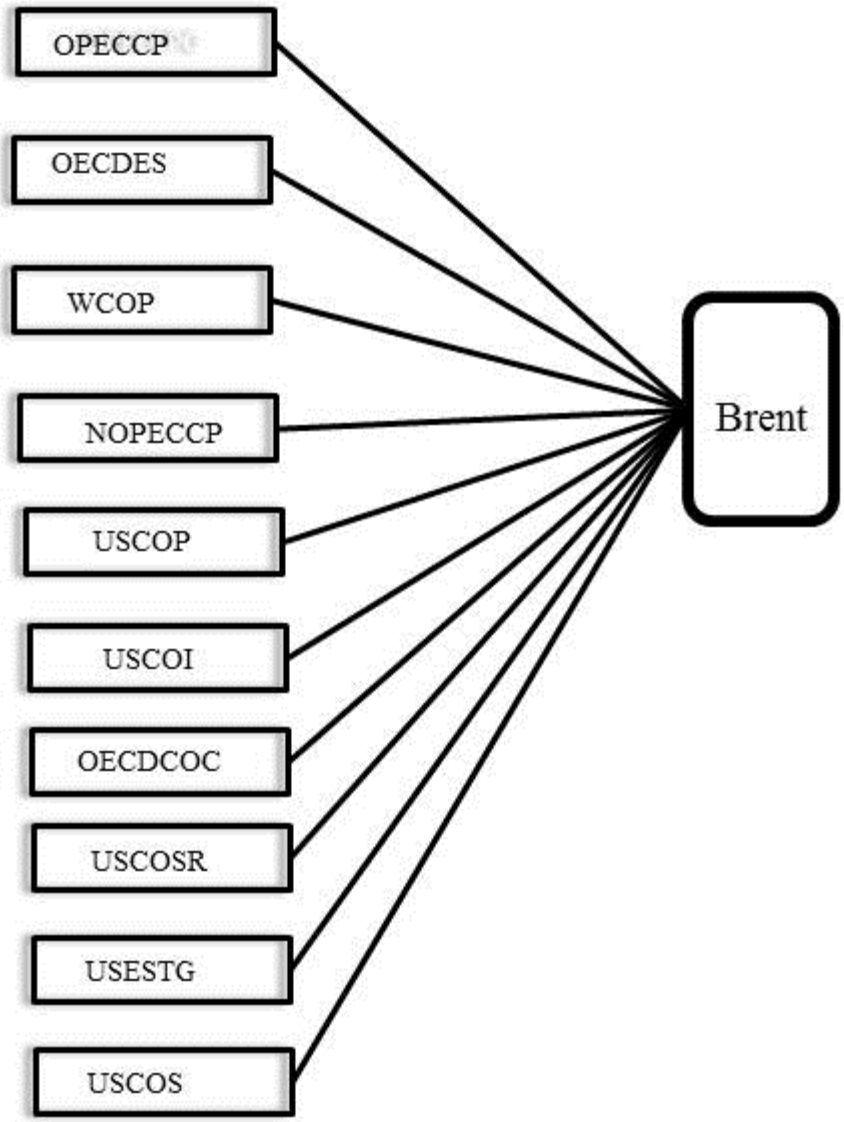


Figure 5.2: Attributes of the dataset

5.4 Transformation of the Experimental Data

Preliminary investigation is required to determine the best opinion about data standardization methods. Therefore, to address these arguments in the domain of crude oil price projection, we compared the projection performances of the two methods (data normalization and raw data). Two databases were created: the first database contained the normalized data, and the second comprised the raw data. In normalizing the data, there are

several ranges of scales that are used by scholars. For example, the commonly use ranges of normalized data are within 1 to 0, 0.8 to 0.8, 2 to 2, 3 to 3 among others. Conversely, using such ranges to scale the original data could be disastrous because the scale data lacks the uniform distribution that is expected by the nonlinear (e.g. sigmoid) activation function. These ranges of data normalization lump on only the positive side without considering the negative side, which could generate poor results. The preferable range would be to scale both the negative and positive sides, for instance the data can be normalized between -1 to 1 (Azoff, 1994). Thus, the dataset in this research was normalized between the range of -1 to 1 using Equation (5.1). Numerical examples of the sample of raw data before normalization are presented in Table 5.2 and the complete dataset is provided in Appendix A. The same sample is shown in Table 5.3 after the normalization procedure.

$$N_r = \frac{M_i - M_{min}}{M_{max} - M_{min}} \tag{5.1}$$

where N_r, M_i, M_{min}, *and* M_{max} are the normalized data, raw data, the minimum value of the data, and maximum value of the data, respectively. This approach enables the sigmoid transfer function in the hidden layer neurons to compute both the negative and positive data points.

Table 5.2: Sample of raw data before normalization

No.	OECDES (mb/d)	OECDCOC (tb/d)	USCOP (tb/d)	USCOS (tb/d)	USCOSR (tb/d)	USCOI (tb/d)	NOPECCP (tb/d)	WCOP (tb/d)	Brent ($/barrel)
1	1562	36417.91	258426	1310	100773	134549	38884.04	55674.96	18.58
2	3454	38864.65	248356	1087	101755	144197	38055.35	55375.71	18.86
3	1580	39522.88	255782	981	103457	164131	38996.7	57939.92	19.86
4	1616	38038.39	254508	954	105394	170802	38851.39	58737.6	18.98
5	3619	39388.79	246163	832	105794	153312	39117.58	58130.08	18.31
6	1637	39437.97	259272	777	105521	159390	39147.31	58325.25	18.76
7	1662	39729.91	251915	758	105296	150389	39168.45	57862.85	17.78
8	3643	39428.72	257844	964	95851	143825	39215.22	57936.91	17.05
9	3637	39661.77	255743	1383	98805	144515	39285.66	57137.66	16.75
10	3571	42468.69	242848	1504	97616	134863	39373.71	57217.71	15.73
11	3496	42379.83	259587	1609	101042	150912	39571.92	57578.92	14.73
12	3520	38609.68	248629	1259	107276	155012	39358.1	57890.1	16.6
13	3605	37528.04	255089	1042	105691	165510	39112.93	57606.93	16.31
14	3608	40004.52	245102	967	107954	159667	38465.74	57271.74	15.54
15	3690	39143.87	249229	890	103275	158100	38799.99	57695.99	14.91
16	3691	40136.57	250459	917	101807	157766	38706.67	58852.67	14.89
17	3681	40202.82	236847	1121	102094	156355	38385.88	59206.88	13.18
18	3692	40605.9	248704	1308	103149	172080	38868.87	60890.87	12.41
19	3656	43416.84	240700	1319	96092	152093	38732.14	61350.14	13.02
20	3588	40745.44	246189	1378	92129	162132	38813.29	61597.29	15.31
21	3646	41106.85	246056	1463	97056	175497	38761.55	58706.81	17.17
22	3696	43349.78	218076	1344	100064	148539	38471.31	58227.58	16.89
23	3537	42956.21	234822	1380	100856	156100	38550.51	58629.78	18.7
24	3575	39519.13	233147	689	106668	172496	38354.68	59059.99	20.32
25	3629	39091.25	242303	577	104770	177604	38204.64	58980.95	18.63
26	3603	40722.96	228731	614	99200	179269	37638.74	59017.97	17.67
27	3681	39036.04	230766	575	101924	192628	38099.91	59536.22	17.62
28	3723	40991.57	233878	524	104758	203517	38524.43	60428.66	16.77
29	3725	40500.17	226436	548	97697	180834	38466.56	60511.7	17.77

Millions of barrels per day (mb/d), thousands of barrels per day (tb/d)

Table 5.3: Sample of normalized data

No.	OECDES (mb/d)	OECDCOC (tb/d)	USCOP (tb/d)	USCOS (tb/d)	USCOSR (tb/d)	USCOI (tb/d)	NOPECCP (tb/d)	WCOP (tb/d)	Brent ($/barrel)
1	-1.0129823	-1.15438	0.983427	0.628341	0.082096	-1.00326	-0.04045	-0.9703	-0.85745
2	0.3516048	-0.81384	0.839683	0.35115	0.145055	-0.90308	-0.24331	-1	-0.85289
3	-1.0000000	-0.72223	0.945686	0.219391	0.254175	-0.69609	-0.01287	-0.7462	-0.83662
4	-0.9740353	-0.92884	0.9275	0.18583	0.378362	-0.62683	-0.04844	-0.6672	-0.85094
5	0.4706094	-0.74089	0.808379	0.034183	0.404007	-0.80843	0.016722	-0.7273	-0.86184
6	-0.9588893	-0.73405	0.995504	-0.03418	0.386504	-0.74532	0.024	-0.7080	-0.85452
7	-0.9408583	-0.69342	0.890486	-0.0578	0.372079	-0.83878	0.029176	-0.7538	-0.87046
8	0.4879192	-0.73534	0.97512	0.19826	-0.23347	-0.90694	0.040625	-0.7465	-0.88234
9	0.4835918	-0.7029	0.945129	0.71908	-0.04408	-0.89978	0.057868	-0.8256	-0.88723
10	0.4359899	-0.31223	0.761059	0.869484	-0.12031	-1	0.079422	-0.8176	-0.90382
11	0.3818969	-0.3246	1	1	0.099343	-0.83335	0.127944	-0.7819	-0.9201
12	0.3992066	-0.84933	0.84358	0.564947	0.499022	-0.79078	0.075602	-0.7511	-0.88967
13	0.4605121	-0.99987	0.935793	0.295214	0.397403	-0.68178	0.015586	-0.7791	-0.89439
14	0.4626758	-0.65519	0.793234	0.201989	0.542491	-0.74245	-0.14285	-0.8123	-0.90692
15	0.5218175	-0.77498	0.852145	0.106277	0.242507	-0.75872	-0.06102	-0.7703	-0.91717
16	0.5225388	-0.63681	0.869702	0.139838	0.148389	-0.76219	-0.08387	-0.6558	-0.91749
17	0.5153264	-0.6276	0.675398	0.393412	0.16679	-0.77684	-0.1624	-0.620	-0.94532
18	0.5232600	-0.57149	0.844651	0.625855	0.234429	-0.61356	-0.04416	-0.4541	-0.95785
19	0.4972953	-0.18027	0.730398	0.639528	-0.21802	-0.82109	-0.07763	-0.4086	-0.94793
20	0.4482510	-0.55207	0.80875	0.712865	-0.47209	-0.71685	-0.05777	-0.3842	-0.91066
21	0.4900829	-0.50177	0.806852	0.818521	-0.15621	-0.57808	-0.07043	-0.6702	-0.88039
22	0.5261450	-0.1896	0.407451	0.670603	0.03664	-0.85799	-0.14148	-0.7177	-0.88495
23	0.4114677	-0.24438	0.646492	0.715351	0.087418	-0.77948	-0.12209	-0.6779	-0.85549
24	0.4388749	-0.72275	0.622582	-0.14357	0.460042	-0.60924	-0.17003	-0.6353	-0.82913
25	0.4778219	-0.7823	0.75328	-0.28278	0.338356	-0.5562	-0.20676	-0.6431	-0.85663
26	0.4590696	-0.5552	0.559546	-0.23679	-0.01875	-0.53891	-0.34529	-0.6394	-0.87225
27	0.5153264	-0.78999	0.588595	-0.28527	0.15589	-0.4002	-0.2324	-0.5882	-0.87307
28	0.5456185	-0.51782	0.633017	-0.34866	0.337586	-0.28713	-0.12848	-0.4998	-0.8869
29	0.5470609	-0.58621	0.526786	-0.31883	-0.11511	-0.52266	-0.14264	-0.4916	-0.87063

Millions of barrels per day (mb/d), thousands of barrels per day (tb/d)

5.5 Attributes

In machine learning, a performance criterion is optimized through historical data. The learning system parameters are tuned through the training dataset. It is expected that during the training phase, the system learns the nonlinear relationship that maps the inputs to the

output vectors to project the answers for a test dataset that the system was not exposed to during the training phase of the modeling process. The projection accuracy of the learning system on the independent test data determines its generalization performance (Alpaydin, 2004). As noted in (Bishop, 1995), further partitioning of the data sets into training, validating and testing improves the ANN projection accuracy. A good method is required to find the optimal solution with a smaller number of iterations and a shorter computation time. In some applications, the computation time of the algorithm can be as important as the projection accuracy (Alpaydin, *et al*., 2004). This research attempts to establish an association among the training dataset, validation dataset, test dataset, computation time and the number of iterations. In this research, our datasets were partitioned into several ratios according to the literature (Witten and Frank, 2005; Sexton *et al*., 2004; Yu *et al*., 2009; Wang *et al*., 2010; Yao and Tan 2000; Abdel-Aal, 2008). The data partition ratios are reported in Table 6.1. Therefore, we formulated the following hypothesis for comparing the performances of the two methods:

Ho$_1$: *There is no significant difference between the projection accuracy of the normalized method and the projection accuracy of the raw method on the training dataset.*

Ho$_2$: *There is no significant difference between the projection accuracy of the normalized method and the projection accuracy of the raw method on the validation dataset.*

Ho$_3$: *There is no significant difference between the projection accuracy of the normalized method and the projection accuracy of the raw method on the test dataset.*

Ho$_4$: *There is no significant difference between the number of iterations required by normalized and the raw method to converge to the optimal solution.*

Ho_5: *There is no significant difference between the computation time of the normalized method and the computation time of the raw method required to converge to the optimal solution.*

Ho_6*: The computation time of the training, validation and iterations do not affect the ANN model performance.*

The ANN used in this research comprised 10 input neurons at the input layer that corresponded to the 10 independent attributes in the datasets. The dependent attribute, the Brent crude oil price, corresponds to the output node (the single output neuron). One hidden layer was used in the network. The number of hidden layer neurons and the activation function at the hidden layer were determined through the commonly used trial-and-error preliminary experimentation using small samples of the datasets. For fair comparison purposes, the preliminary experiments were conducted with the two separate data sets, and the optimal numbers of hidden layer neurons were found to be 10 and 12. The architectural configurations of the ANN for the normalized method was (10 – 10 – 1), while for the raw method, it was (10 – 12 – 1). In the hidden layer, the sigmoid activation function was used.

In the output layer, the linear transfer function was used (refer to chapter 4, section 4.2 for justification). The ANNs were created and trained with the Levenberg–Marquardt learning algorithm, and the network model was iteratively trained to minimize the MSE between the crude oil price projected by the ANN models and the actual price. In each iteration, the gradient of the MSE was used to adjust the ANN weights and biases. The validation dataset validates the generalization ability of the ANN during training. In addition, the training process was stopped when the validation MSE started to increase to avoid over-

fitting the data. Finally, the optimal weights and biases were returned. The test dataset was independently used to test the overall projection accuracy of the ANN models. We performed thirty six (36) experiments with varying data partitioning ratios to ensure consistency in the findings. Sixteen experiments were conducted with normalized data using the ANN with a network structure of 10 input neurons, 10 hidden neurons and 1 output neuron. Sixteen experiments were also performed with raw data using the ANNs configurations with 10 input neurons, 12 hidden neurons and 1 output neuron. The same data partitioning ratios were used in both of the methods for a fair comparison. The training, validation and test MSE at each experimental trial were recorded, as was the number of iterations and the computation time (Nanosec) required to converge to the optimal solution.

5.6 The Selection of Input Attributes Using Genetic Algorithm

Our goal was for the GAs to genetically evolve different architectures and parameters of ANN so that an optimal ANN structure that projects Brent crude oil price with the minimum error can be obtained. It was also required to extract the minimum subset of optimal input attributes. The dual goals led to the formulation of the fitness function defined as expressed by Eqn. 5.1 adapted from (Mantzaris *et al*., 2011):

$$fitness = \left(\sum_{k=0}^{N_1} \sum_{l=0}^{N_2} (T_{lk} - P_{lk})^2 \right) (N_1 N_2)^{-1} + N_3^{-1} \tag{5.1}$$

Where N_1 is the total number of the dataset, N_2 is the number of the input neurons. T_{lk} and P_{lk} are target Brent crude oil price for observation *l* at neuron *k*, and projected Brent crude oil price for observation *l* at neuron *k* and N_3 is the number of training datasets. Fitness of zero value indicates perfect Brent crude oil price projection; thus, the closer the fitness

value is to zero (0), the better is the projection. The GA used a binary encoding technique to build chromosomes of a fixed length (37) to represent each variety of ANN in the population. A string 0's and 1's represent an ANN, as shown in Figure 5.3.

1011110000111111011110011110110001 11

Figure 5.3: Chromosomes representing ANN

The GA started evolving with an initial population size of 50 ANNs, created in the form of a binary string. The chromosomes were randomly selected by roulette wheel from the population, based on the fitness value of each of the respective chromosomes to perform reproduction in order to create a new population (refer to Table 6.6 for the fitness values). Crossover probability of 0.6 and mutation probability of 0.001 were adopted from (Grefenstette, 1986) and applied to the new population because they control exploitation in the search space. Several crossover functions including single-point, uniform, two-point and scattered crossovers were experimentally tried. The two mutation functions, namely Gaussian and uniform (refer to section 4.4.5), were also tried experimentally so as to choose the most effective function. TailSwap (refer to section 4.4.4) was the method used for mating. The GA is specified to select the optimal input attributes from 1 to 10 since the maximum number of major input attributes related to the crude oil price as identified from the literature was 10 (refer to chapter 2, Tables 2.1 and 2.2). The selected input attributes serve as input to the ANN which represents the input neurons. Upper and lower limits of the hidden neurons were set at 1 and 256 respectively. One hidden layer was specified since this was sufficient to approximate any function. Data was partitioned into 70%: 15%: 15%. The GA optimized the weights of the network within the range of value limited to -1

and 1. Here, the GA genetically evolved several ANNs on the same dataset in order to realize the optimal ANN structure with the best combination of input attributes that yielded maximum accuracy within the shortest convergence speed. The GAs operation was set to terminate when it had run for seven generations without improvement in the fitness value of the optimal chromosome.

5.7 The State of the Art Methods of Input Attribute Selection

In the experiment, all the ten (10) inputs were used without selection of the most relevant inputs. Conventional methods were used for the projection of the crude oil price for the purpose of evaluating the effectiveness of our approach. The commonly used backpropagation ANN in the domain of crude oil price projection, as revealed from the literature reviewed (refer to chapter 3), was chosen as the benchmark. The BPNN parameters were determined through the commonly used trial-and-error technique.

The Karhunen-Loève Transform (KLT) commonly referred to as the PCA was used to reduce the dimension of the input attributes and the principal components selected through KLT were used as inputs to the backpropagation (BPNN) (KLTBNN); while others were rejected. The attributes selected by the KLT are WCOP, OPECCP, OECDES, OECDCOC, NOPECCP, USCOP and USESTG (please refer to section 6.2.7 for details). The KLTBNN was trained to build a model for the projection of Brent crude oil price.

The correlation coefficient was applied to determine the input attributes with a positive relationship with the Brent crude oil price. The input attributes with a positive correlation coefficient relationship were accepted and used as inputs to create the BPNN (rBPNN) model for the projection of Brent crude oil price (please refer to discussion of Table 6.7 in section 6.2.6 for details on empirical examples).

We used the trial and error (TE) method to select the optimal input attributes. Several combinations of the input attributes were experimented and minimum subset with superior performance was selected as the best. The best attributes were used as inputs to the BPNN (TEBPNN).

In the experiments involving the BPNN and the Haar Wavelet Transform (HWT), the HWT decomposition (please refer to section 3.4.3), the original Brent crude oil price data were not directly applied to the BPNN. The HWT was applied to decompose the original Brent crude oil price data into sub-series of scales so that the information could be captured on various scales (please refer to section 6.2.9 for empirical examples). Each of the sub-series was used to build a BPNN model for the projection of the Brent crude oil price. The HWT decomposition of the Brent crude oil price was used as training data to build a BPNN model (HWTBPNN) to project Brent crude oil price. The entire projection results yielded by the several models based on separate sub-scales were recombined by a HWBPPNN to produce an ensemble projection result (Refer to section 6.2.9 for details). It is well known that the data partition ratio significantly affects the result produced by ANNs, so our experiments with both the GA and the state of the art methods were repeated several times with varying data partition ratios in order to ensure reliable results.

5. 8 The Proposed Neuro-Genetic Model

To build a Neuro-Genetic model for the projection of crude oil price, it is required to minimize the fittest (the error between projected and actual price) and to maximize the percentage of accuracy. The dataset prior to the uncertainties (regular oil price fluctuation (ROPF see Figure 5.5)) are used through which the dataset for the initial modeling were partitioned into 80% for training and 20% for testing the generalization ability. The ANN

and GA parameters require initialization for the GA to start execution. The input neurons were set to five (5) because GA selected five (5) attributes as the optimal relevant input (refer to section 5.6 for details) and the inter correlation coefficient relationship (refer to section 3.5 for details about correlation coefficient) among the input attributes were investigated. The weights were initialized between -10 to 10. The activation functions in the hidden layer and output layer were set to tansig and logsig, respectively. The setting of the best parameters to GA values was determined by trial-and-error. We used initial experimental trials to obtain the optimal parameter settings. The process begins with the generation of random population of chromosomes in which the chromosome represents the set of Neuro-Genetic model parameters. The fitness of each chromosome in the population is evaluated. The GA continues evolving iteratively until the best fitness value in the population cannot improve before the evolution stop and return the optimal weights, bias and *hn* of Neuro-Genetic model as the best converged solution.

The optimal training parameters for the Neuro-Genetic model is reported in Table 5.4, the parameter settings are critical in the implementation of GA.

Table 5.4 The parameter values used in the Neuro-Genetic model

Parameter	Setting
Population size	20
Crossover rate	0.5
Mutation rate	0.01
Generation	200
η	0.6
Output layer activation function	logsig
Hidden layer activation function	tansig
Hidden layers	1
Input nodes	5
Hidden layer nodes	5
Structure of the Neuro-Genetic model	5-5 -1

To improve the projection accuracy and to capture the impact of uncertainties on crude oil price the Neuro-Genetic model has to be retrained with a relatively new subset dataset distorted by certain uncertainties. Therefore, the Neuro-Genetic model was progressively retrained with a relatively new dataset to learn and capture the new pattern in the dataset. The retraining is performed using the data distorted by uncertainties (see Figure 5.5) so that the model could capture changes that occurs in the dataset as a result of distortion through these uncertainties. Completely replacing the initial dataset with a new dataset can reduce the training data which would retard projection accuracy and a large sample of data is required for building a robust model. For that reason, we add a new set of data to the existing dataset while removing the old dataset to maintain a dataset with relatively updated information. The number of new datasets added into the initial data set is equal to the number of old datasets that is removed. For example, adding recent four months dataset warrant the removal of four months oldest datasets.

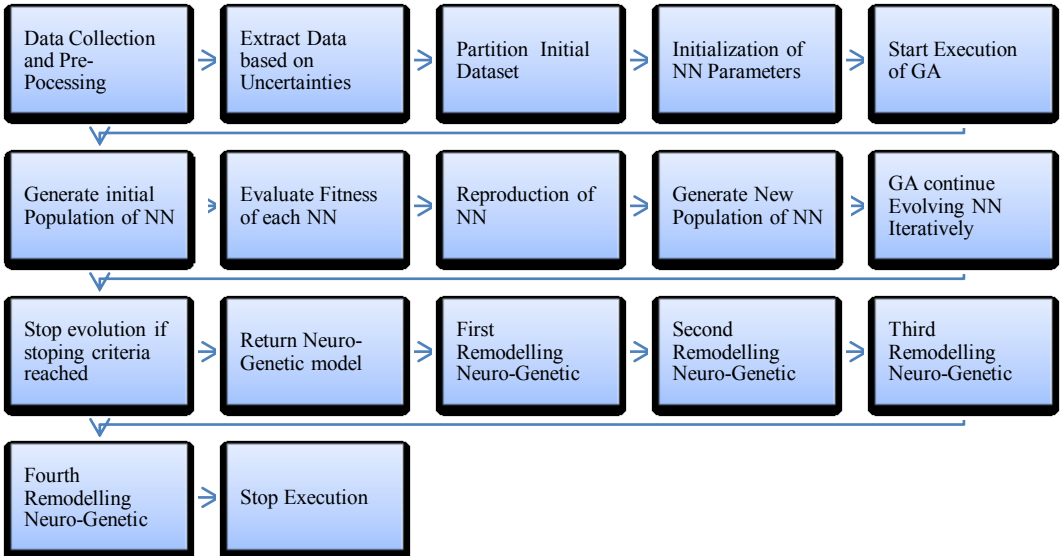


Figure 5.4: The proposed framework for the Neuro-Genetic model

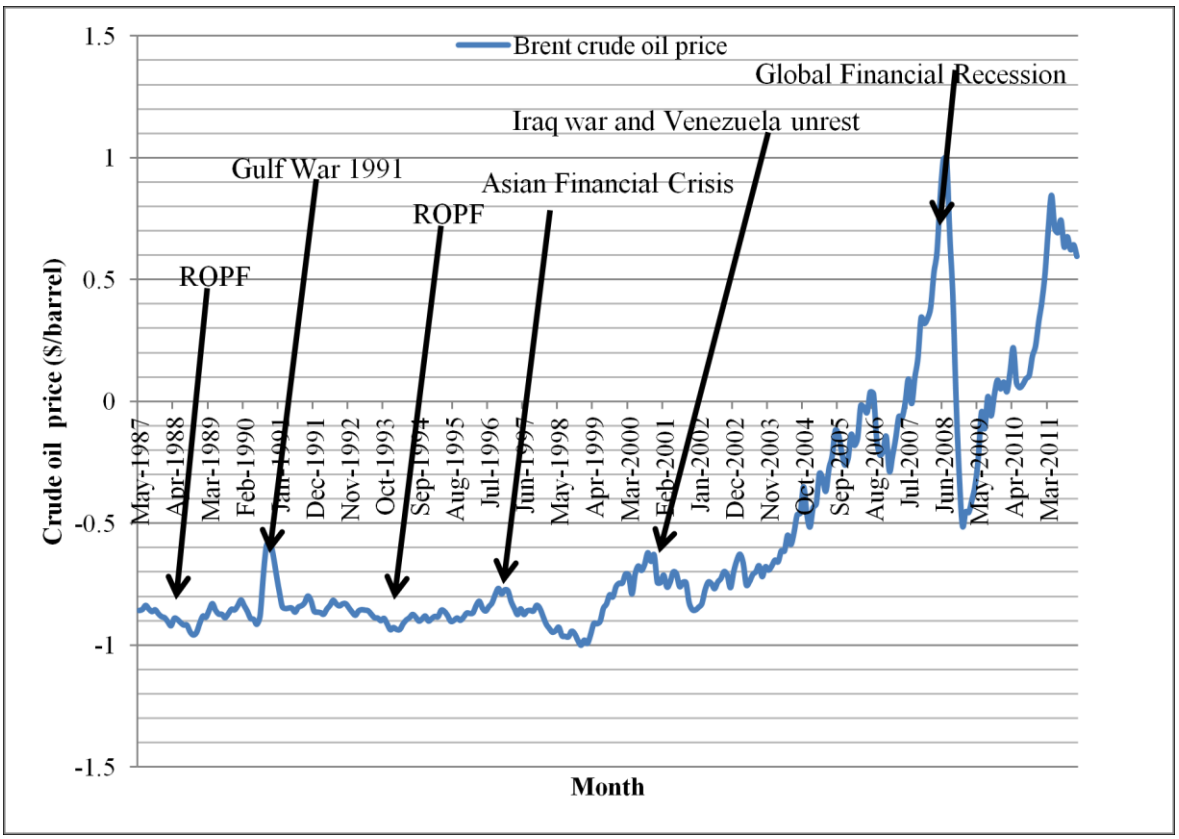


Figure 5.5: Indicating the impact of uncertainties on crude oil price and regular oil price fluctuation (ROPF) over a period of 1987 to 2011.

The Neuro-Genetic model was first developed with a dataset from May, 1987 to July, 1991 with ROPF as shown in Figure 5.5 prior to the uncertainties considered in our case study. The new dataset (August, 1990 to February, 1991) distorted after the first Gulf War of 1991 as indicated in Figure 5.5 were added into the dataset and the Neuro-Genetic model is retrained leaving out the oldest dataset. This process was periodically repeated with the dataset (July, 1997 to August, 1998) distorted during the Asian Financial Crisis of 1997 (see Figure 5.5). The Iraq war and Venezuela unrest occur in the same period (refer to Table 2.1), as a result, the two uncertainties were combined to have one period of distorted dataset (December, 2002 to May, 2003) as shown in Figure 5.5. Lastly, the dataset (December, 2007 to Jun, 2009) distorted during the Global Financial Recession of 2007 as

indicated in Figure 5.5 were used for retraining the model. In each of the retraining, the Neuro-Genetic model captured the impact of the uncertain event, learned new patterns, and subsequently projected the crude oil price. The complete procedure of the proposes methodology for the development of Neuro-Genetic model is presented in a block diagram in Figure 5.4 and the pseudo-code is presented in Figure 5.6.

1. ***Initializes*** random population of chromosomes representing W, β, and *hn*
2. *Load* the initial dataset prior to uncertainties (May 1987 to July 1991)
3. ***Evaluate*** the fitness of each chromosome in the population
4. While ***fitness*** = stopping criteria
5. Compute *ac* and go to step 18 otherwise
6. Pass the GA ***chromosomes*** as W, β , and *hn* to the ANN
7. ANN runs using the initialized W, β, and *hn* with the GA operators
8. Generate new population using ***roulette-wheel*** selection
9. Create new chromosomes by mating using ***crossover***
10. ***Mutate*** the chromosomes
11. W, and β are updated through Eqn. (5.4) to (5.7)

$$\nabla W_{k(ji)} = W_{x(ji)} + \eta fitness_j h_j \quad (5.4)$$

$$\nabla \beta_{k(ji)} = \beta_{x(ji)} + \eta fitness_j h_j \quad (5.5)$$

$$\nabla W_{k(ji)} = W_{x(ji)} + \eta fitness_j net_i \quad (5.6)$$

$$\nabla \beta_{k(ji)} = \beta_{x(ji)} + \eta fitness_j net_i \quad (5.7)$$

12. Compute crude oil price projec by ***Neuro-Genetic model***
13. Compute fitness
14. Compute *ac*
15. Minimize fitness by adjusting W, β, and *hn* to maximize *ac*
16. Evaluate the fitness of each individual in the population
17. Computation of W, β, and *hn* is performed repeatedly in each generation
18. ***Stop*** *Evolution*
19. Return the optimal chromosome, which translate to ***Neuro-Genetic model*** with the optimal W, β, *hn*
20. Remove the oldest dataset and add a new dataset distorted by the Gulf War and remodel
21. Repeat step 20 for the new dataset distorted by Asia Financial Crises
22. Repeat step 21 for the new dataset distorted by the Iraq War and Venezuela Unrest
23. Repeat step 22 for the new dataset distorted by Global Financial Crises
24. In each step 19, 20, 21, 22, 23 do step 25
25. Compute *fitness*, *ac,* and record *CPU processing time*
26. Stop *remodelling*
27. ***End.***

Figure 5.6: Pseudo-code of the Neuro-Genetic

For the purpose of evaluation, the backpropagation ANN, and SVM based modelling approaches were also implemented. The NN and SVM were modelled to project crude oil price using the state of the art approach (i.e. without retraining).

5.9 Summary

We investigated the performance of ANNs model based on normalized data and raw data. The effects of normalized and raw data on the training dataset, validation dataset, testing dataset, the number of iterations and computation time on the ANNs model were experimentally investigated. The performances of the two methods (normalization and raw) with respect to data standardization were compared. GA was used to select input attributes to serve as inputs to ANN. Three benchmarks, namely principal component analysis, correlation, and trial-and -error were used to select input attributes, whereas, Haar wavelet transformation was used to decompose the crude oil price and applied to develop a BPNN model; we observed the accuracy and convergence speed of each method. This chapter presents an algorithm and experimental procedure for an alternative AI approach for the projection of crude oil price while considering the impact of uncertainties at different times in order to meet the practical application of crude oil price projection.

CHAPTER SIX

RESULTS AND DISCUSSION

6.1 Introduction

This chapter discusses the results obtained from the experimental description presented in chapter five (5). The experiments were for the comparative performance between raw data and normalized data. The selection of input attributes (refer to Tables 2.3 and 6.7) using GA and the state of the art methods of attribute selection, and developer of Neuro-Genetic model for the projection of crude oil price while considering the impact of uncertainties on the crude oil price. In addition, the chapter is divided into two major parts: preliminary results that contain experimental trials for the purpose of choosing the best option on data preprocessing and selection of input attributes. The second part contains the final results of the crude oil price projection using Neuro-Genetic model based on the preliminary results while considering the impact of uncertainties on crude oil price. The second part is further divided into training and testing results.

6.2 Preliminary Results

6.2.1 Comparison between Raw and Normalization Data Standardizations Methods

This section presents the results that were obtained from the experiments performed for comparing raw and normalized data. The experiments were not designed to gain statistically valid quantitative results or to evaluate the performance of the ANN. This experiment is not hypothesis-testing with the aim of generalizing to the general population; instead, it is to learn and construct an initial working hypothesis for addressing normalized and raw data in the domain of crude oil price projection. The differences in the two

methods were determined by ANOVA, and the effects of training, validation, computation time and iterations on the performance of the ANN were analyzed using the Multiple Linear Regression Model (MLRM). Different data partitioning ratios were to build the ANN model for projecting crude oil price. Table 6.1 reports the complete computation results obtained from the two compared methods. The computation time (Nanosec on HP L1750 model, 4Gb RAM, 232.4 GB HDD, 32- bit OS, Intel (R) Core (TM)2 Duo CPU @ 3.00 GHz) and the number of iterations required in each experiment to converge are also reported.

Table 6.1: Experimental results for the normalized and raw data methods

Exp.	TR-V-T Partition	NM Training (MSE)	NM Validation (MSE)	NM Test (MSE)	RM Training (MSE)	RM Validation (MSE)	RM Test (MSE)	NM Iteration	NM CT (Nanosec)	RM Iteration	RM CT (Nanosec)
1	80-10-10	0.01	0.01	0.03	31.42	57.33	114.62	23	6.97E+08	22	6.67E+08
2	70-15-15	0.01	0.04	0.02	36.4	105.38	128.14	17	5.15E+08	20	6.06E+08
3	50-20-30	0.01	0.03	0.03	142.44	199.46	216.95	13	3.94E+08	12	3.64E+08
4	60-20-20	0.01	0.03	0.06	54.56	71.09	86.47	23	6.97E+08	19	5.76E+08
5	65-15-20	0.02	0.03	0.02	89.41	81.42	90.99	23	6.97E+08	15	4.55E+08
6	90-05-05	0.01	0.002	0.01	140.62	122.43	260.38	17	5.15E+08	10	3.03E+08
7	40-30-30	0.02	0.17	0.03	289.17	434.14	443.78	18	5.45E+08	19	5.76E+08
8	80-5-15	0.01	0.01	0.02	69.66	60.4	103.42	20	6.06E+08	25	7.58E+08
9	70 -10-20	0.02	0.01	0.24	174.61	168.44	210.28	19	5.76E+08	10	3.03E+08
10	50-15-35	0.12	0.03	0.04	24.64	72.33	111.68	18	5.45E+08	28	8.48E+08
11	60-15-25	0.22	0.01	0.02	75.54	56.07	95.63	17	5.15E+08	18	5.45E+08
12	85-5-10	0.2	0.02	0.03	73.11	20.43	59.33	13	3.94E+08	14	4.24E+08
13	75 -5-10	0.02	0.03	0.02	50.88	109.12	97.27	17	5.15E+08	13	3.94E+08
14	45-20-35	0.02	0.02	0.29	165.54	132.11	161.98	11	3.33E+08	10	3.03E+08
15	60-10-30	0.11	0.03	0.03	113.09	94.05	100.84	20	6.06E+08	12	3.64E+08
16	55-15-30	0.06	0.04	0.05	52.62	77.83	105.56	8	2.42E+08	13	3.94E+08

6.2.2 Results of the ANOVA for Normalized Vs. Raw data

The data presented in Table 6.1 were analyzed using ANOVA, and the results are reported in Table 6.2.

Table 6.2: ANOVA results (F and P-value) at a 95% confidence interval for mean MSE

Training				Validation				Testing			
Mean MSE of NM	Mean MSE of RM	F	Sig.	Mean MSE of NM	Mean MSE of RM	F	Sig.	Mean MSE of NM	Mean MSE of RM	F	Sig.
0.05	98.98	32.28	0.000	0.03	116.38	23.59	0.000	0.06	149.21	38.7	0.000

The output of the analysis indicates a statistical significant difference [F (df = 1, 30, P < 0.05) = 32.28]; hence, the first null hypothesis which stated that "*there is no significant difference between the projection accuracy of the normalized method and the projection accuracy of the raw method on the training dataset"* is rejected, where the projection accuracy of the normalized method is significantly better than that for the raw method on the training dataset (mean MSE: NM =0.05; RM = 98.98). Furthermore, a statistically significant difference for validation was also found between the normalized method and the raw method [F (df = 1, 30, P < 0.05) = 23.59]; thus, the second hypothesis which stated that "*There is no significant difference between the projection accuracy of the normalized method and the projection accuracy of the raw method on the validation dataset"* is rejected. This finding indicates that the projection accuracy of the normalized method is significantly better than the raw method on the validation dataset (mean MSE: NM =.03; RM = 116.38). Moreover, a statistically significant difference between normalized method and raw method was also found during testing [F (df = 1, 30, P < 0.05) = 38.69], whereby the projection accuracy of the normalized method is significantly better than that of the raw method on the test dataset (mean MSE: NM =.06; RM = 149.21).

In summary, the performance accuracy of the normalized method in terms of MSE is significantly better than that of the raw method on both in-sample and out-sample projections of crude oil price. The accuracy of the results is most likely because the normalized data are in approximately equal proportion to each other in the datasets, which makes the model for handling the normalized data behave more efficiently. Moreover, computation in the neurons was efficiently performed without overflow. Unlike the normalized data, the raw data have some attributes that are thousands of times larger than others, for example, the world crude oil consumption is in millions of barrels, whereas the highest value of the attribute Brent crude oil price is $123.26; the values of these two attributes are not in approximately equal proportion. The poor performance observed on the raw method is likely due to the sigmoid activation function used in the hidden layer of the ANNs projection model, which is built through the raw data becoming saturated with large values since the sigmoid activation function has the possibility of becoming saturated when input vectors are approximately greater than (exp (−3) 0.05) as pointed out in (Beale *et al*., 2013). Another probable cause could be the numeral overflow in the neurons; the occurrence of this phenomenon can lead to degradation in the model projection accuracy because some vital values might be missing during computation in the neurons. Data normalization is of great advantage to the crude oil price projection problem, which contradicts the results obtained by (Shanker, 1996), who stated that data normalization has no effect on the ANN performance. It was discussed in (Beale *et al*., 2013) that the ANN is trained for generalization within a specific problem domain; as such, it does not have the capability to accurately extrapolate into another domain. This limitation could most likely

be the reason for the contradiction because the work of (Shanker, 1996) was a classification problem on American telephone and telegraph company data.

It was found in (Jin, 2005) that a model might introduce false optima, while it performs very well on the training dataset. Let us assume here that a significant performance is observed only on the test dataset, whereas training and validation indicate otherwise, or that a significant performance is observed on the training data set, whereas the validation and test dataset showed no significance. In addition, we assume that the significance is observed only on the validation dataset. Then, conclusions cannot be drawn from the results due to a lack of consistency in the performance accuracy because more than one factor must be considered and the most significant is the accuracy on both the training dataset and the test dataset, as argued in (Jin, 2005).

If the performance on the training and test dataset indicated significance and showed otherwise on the validation dataset, then it can be concluded that the accuracy is consistent because some users might not even validate and still realize accurate performance with their model (Hernandez *et al*., 2013). Once more, if a significant performance is observed on the validation and test dataset and not on the training set or a significance performance is observed on the training and validation sets but not on the test dataset, then the results can be considered to be inconsistent in their performance because one of the major factors required for the performance accuracy of the model is not significant. The accuracy of the normalized method is consistent, as established in our research findings.

6.2.3 Regression Analysis

The Regression (R) values throughout the experiments were also recorded for both the NM and RM data standardization methods which are reported in Table 6.3. The R approach

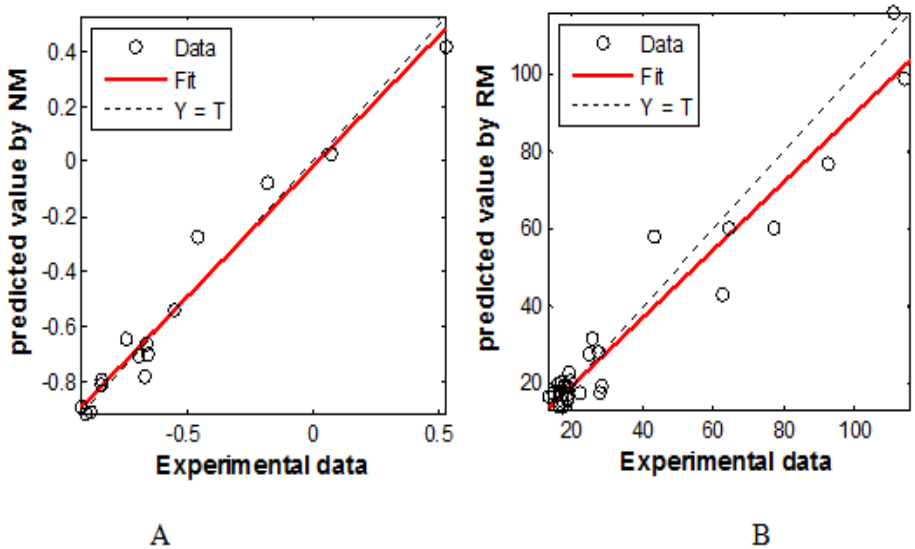
reveals the fit between the experimental data and the network output, which shows the directional movement of the two sets of data.

Table 6.3: Regression results

Exp.	TR-V-T Partition	NM TR	NM V	NM T	RM TR	RM V	RM T
1	80-10-10	0.96977	0.93364	0.96320	0.98176	0.95730	0.93318
2	70-15-15	0.96929	0.96979	0.95211	0.91404	0.91339	0.85849
3	50-20-30	0.97688	0.88203	0.95034	0.98324	0.96601	0.90345
4	60-20-20	0.96471	0.94741	0.94530	0.94716	0.94210	0.95719
5	65-15-20	0.96670	0.99606	**0.98168**	0.93868	0.91307	0.94079
6	90-05-05	0.95841	0.96840	0.93917	0.80515	0.68475	0.71999
7	40-30-30	0.97245	0.98144	0.96062	0.95633	0.97752	0.93855
8	80-05-15	0.96203	0.97970	0.57862	0.97606	0.94311	0.95757
9	70 -10-20	0.97400	0.88660	0.90233	0.98480	0.94656	0.93756
10	50-15-35	0.97720	0.96779	0.94108	0.94446	0.96237	0.96151
11	60-15-25	0.95307	0.97652	0.93720	0.95998	0.98695	**0.96786**
12	85-05-10	0.96602	0.93408	0.91862	0.96830	0.95737	0.94629
13	75 -05-20	0.96148	0.96722	0.92131	0.95214	0.91396	0.90638
14	45-20-35	0.96159	0.93574	0.90517	0.88985	0.92085	0.90923
15	60-10-30	0.97208	0.95414	0.92112	0.93992	0.94048	0.92657
16	55-15-30	0.89005	0.89500	0.87988	0.98098	0.93647	0.90384

Training-Validation-Testing (TR-V-T), Normalized Method Training (NM TR), Normalized Method Validation (NM V), Normalized Method Testing (NM T), Raw Method Training (RM T), Raw Method Validation (RM V), Raw Method Testing (RM T).

The whole regression plot obtained during the simulations cannot be presented in the thesis because they are very large in number; as such, the best fit plots for each of the methods are shown in Figure 6.1. The optimal R on the test dataset for the normalized method, which occurred in experiment 5, is depicted in Figure 6.1A to show the fit between the projected and experimental Brent crude oil price datasets. The best value of R for the raw method was obtained in experiment 11, and the fit on the test dataset is shown in Figure 6.1B.



A B

Figure 6.1: (A) Regression plot of Normalized data on test dataset, (B) Regression plot for raw data on test dataset

In the scatter plot in Figure 6.1, a few data points are observed to exhibit poor fits because they are relatively far away from the best linear fit line.

Table 6.4: ANOVA results (F and P-value) for regression at a 95% confidence interval for the mean R

Training				Validation				Testing			
Mean R of NM	Mean R of RM	F	Sig.	Mean R of NM	Mean R of RM	F	Sig.	Mean R of NM	Mean R of RM	F	Sig.
0.9622	0.9452	1.863	0.182	0.9485	0.9289	1.028	0.319	0.9124	0.9168	0.026	0.874

The experimental data in Table 6.3 were analyzed, and the results are presented in Table 6.4, which signifies that there is no significant difference [F (df = 1, 30, P < 0.05) = 1.863]; hence, the R fit of the normalized method is (with respect to this significance test) equal to that of the raw method on the training dataset (mean R: NM = 0.9622; RM = 0.9452). In addition, no statistically significant difference in the validation dataset was found between the two methods [F (dF = 1, 30, P < 0.05) = 1.028], which indicates that the R of the

normalized method and the raw data method on the validation dataset are statistically equal (mean R: NM = 0.9485; RM = 0.9289). Additionally, no statistically significant difference was established between the R value of the normalized method and the raw method [F (df = 1, 30, $P < 0.05$) = 0.026], where the R of the normalized method and raw method are statistically equal on the test dataset (mean R: NM = 0.9124; RM = 0.9168).

It has been established in this research that the R values for both of the NM and RM data standardization methods that are compared are statistically equal to the training, validation and test dataset. The result is surprising considering their respective performance in terms of the MSE; as a result, it might be expected that the normalized method has a significantly better value of R due to its performance in terms of the MSE. However, the results are in accordance with the statement of NeuroSoluitions (2012), which notes that the MSE determines how well the neural network output matches the target values, but does not necessarily indicate the directional movement of the two sets of data. The MSE can be changed without the direction of the dataset being changed, which implies that the MSE can be very good while R can have a poor value. Conversely, MSE can be poor, whereas R can have a very good value, as occurs in the case of the raw method. Both the MSE and the R value can be very good, as established in the normalized method. Additionally, there is the possibility of having both the values of MSE and R performing poorly. Therefore, it can be deduced that the findings indicating that the R value of RM and NM are equal whereas MSE of NM is significantly better than the MSE of RM corroborate with the literature (NeuroSoluitions, 2012).

Table 6.5: ANOVA results (F and P-value) for the number of iterations and the computation time at a 95% confidence interval for the mean number of iterations and the computation time

Number of Iterations				Computation Time (Nanosec)			
Mean for NM	Mean for RM	F	Sig	Mean for NM	Mean for RM	F	Sig
17.31	16.25	0.37	0.55	4.92E+08	5.25E+08	0.37	0.55

The fourth hypothesis is supported by the output of the ANOVA analysis, as presented in Table 5.7, which indicates that there is no statistical significant difference for the [F (df = 1, 30, P > 0.05) = 0.37]. In other words, there is no significant difference between the number of iterations of the normalized method and the number of iterations of the raw method required to converge to the optimal solution. Similarly, the fifth hypothesis states "*There is no significant difference between* the *computation time of* the *normalized method and* the *raw method that is required to converge to* the *optimal solution",* which is supported by the output of the one-way ANOVA analysis, which shows that there is no statistical significant difference [F (df = 1, 30, P > 0.05) = 0.37]. The results signify that the computation time for the two methods is statistically equal and that the differences observed are merely random. It was expected that the computation time and the number of iterations of the normalized method could be significantly lower than those of the raw method because the values are normalized to a smaller range, which can easily be computed by the algorithm, whereas some of the raw data values are very large and are expected to take more computation time. In contrast, the results proved otherwise, signifying that large numbers have no significant effects on the ANNs network computation time and the number of iterations required to obtain an optimal value of the MSE and R.

6.2.4 Selection of Input Attributes using Genetic Algorithm

In this section, 314 architectures of the ANN were evolved through the genetically evolving GA selection of input attributes (GEGRNNGA) operation in 7862 seconds. The ten (10) best solutions that emerged based on fitness values with their corresponding parameters are reported in Table 6.6. The minimum and maximum chromosome training passes for each chromosome were 20 and 50 respectively. The best solution was found in generation 1 after a runtime of 8s (see Table 6.6). The architecture of the ANN used in the experiment has 5 input neurons, 148 hidden layer neurons, and one output neuron (5 – 148 – 1) as shown in Table 6.6 rank 1. The hidden layer used a summation transfer function. The output neuron used a direct transfer function. Population size was 50, crossover probability was 0.6, and mutation probability was 0.001. One point crossover function and a uniform mutation function were used. The following input attributes were selected by the GA as the optimum combinations: WCOP, OECDES, OECDCOC, OPECCP and NOPECCP.

Table 6.6: The ten best ten genetically evolved ANNs with the input attribute selection by GA

Rank	Fitness	NIFSGA	HLNs	TMSE	TrMSE	VMSE	CS(s)	GF
1	**0.000043**	**5**	**148**	**0.000142**	**0.00033**	**0.00041**	**08**	**1**
2	0.0169	3	148	0.0189	0.0098	0.00576	28	1
3	0.0183	4	148	0.0217	0.0048	0.00649	1322	5
4	0.0183	3	148	0.0227	0.0091	0.00672	954	4
5	0.0204	2	12	0.0227	0.011	0.0562	6392	97
6	0.0206	2	148	0.0212	0.0181	0.0743	230	1
7	0.0228	2	148	0.0264	0.0082	0.0867	249	1
8	0.0232	3	148	0.0261	0.012	0.0067	185	1
9	0.03666	4	67	0.0428	0.0454	0.0451	1417	20
10	0.0306	6	23	0.00411	0.043	0.0011	7693	10

Number of input attributes selected by GA (NIFSGA), Hidden layer neurons (HLNs), Test MSE (TMSE), Training MSE (TrMSE), Convergence speed (CS), Generation found (GF)

From Table 6.6 it was observed that the GA does not just select minimum inputs without considering their performance and influence on the Brent crude oil price. As indicated in column 3, there are chromosomes with a smaller number of inputs, such as two which is less than the optimal five inputs selected by the GA, yet their performance is inferior to the optimal solution.

6.2.5 BPNN Parameters

The parameters of the BPNN obtained after the experimental trials are: 1 hidden layer, the sigmoid activation function in hidden and linear activation in the output layer. Momentum of 0.5, learning rate of 0.07, Levenberg–Marquardt learning algorithm, the number of hidden layers is 1. The number of input neurons depends on the input selection technique used. For the KLT, seven (7) input neurons, for r seven (7) input neurons, for TE nine (9) input neurons and HWT five input neurons (5) were used.

6.2.6 Correlation

The correlation coefficient relationship between each of the input features and the Brent crude oil price was found and the results are reported in Table 6.7. It was found that seven input attributes had a positive significant relationship with the Brent crude oil price. The results in Table 6.7 rejected three attributes with negative relationship and accepted seven attributes with positive relationship. Those seven attributes are the inputs to the rBPNN.

Table 6.7: The correlation coefficient relationship between the input attributes and the Brent crude oil price

Attributes	Brent crude oil price
OECDES	0.478**
OECDCOC	0.371**
USCOP	-0.614**
OPECCP	0.469**
USCOS	-0.407**
USCOSR	-0.444**
USCOI	0.525**
NOPECCP	0.706**
WCOP	0.764**
USESTG	0.405**

** Correlation is significant at the 0.01 level (2-tailed)

Table 6.7 showed that the attributes with negative correlation coefficient cannot be used to project the Brent crude oil price because the negative attributes have no influence on the directional movement of the Brent crude oil price. Thus, the negative attributes cannot determine the fluctuation of the Brent crude oil price. On the other hand, the attributes that form positive relationship indicates that they can influence the fluctuation of the Brent crude oil price in view of the fact that they move in the same direction. Therefore, they can be used as input attributes to ANN for the projection of the Brent crude oil price since any fluctuation of the attributes also influence the fluctuation of the Brent crude oil price and vice versa (refer to section 3.5 for explanation of correlation coefficient).

6.2.7 Principal Component Analysis

Figure 6.2 presents the results of transforming the 10 input attributes (identified in section 2.5.2) using the KLT. It indicates that seven (7) (KLT are WCOP, OPECCP, OECDES, OECDCOC, NOPECCP, USCOP and USESTG) of the 10 components accounted for 96% variance as shown in Figure 6.2. These were extracted and the other components discarded.

The seven components were used as inputs to build the KLTBPNN model for the projection of Brent crude oil price.

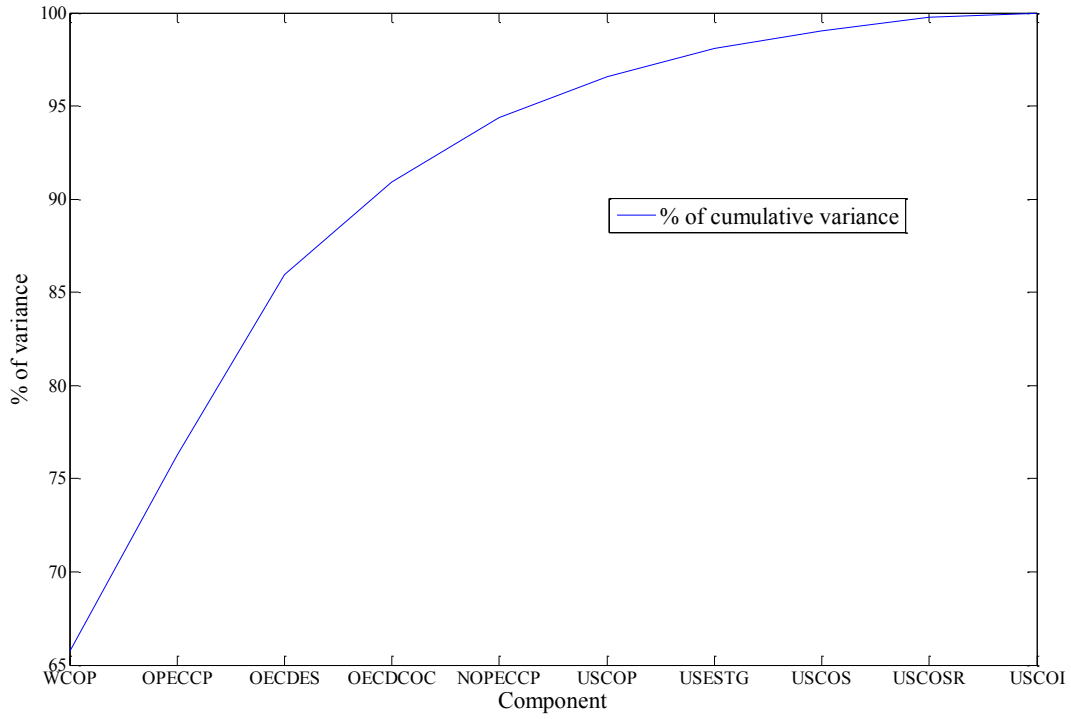


Figure 6.2: KLT of the dataset with variance of each component

6.2.8 Trial and Error

After several trials, the TE method was able to select nine (9) input attributes: OECDES, OECDCOC, USCOP, USCOS, WCOP, USCOSR, USESTG, OPECCP and NOPECCP. The attributes were used as inputs to the TEBPNN and this was trained to develop a model for projecting Brent crude oil price.

6.2.9 Haar Wavelet Transform

The decomposition of the Brent crude oil price data signal is illustrated in Figure 6.3. There were five (5) HWT decomposition levels, as clearly indicated in the figure. The

crude oil price time series data are decomposed into five (5) levels as $d_{(j)}$, where j =1, 2, 3, 4 and 5, i.e. the five (5) decomposed time series data based on the HWT coefficient at several scales. The a_5 is the decomposed crude oil time series data based on the scale coefficients. The monthly Brent crude oil price (both original and extracted portions) in dollars per barrel is represented on the horizontal axis, whereas the vertical axis represents the amplitude of scaling coefficient (hertz).

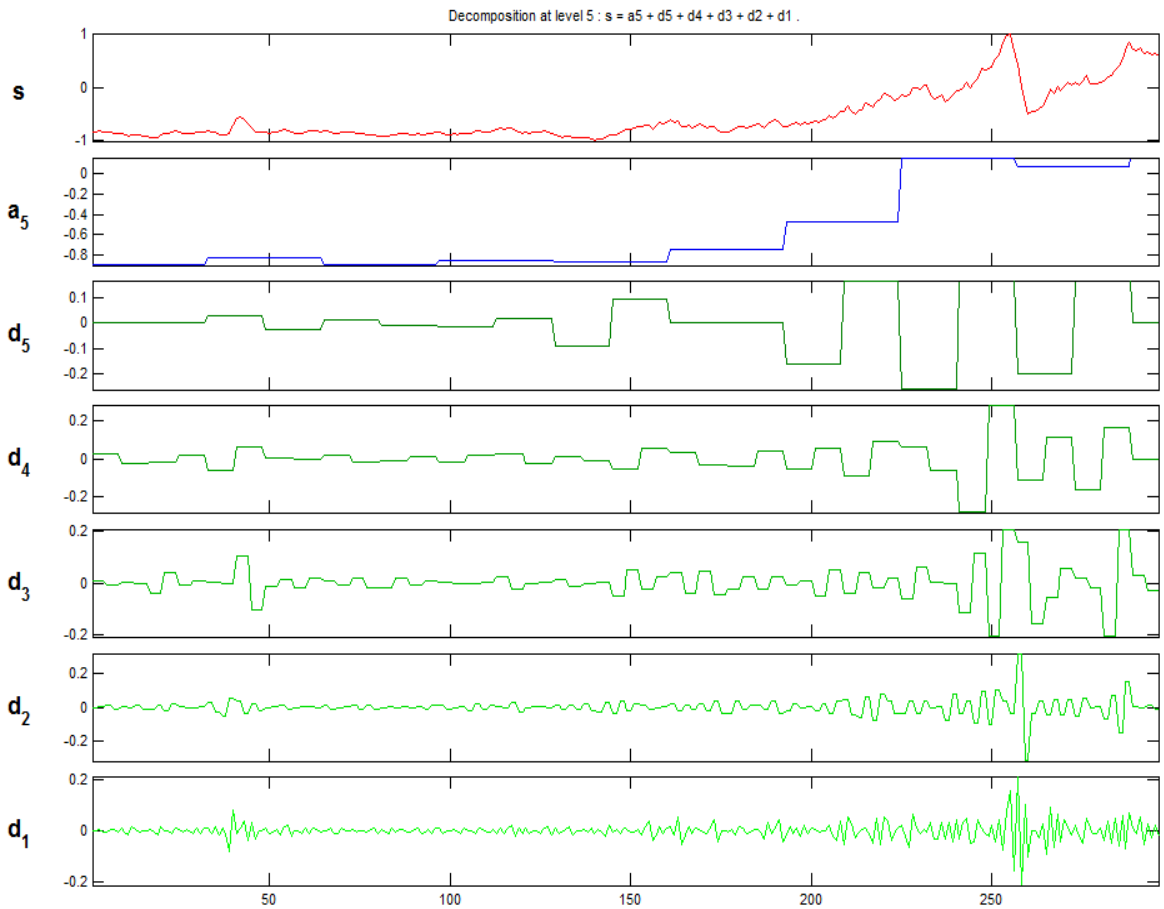


Figure 6.3: HWT scale coefficients of Brent crude oil price time series data

6.2.10 Comparing Performance of the GA with the State of the Art Methods

The performance accuracy of projecting the Brent crude oil price for each method was recorded while considering convergence speed. The MSE of every method in training, validation and testing datasets were observed. The MSE results are plotted in Figures 6.4 (training), 6.5 (validation), 6.6 (test), and 6.7 (convergence speed). A total of sixty (60) experiments was conducted in which twelve (12) were performed with each of the methods. The lines plotted in the Figures 6.5, 6.6, and 6.7 represent the errors between observed and projected Brent crude oil price. In Figure 6.4, the horizontal axis present the data partition ratio and the MSE is on the vertical axis. Each data partition ratio corresponds to the MSE for each of the models (rBPNN, KLTBPNN, TEBPNN, GEGRNNGA, and HWTBPNN) represented by lines.

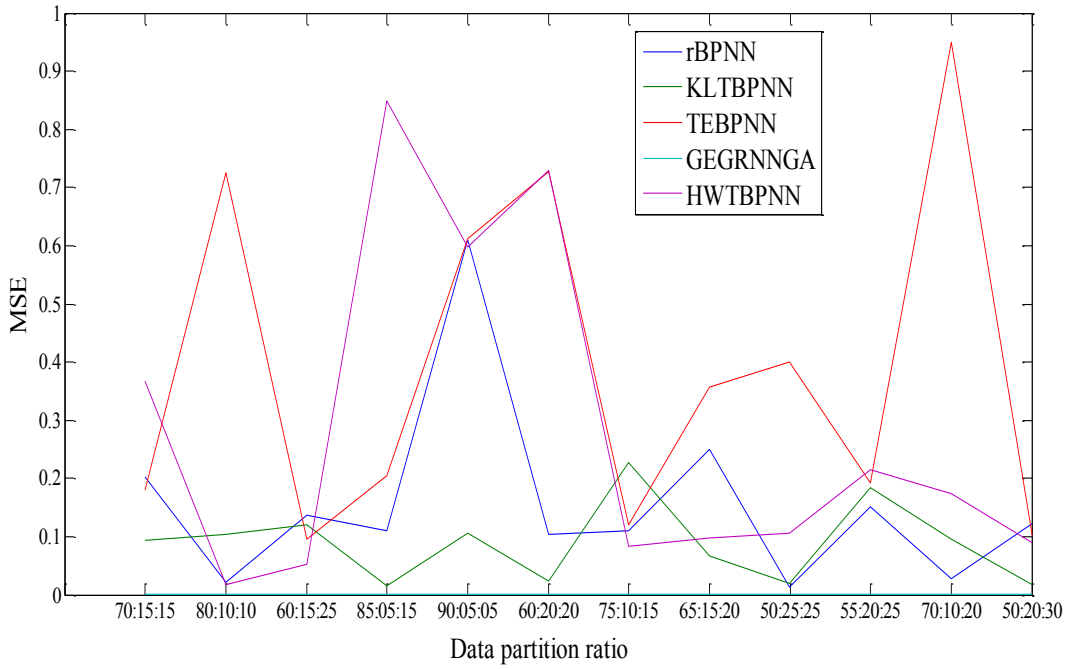


Figure 6.4: Comparing training MSE of the models

Figure 6.4 shows the MSE of the methods after training; it clearly indicates that the partitioning of training data has an impact on the results. The pattern which the graphs follow, points out the way in which the results are affected by the training data. From Figure 6.4, it is seen that the proposed GA has the lowest MSEs of all the methods for all data partitioning ratio; the TEBPNN displays the worst rate of MSEs. However, the performance displayed by an algorithm in the training phase does not necessarily guarantee its performance in the validation and test phases; it can be very accurate in the training phase but poor in validation or testing as pointed by (Ji, 2005).

To really determine the performance of the GA, a statistical test with confidence interval is required to conduct further analysis, to indicate whether the performance is statistically significant. The MSE and convergence speed of the five methods were analyzed using ANOVA and the results of the post hoc test are presented in Tables 6.8 – 6.11. The values

in Tables 6.8 – 6.11are the mean difference of the ANOVA Tukey post hoc analysis. The ANOVA Tukey post hoc test finds the differences among MSE means of the models. This is typically performed if the models are more than two and significant differences exist. The difference means score are compared to a *p-value* to find out if the difference is significant.

Table 6.8: ANOVA post hoc results at 95% confidence interval for comparing training MSE performance among the models

	KLTBPNN	TEBPNN	GEGRNNGA	HWTBPNN
rBPNN	-0.01383	-0.37893	0.00944*	-0.08213
KLTBPNN		-0.36509	0.02328*	-0.06830
TEBPNN			0.38837*	0.29679
GEGRNNGA				0.09158*

* The mean difference is significant at the 0.05 level.

From Table 6.8, it is clear that the performance of GEGRNNGA in the training phase of the modeling is significantly better than that of the other methods.

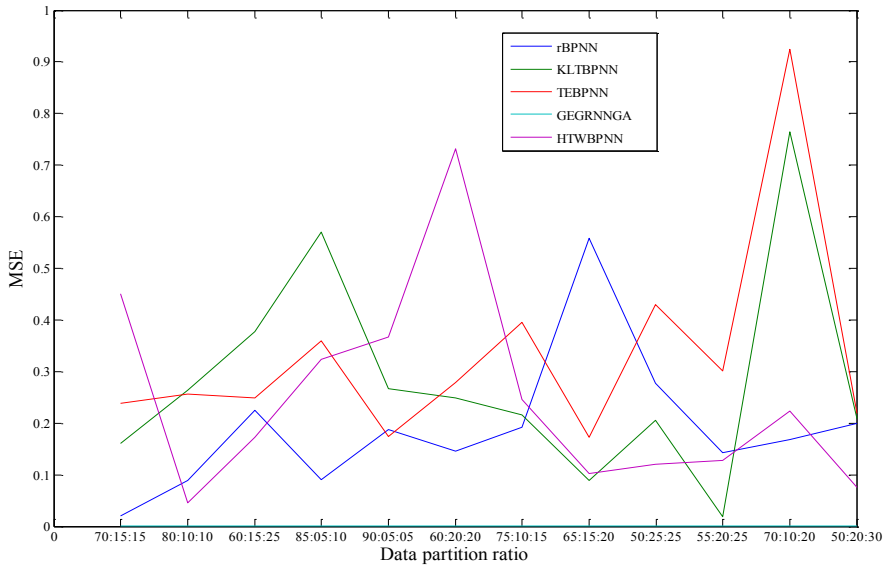


Figure 6.5: Validation MSE of the comparable models

It can be seen from the plots of Figures 6.5 and 6.6, that the TEBPNN performs poorly in both validation and testing, whereas the proposed GEGRNNGA performs better than the other methods since the pattern of the line produced is at the minimum level of the MSE values.

Table 6.9: ANOVA post hoc results at 95% confidence interval for comparing validation MSE performance among the methods

	rBPNN	KLTBPNN	TEBPNN	GEGRNNGA	HWTBPNN
rBPNN		0.00040	-0.24741	0.01629*	-0.03282
KLTBPNN			-0.24782	0.01588*	-0.03322
TEBPNN				0.26370*	0.21460
GEGRNNGA					0.04910*

* The mean difference is significant at the 0.05 level.

The results in Table 6.9 show that the performance of the proposed GA is significantly better than that of the rBPNN, KLTBPNN, TEBPNN, and HWTBPNN during validation.

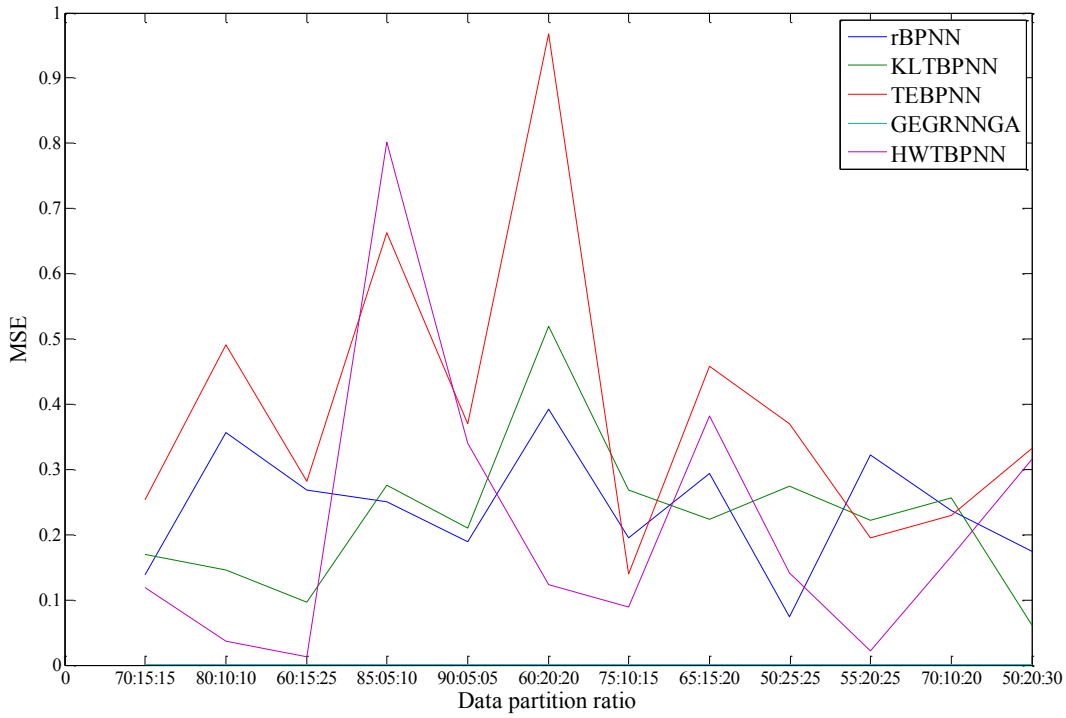


Figure 6.6: Comparison of testing MSE of the models

From Figure 6.7, we can see that the proposed GEGRNNGA has a remarkable convergence speed, much better than that of the four benchmarks (rBPNN, KLTBPNN, TEBPNN, and HTWBPNN) while the TEBPNN has the poorest convergence speed.

Table 6.10 ANOVA post hoc results at 95% confidence interval for comparing test MSE performance among the methods

	rBPNN	KLTBPNN	TEBPNN	GEGRNNGA	HWTBPNN
rBPNN		0.00165	-0.37002	0.02596*	0.00247
KLTBPNN			-0.37167	0.02431*	0.00412
TEBPNN				0.39598*	0.36755
GEGRNNGA					0.02842*

* The mean difference is significant at the 0.05 level.

Table 6.6 indicates that the performance of the proposed GEGRNNGA is significantly better than that of rBPNN, KLTBPNN, TEBPNN and HTWBPNN on the test dataset. It can be summarized that the proposed GEGRNNGA has a superior performance in the projection of Brent crude oil price in both in-sample and out-sample datasets. The most likely cause of this result is the ability of the GA to thoroughly search a very large space and identify the optimal combination of input attributes that are significant and relevant in the projection of Brent crude oil price. The GA was able to offer a minimal but unique subset of inputs compared to the initial 10 input attributes. The redundant and irrelevant inputs were discarded by the GA due to their insignificant contribution to the projection accuracy, therefore improving the performance of the GEGRNNGA. Irrelevant and redundant input features retard the performance of NNs; as such, the projection accuracy of the Brent crude oil price exhibited by the proposed approach as a result of using five out of the original ten input features considered in this research is not misleading. The probable

reason for the poor performance exhibited by TEBPNN is the limitations of the method in searching for the ANNs models.

The lines plotted in Figure 6.7 represent the convergence speed of each of the models in reaching the optimal solution. The figure indicates that the GEGRNNGA has the fastest convergence speed, and TEBPNN the slowest. The poor convergence speed (a measure of the time taken to converge to the optimum solution) of TEBPNN can probably be attributed to the fact that the trial and error technique requires a long time to the destination.

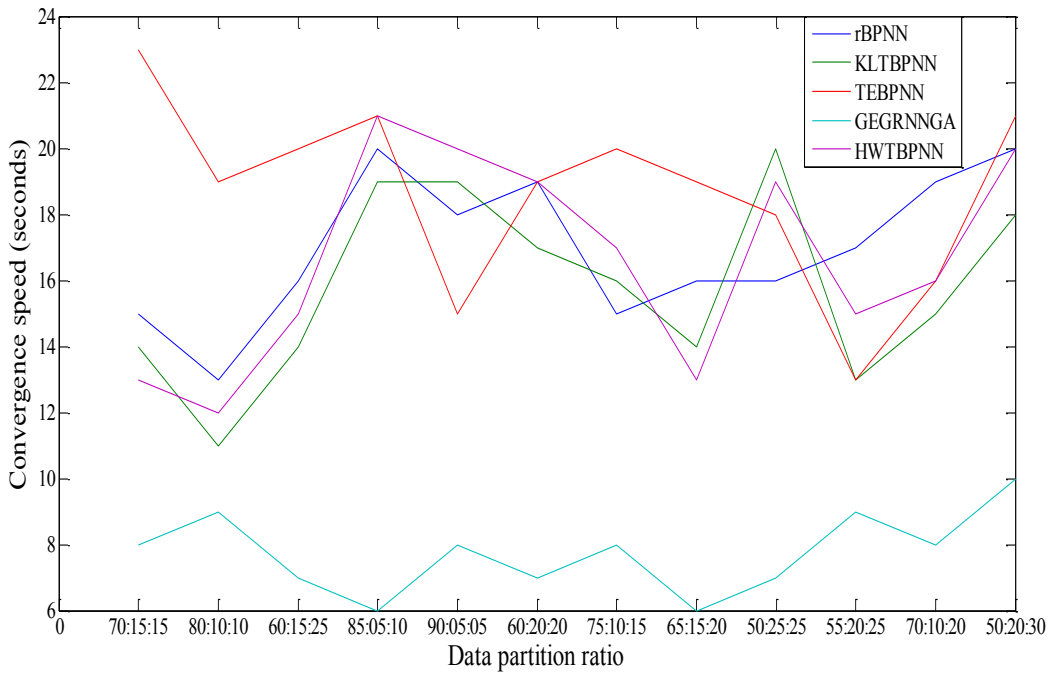


Figure 6.7: Comparing convergence speed of the models

Table 6.11: ANOVA post hoc results at 95% confidence interval for comparing convergence speed among the models

	rBPNN	KLTBPNN	TEBPNN	GEGRNNGA	HWTBPNN
rBPNN		1.16667	-1.66667	10.91667*	0.33333
KLTBPNN			-2.83333	9.75000*	-0.83333
TEBPNN				12.58333*	2.00000
GEGRNNGA					-10.58333*

* The mean difference is significant at the 0.05 level.

Table 6.11 shows that the GEGRNNGA converges significantly faster than the rBPNN, KLTBPNN, TEBPNN, and HWTBPNN, probably because the GA has the capability to automatically search for the best parameters of the NN in the shortest possible time without risk of being trapped in local minima. The GA was able to successfully select five (5) of the most relevant input attributes and it was used for the development of the Neuro-Genetic model.

6.3 Final Results: Neuro-Genetic Model

6.3.1 Training Performance Results

The GARGNNGA was used to select the optimal attributes to serve as inputs to the Neuro-Genetic model using normalized data because in terms of MSE, the normalized data were found to be significantly more accurate than using the raw data in crude oil price projection as discovered in the preliminary findings of this research. The accuracy of the results is most likely because the normalized data are approximately equal in proportion to each other in the dataset, which makes the NN model for handling the normalized data behave more efficiently than the raw method. Moreover, computation in the neurons was efficiently performed without overflow. Unlike the normalized data, the raw data have some attributes that are thousands of times larger than others, for example, the world crude oil consumption is in millions of barrels, whereas the highest value of the attribute Brent

crude oil price is $123.26; hence the values of these two attributes are vastly un-proportional.

The correlation coefficient among input attributes selected by the GA were investigated as required (refer to section 3.6 for justification) and results are reported in Table 6.12. The result shows that the input attributes can be ideal for the projection of crude oil price in view of the fact that better projection requires positive correlation coefficient relationship among the attributes.

Table 6.12: Inter correlation matrix

	OECDES	OECDCOC	OPECCP	NOPECP	Brent
OECDES					0.478
OECDCOC	0.543				0.371
OPECCP	0.408	0.616			0.469
NOPECP	0.410	0.600	0.704		0.706
WCOP	0.640	0.796	0.702	0.870	0.764

The proposed algorithm described in chapter 5, section 5.8 was implemented in MATLAB 2012b on a machine (RAM 2Gb, HDD 305Gb, System Type: windows 7 64 bit OS, Processor: Intel (R) Core (TM) i3-2350M CPU @ 2.30GHz) and the MATLAB source code can be found in Appendix B.

The performances of the training learning curve obtained from the modelling are shown in Figure 6.8 (initial modelling), Figure 6.9 (first retraining), Figure 6.10 (second retraining), Figure 6.11 (third retraining), and Figure 6.12 (fourth retraining) showing performance in each of the modelling phase. The graphs show results for using GA in the development of Neuro-Genetic model for the projection of crude oil price. The straight lines in the graphs indicate convergence to the optimal solution. The learning curves were found to be smooth

without any oscillation as can clearly be seen in the convergence curves. This phenomenon signified successful convergence as pointed out by (Nawi *et al*., 2014).

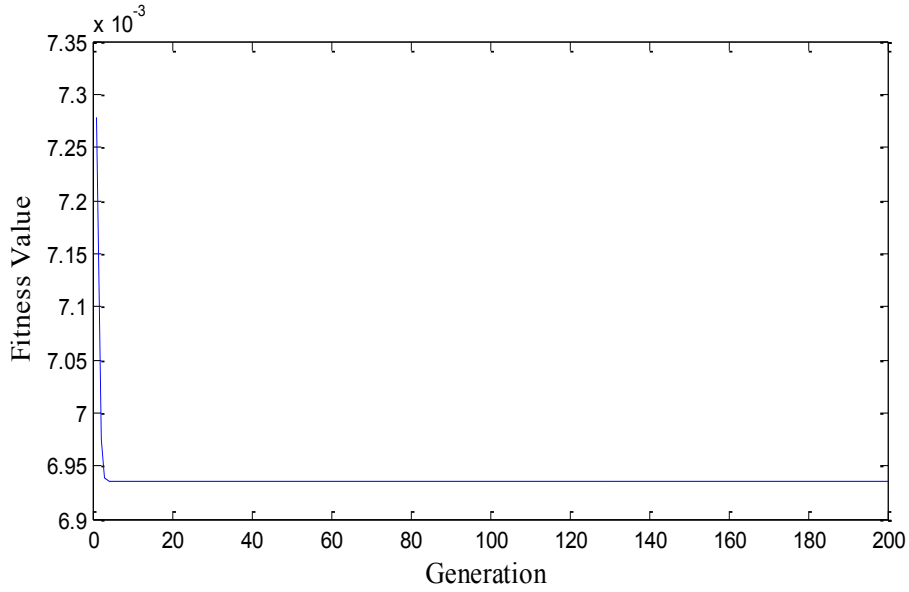


Figure 6.8: Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Initial training prior to uncertainties)

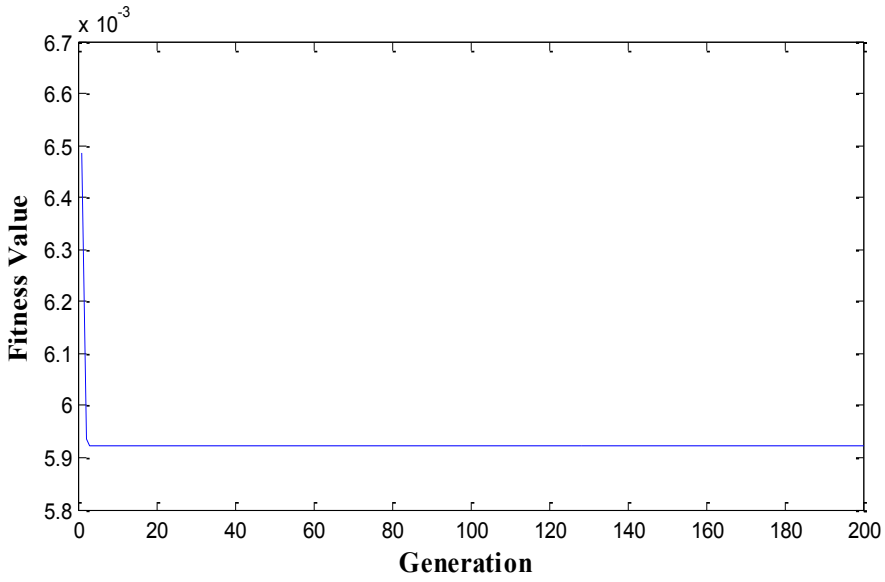


Figure 6.9: Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (First Retraining)

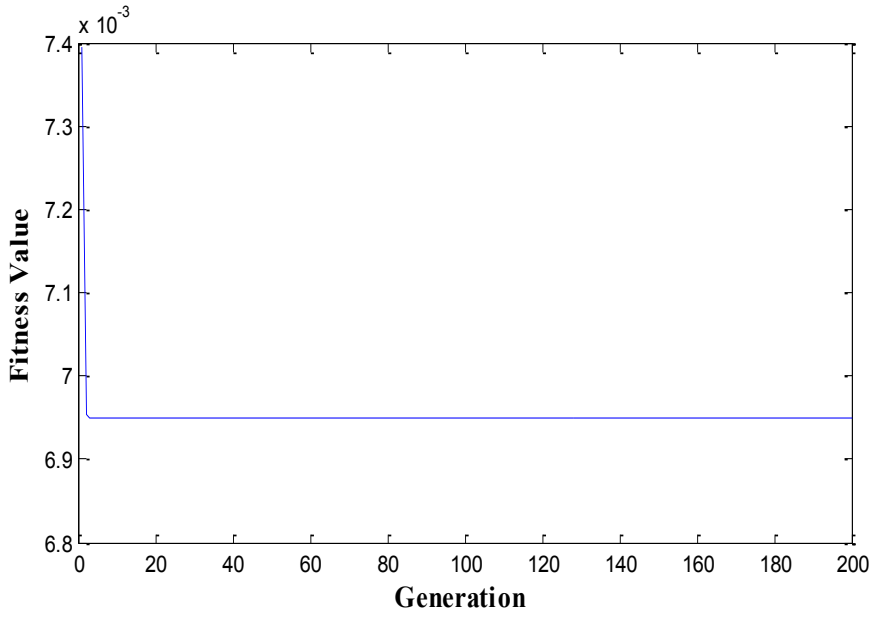


Figure 6.10: Search results for retraining the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Second Retraining)

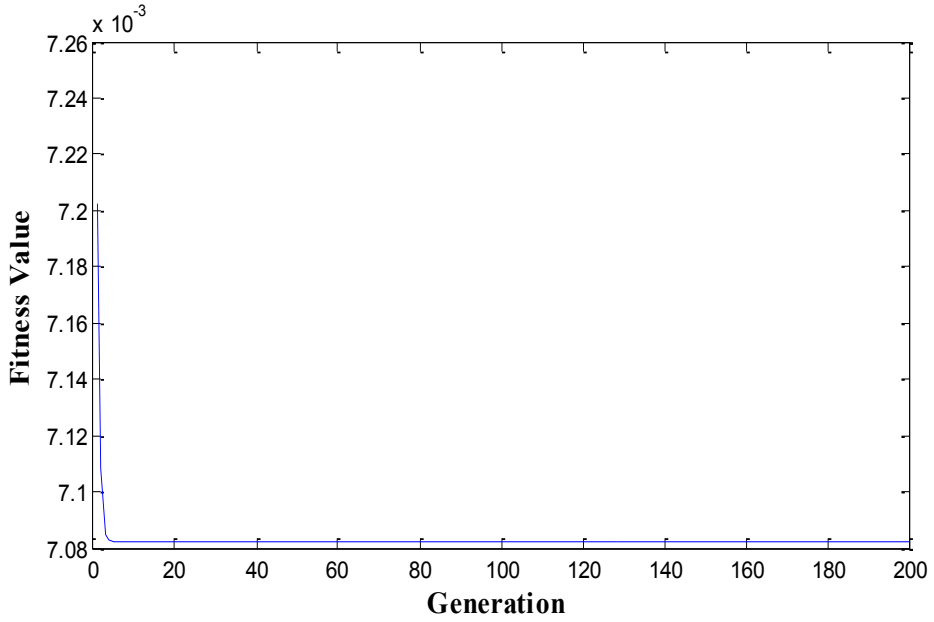


Figure 6.11: Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Third Retraining)

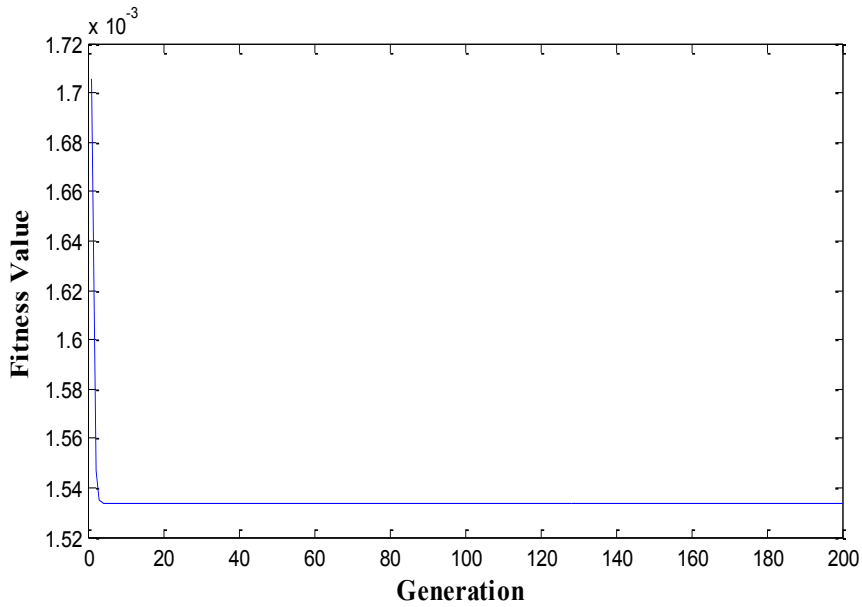


Figure 6.12: Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Fourth Retraining)

6.3.1 Test Performance Results

The following experiments were conducted to evaluate the generalization and the effectiveness of the Neuro-Genetic models developed using the test dataset. The results obtained are depicted in Figure 6.13 (initial training), Figures 6.14 (first retraining), 6.15 (second retraining), 6.16 (third retraining), and 6.17 (fourth retraining) showing both crude oil price projected by the Neuro-Genetic model and the actual price.

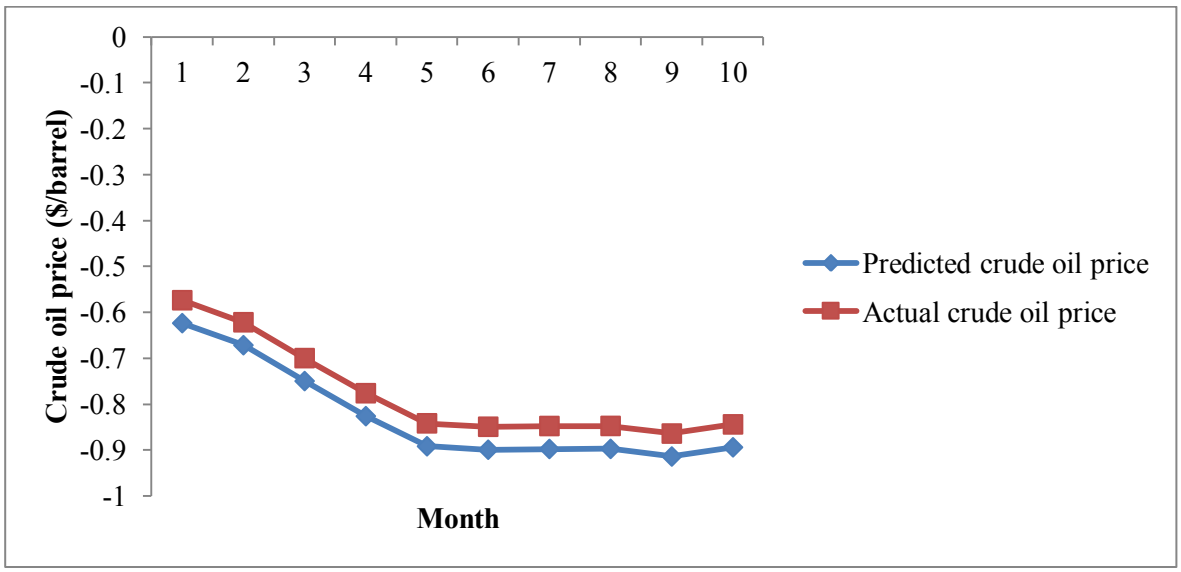


Figure 6.13: Comparing crude oil price projected by Neuro-Genetic model with an actual crude oil price (Initial training prior to uncertainties)

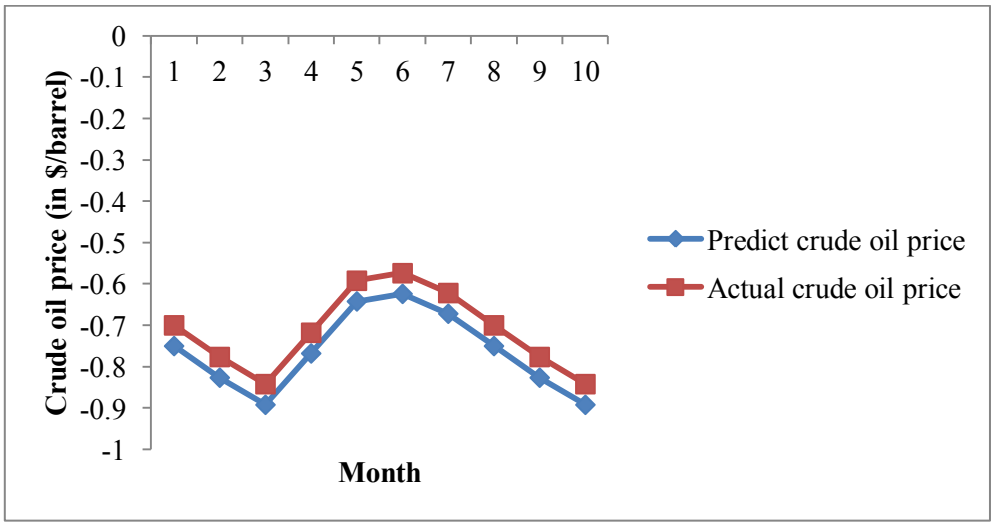


Figure 6.14: Comparison of crude oil price projected by Neuro-Genetic Model with an actual crude oil price (First Retraining)

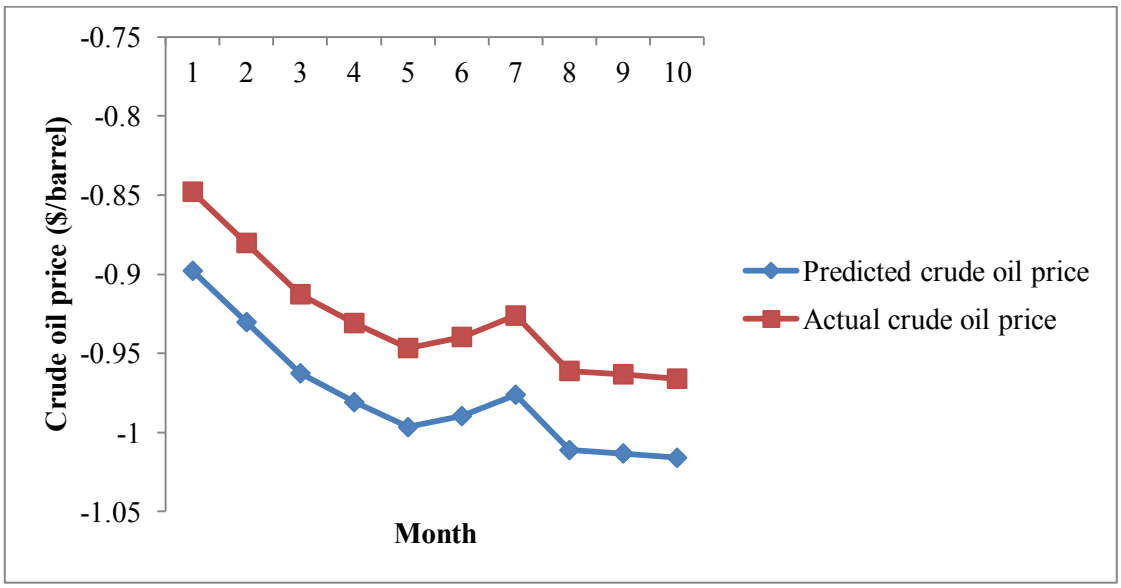


Figure 6.15: Comparing crude oil price projected by Neuro-Genetic model with actual an crude oil price (Second Retraining)

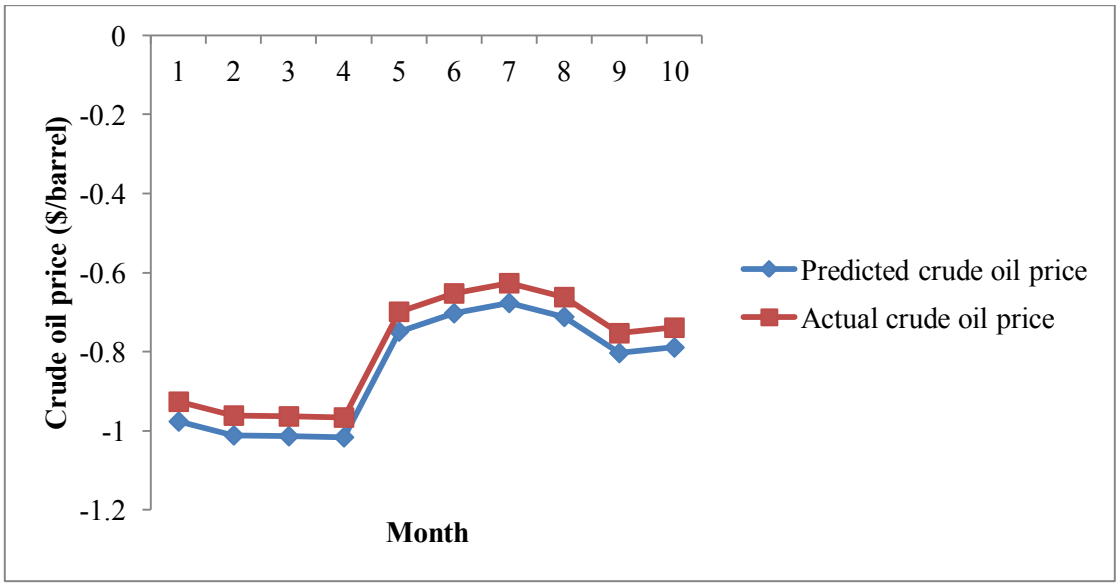


Figure 6.16: Comparing crude oil price projected by Neuro-Genetic model with an actual crude oil price (Third Retraining)

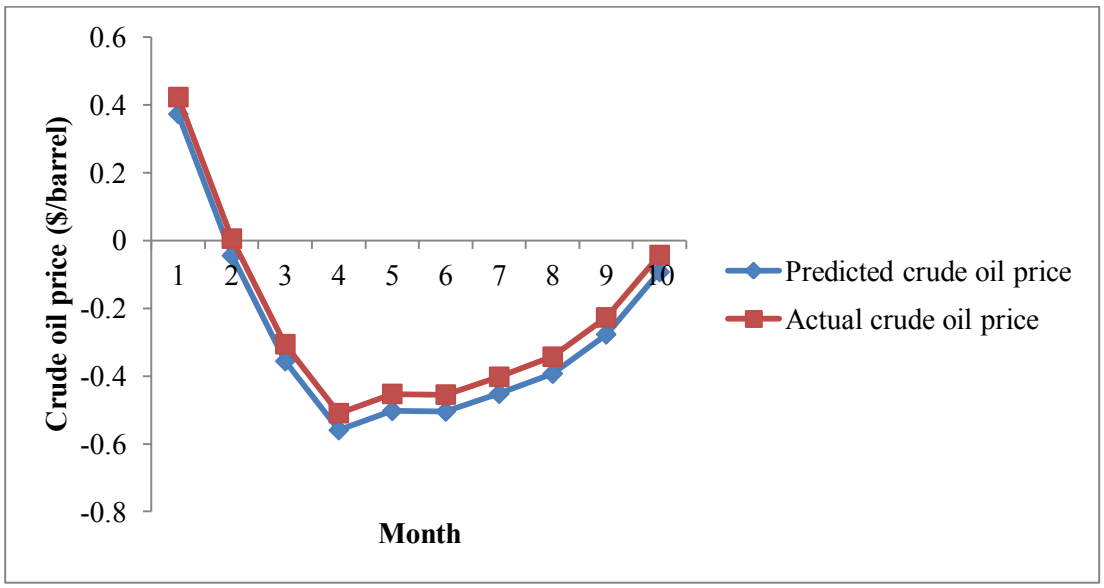


Figure 6.17: Comparing crude oil price projected by Neuro-Genetic model with an actual crude oil price (Fourth Retraining)

It is obvious from Figures 6.13 – 6.17 that the Neuro-Genetic model was able to capture the patterns at different periods of time. This means that whenever there is an uncertainty which distort crude oil price as a result of its impact, the model can effectively be retrained to capture the new pattern which occurs due to the impact of the uncertainty, and project the price of crude oil with high level of accuracy. In the future, we expect uncertainties related to crude oil to occur at different period of time with different magnitude of impact as earlier stated, thus, as the crude oil price are distorting the model can continue to be retrained with an additional new set of data which might have been distorted by the impact of uncertainties. This could allow the Neuro-Genetic model to continue to learn new patterns as new information flow into the oil market.

It can be observed in Figures 6.13 – 6.17 that there is a polarization resulting in a constant shift of projected values with respect to the actual ones. This is because a constant window of the dataset was maintained throughout the retraining process. When new datasets were

added, the oldest dataset were removed to maintain the constant window of the dataset. For example, when the new dataset from August 1990 to February 1991 (7 months), which was distorted during the first Gulf War of 1991, was added, the dataset of the oldest 7 months were removed from the dataset to maintain a constant shift of the projected crude oil price with respect to the actual crude oil price. Maintaining the same size of dataset avoids bias based on the volume of the dataset in view of the fact that the size of dataset affects the robustness and performance of ANN. Replacing the initial dataset with a completely new dataset distorted by uncertain events will reduce the training data, which can negatively affect the performance of the Neuro-Genetic model, and a relatively large amount of sample data are required to build a robust Neuro-Genetic model. So, maintaining constant window of data enables the dataset to maintain a constant volume for the different uncertainties.

The projection accuracy of the Neuro-Genetic model at various stages of the retraining are reported in Table 6.13. It is indicated from the performance evaluation in Figures 6.13 – 6.17 that the Neuro-Genetic model was able to learn the new changes in the dataset which occur due to the impact of uncertainties, capture the pattern despite being distorted and project the crude oil price.

Table 6.13: Performances of the Neuro-Genetic model compared with those of the state of the art approach based on SVM, and ANN

Models	ac (%)	fitness	CPUPT(sec.)
NGM: Initial training	**99.65**	**0.006935**	**9.70**
NGM: First Retraining	**99.70**	**0.005922**	**10.94**
NGM: Second Retraining	**99.52**	**0.009567**	**9.00**
NGM: Third Retraining	**99.63**	**0.007319**	**10.22**
NGM: Fourth Retraining	**99.92**	**0.001534**	**8.63**
NN: Initial training	56.57	0.068511	15.34
NN: First Retraining	57.33	0.053212	16.11
NN: Second Retraining	57.49	0.050121	15.76
NN: Third Retraining	56.88	0.062331	16.91
NN: Fourth Retraining	57.04	0.059121	15.21
SVM: Initial training	56.42	0.071581	20.11
SVM: First Retraining	56.91	0.061744	21.34
SVM: Second Retraining	56.14	0.077219	19.56
SVM: Third Retraining	55.99	0.080135	22.34
SVM: Fourth Retraining	57.05	0.058828	20.87
Models without retraining			
NGM	**99.81**	**0.002424**	**11.23**
ANN	45.20	0.096000	16.42
SVM	43.50	0.130000	21.76

Neuro-Genetic Model (NGM), CPU processing time (CPUPT), percentage accuracy (ac)

The results in Table 6.13 are for comparing the performance of the Neuro-Genetic model with those of SVM, and backpropagation ANN for training the crude oil price projection. The comparison results (Table 6.13) indicated that the proposed methodology presented in the research was found to be better than the ANN and SVM in terms of all the performance metrics (ac, fitness and cpu processing time). The Neuro-Genetic outperforms the ANN and SVM in both retraining and without retraining of the dataset as shown in Table 6.13. Subsequently, The ANN performs better than the SVM in retraining and without retraining of the dataset. It can be seen that the minimum *fitness* result to maximum *ac* and vice versa (refer to section 4.6 for details on *fitness* and *ac*). The effectiveness of the Neuro-Genetic model in capturing the changes which occur in the data can best be attributed to the retraining procedure which probably allow the model to capture new information that occur as a result of changes in the data. The retraining and leaving out of the oldest subset of the dataset has the advantage of preserving memory and computational resources over

the approach of continuing adding data without removing the old subset dataset which could consume a significant amount of computing resources in the long run.

6.4 Statistical Test

The statistical analysis is implemented in SPSS version 20 on the machine earlier described. To ensure that the accuracy achieve by the Neuro-Genetic model is not just a chance effect. The research used a *t-test* to measure the significance of the Neuro-Genetic model. We have two independent samples from two populations, thus, the *t- test*, 2-sample *t-test* was employed for testing the difference of the two means (project crude oil price and actual crude oil price). The *t-test* is performed under the hypothesis that the project crude oil price and the actual crude oil price are equally formulated as follows:

***H_{o7}*:** There is no significance difference between the crude oil price as projected by the Neuro-Genetic model and the actual crude oil price.

Table 6. 14: The *t-test* results of the Neuro-Genetic model

Neuro-Genetic Modelling	mpcop	maacop	md	t	*p*	df
Initial training	-0.8266	-0.7766	-0.05	-1.045	0.310	18
First Retraining	-0.764	-0.714	-0.05	-1.157	0.262	18
Second Retraining	-0.977	-0.927	-0.05	-2.918	0.009	18
Third Retraining	-0.845	-0.795	-0.05	-0.784	0.443	18
Fourth Retraining	-0.281	-0.231	-0.05	-0.389	0.702	18

Mean difference (md), mean project crude oil price (mpcop), mean actual crude oil price (macop), degree of freedom (df)

The result is based on *p-values* which is typically 0.05 at 95% confidence interval. The t-test results shown in Table 6.14 indicated that the t-test is not significant ($p>0.05$). Thus, H_{07} is accepted and concludes there is no significance difference in the means between the crude oil price projected by the Neuro-Genetic model and the real price. This means that

the crude oil price projected by the proposed model is statistically equivalent to the actual price. Thus, the model has the potential for representing the real world system because it was able to produce results that are statistically the same as the actual real world system. As a result, the model can be used as an alternative for the projection of crude oil price and consequently project the future behavior of the crude oil market. The Neuro-Genetic model which incorporates uncertainties is highly needed in today's era of high price of crude oil, since some regions that significantly produce crude oil are prompt to uncertainties such as the Middle East region, Africa etc. Consequently, the proposed model if applied for the projection of crude oil price, can provide more accurate estimation than the models that are not retrained to capture the impact of uncertainties because a significantly new information that probably might have distorted the price may be missing out, if not retrained. The statistical validation of our results is required for real life practical applications as argued by Demšar (2006). Retraining as changes occur in data, as practiced in this research, has real world potential practical applications. We can conclude that the model can probably meet realistic, practical needs of the application of crude oil price projection in the real world as compared to the approaches discussed in the literature by researchers (Khashman and Nwulu, 2011b; Movagharnejad *et al*., 2011; Yu *et al*., 2008c; Mehdi, 2009; Shouyang *et al*., 2005).

The performance of the Neuro-Genetic model is better than the ANN and SVM as shown in Table 6.13, hence, to determine whether the difference between the performance of the Neuro-Genetic and that of ANN and SVM significant, we employed the ANOVA to explore the difference among the three compared methods as the *t-test* is restricted to only two samples. The following hypothesis were formulated for the test using ANOVA.

H_{08}: There are no significant differences in the projection accuracies of the three (3) methods.

H_{09}: There are no significant differences in the CPU processing time of the three (3) methods.

Table 6.15 ANOVA Result for comparing the performance accuracy of Neuro-Genetic model, ANN, and SVM

	Sum of Squares	df	Mean Square	F	P
Between Groups	0.053	2	0.027	168.866	0.000
Within Groups	0.002	12	0.000		

From Table 6.15, the value of F (df = 5, 174, $p<0.05$) = 43.32 is significant. The ***H_{08}*** hypothesis is rejected. Thus, the ANOVA test shows that there are significant differences among the performance accuracy of the three (3) compared methods. Therefore, the model that causes the significant difference has to be identified by performing Tukey post hoc (refer to section 6.2.10 details) multiple comparison test and the results are shown in Table 6.16. This indicated that there is a significance difference among the comparative methods, thus, the Tukey post hoc is required to identify the method that is significantly better than the other methods.

Table 6.16: Tukey post hoc multiple comparison test for Neuro-Genetic, ANN, and SVM

(I)	(J)	Mean Difference (I-J)	*p*	95% Confidence Interval Lower Bound	Upper Bound
Neuro-Genetic	ANN	-0.110536*	0.000	-0.13176	-0.08931
	SVM	-0.138136*	0.000	-0.15936	-0.11691
ANN	Neuro-Genetic	0.110536*	0.000	0.08931	0.13176
	SVM	-0.027600*	0.012	-0.04882	-0.00638
SVM	Neuro-Genetic	0.138136*	0.000	0.11691	0.15936
	ANN	0.027600*	0.012	0.00638	0.04882

* The mean difference is significant at the 0.05 level.

The Tukey post hoc multiple comparison test results indicated that significant differences occur between Neuro-Genetic model and the other two (2) models (ANN and SVM) as shown in Table 6.16. This Means that the accuracy achieved by the Neuro-Genetic model is significantly better than that of the ANN and SVM.

Table 6.17: ANOVA test results for comparing CPU processing time of Neuro-Genetic, ANN, and SVM

	Sum of Squares	df	Mean Square	F	*p*
Between Groups	275.406	2	137.703	46.704	0.000
Within Groups	35.381	12	2.948		

Table 6.17 shows that the ANOVA test for the differences among CPU processing time for three of the models (NG, ANN and SVM) is significant (F (df = 5, 174, $p<0.05$) = 43.32). The H_{09} hypothesis is rejected. Thus, the ANOVA test shows that there are significant differences among the CPU processing time of the three (3) compared methods (Neuro-Genetic, ANN, and SVM). Therefore, further analysis of post hoc multiple comparison test was performed to identify the model that causes the significant difference and the results is presented in Table 6.18.

Table 6.18: Tukey post hoc multiple comparison test of CPU processing time for the Neuro-Genetic model, ANN, and SVM

I	J	Mean Difference (I-J)	*p*	95% Confidence Interval	
				Lower Bound	Upper Bound
Neuro-Genetic	ANN	-6.696340*	0.000	-9.59362	-3.79906
	SVM	-10.347540*	0.000	-13.24482	-7.45026
ANN	Neuro-Genetic	6.696340*	0.000	3.79906	9.59362
	SVM	-3.651200*	0.014	-6.54848	-0.75392
SVM	Neuro-Genetic	10.347540*	0.000	7.45026	13.24482
	ANN	3.651200*	0.014	0.75392	6.54848

*The mean difference is significant at the 0.05 level.

The Tukey post hoc multiple comparison test results in Table 6.18 shows that significant differences occur between Neuro-Genetic model and the other two (2) models (ANN and SVM). This Means that the CPU processing time of the Neuro-Genetic model is significantly better than that of ANN and SVM. In summary, the projection accuracy of the Neuro-Genetic model proposed in this study significantly outperforms the ANN and SVM in terms of accuracy and CPU processing time. It was painted by Arciniegas and Rueda (2008) that a small improvement in projection accuracy could yield a significant increase in generating revenue. The Neuro-Genetic model is faster in converging to the optimal solution and this could reduce the time required to build an accurate projection model. Time constraints are significant factors in investment, governmental planning, critical managerial decision making and risk management, thus, small delays may cause millions of dollars to be lost. The convergence speed exhibited by the Neuro-Genetic model would therefore be valuable in the business of making profit and strategic planning. Our proposed model makes it possible for timely decisions to reach informative conclusions with a high degree of accuracy in critical business issues.

6.5 Summary

The results of the preliminary and the final experiments of the research are presented. The preliminary experiments include: data standardization, selection of input attributes using GA and other state of the art methods such as ANN, correlation coefficient, HWT as well as trial and error. The inter correlation coefficient relationship between the attributes including Brent crude oil price was found to be positive. The final experiment involved the development of the Neuro-Genetic model for the projection of crude oil price while considering the impact of uncertainties.

CHAPTER SEVEN

CONCLUSIONS AND FUTURE WORK

7.1 Introduction

The main purpose of this research is to investigate the use of AI techniques for the projection of crude oil price while considering the impact of uncertainties at different times in order to meet the practical application of crude oil price projection. Six (6) uncertainties that have proven to significantly affect crude oil price were selected as case studies for the research. Historical data were collected from EIAUSDE. Experiments were conducted to ascertain the appropriate data standardization method and to select the optimal input attributes using GA for building the Neuro-Genetic model in order to achieve the following objectives as stated in section 1.4:

iv. To develop a Neuro – Genetic Model in order to project the price of crude oil.

v. To project crude oil price while considering the impact of unexpected events using a Neuro – Genetic model.

vi. To compare the performance of the Neuro – Genetic model with support vector machine and neural network.

Section 7.4 shows how the stated objectives were achieved.

The answer to the following research questions as stated in section 1.5 are provided in this section:

viii. Which data standardization method is preferable in the domain of crude oil price projection?

ix. Can GA perform better than the statistical methods, trial and error, and manual methods in selecting input attributes?

x. Can the proposed model capture the impact of unexpected events on crude oil price?

xi. Is the crude oil price project by Neuro-Genetic and the actual crude oil price equal?

xii. Can the Neuro – Genetic model performs better than the comparative methods?

xiii. Is there a significant difference between the projection accuracy of the Neuro – Genetic model and the comparative methods?

7.2 Data Standardization

We reported an empirical study on the performance of an ANNs model based on normalized data and raw data. The effects of normalized and raw data on the training dataset, validation dataset, testing dataset, the number of iterations and computation time on the ANNs model were investigated in a preliminary experiment. Significant conclusions are summarized as follows:

i. Data normalization is of great benefit to researchers in the domain of crude oil price projection. The accuracy of the ANN was significantly better using normalized data in both the in-sample and out-sample performances.

ii. The regression values of both the normalized and raw methods are statistically equal. This implies that the fit (directional movement) of the original and the projected data for normalized data as well as for the raw data is the same.

iii. The computation time and number of training iterations required to build the ANN model are statistically equal for both of the data standardization methods.

This means that the convergence speed required by the normalized and raw data to converge to the optimal solution are statistically the same (refer to Table 6.1 for values).

7.3 Selection of Inputs Attributes by Genetic Algorithm

Input attributes were selected in order to discard irrelevant attributes so as to improve the accuracy and convergence speed. The effectiveness of GA in the selection of input attributes was investigated. Four state of the art methods, namely, KLT, r, TE and HWT, were used to select input attributes and develop a BPNN model; we observed the projection accuracy and convergence speed. The proposed GEGRNNGA was found to significantly outperform the benchmark models in both the projection accuracy and convergence speed.

7.4 Neuro-Genetic Model

An AI model (Neuro-Genetic) was developed to project crude oil price at different period while considering the impact of uncertainties. Statistical t-test results show that the crude oil price projected by the Neuro-Genetic model and the actual price were found to be statistically equal. The performance of Neuro-Genetic model was compared with that of ANN and SVM models based on the state of the art method. ANOVA results indicated that the Neuro-Genetic model is significantly better than the back-propagation ANN and SVM in both accuracy and CPU processing time. The Neuro-Genetic model was able to learn the new changes in the data which occur due to impact of uncertainties, capture the pattern despite being distorted and project the crude oil price.

The Neuro-Genetic model was able to learn patterns from crude oil price datasets that were distorted by the impact of uncertainties such as the Gulf War, Asia Financial Crises, Iraq War, Venezuela Unrest and Global Financial Crises. The impact of the uncertainties was captured through retraining by using the distorted dataset at different times of the events. The occurrence of uncertainties in future are expected, evaluating the impact of uncertainty is extremely difficult as earlier stated, and even if a similar uncertain event occurs again in the future, the impact of this uncertain event on crude oil price would not be the same. Thus, the Neuro-Genetic model proposed in this research provides an approach to crude oil price projection, in view of the performance exhibited by the Neuro-Genetic model and the ability of the model capture the impact of uncertainties at different period of time. This could provide a more realistic, practical application of the crude oil price projection. The Neuro-Genetic model can be used as an advisory tool by policy makers in the formulation of food subsidy policies since accurate projections of crude oil price could contribute to a better policy formulation which can reduce the devastating effect of increase in food price, hence suppress the hardship of inadequate food supply in a country. A country like Brazil that focuses attention on alternative source of energy like the USA, could find the model suitable for use as an advisory mechanism when formulating subsidy policies on wheat, corn, sugar, etc. The increase in the crude oil price or decrease can lead to fluctuations in the stock market. Thus, the Neuro-Genetic model can be employed by investors in the creation of effective stock market risk management frameworks and strategies that can absorb a shock from crude oil price volatility, which may likely exert negative impact on the stock market. Governments of several countries across the globe can use the Neuro-Genetic model as an alternative or complementary tool for the estimation of national

budgets and planning. Government revenue generation, national budget, and planning of many countries (refer to section 2.5.1) across the world heavily depend on the expected price of crude oil. That is why, crude oil price has to be effectively monitored. Thus, the price of crude oil is very critical for these countries because any significant negative fluctuation can affect the country's national planning and budget and in some cases result in budget deficits. If there is an accurate projection of the crude oil price behavior, adequate measures can be put in place to absorb the negative impact. A significant percentage of the national budget proposal of some countries (refer to section 2.5.1 for examples of the countries) is based on the international crude oil price benchmark; such countries could find this Neuro-Genetic model for the projection of international crude oil price very helpful in their national planning and budget for effective monitoring of fluctuations in the crude oil market.

7.5 Future Work

In the future, we intend to build a crude oil price projection model that can make projections for different time horizons (daily, weekly, monthly, and yearly) and which is not limited to a particular time horizon. The model will also comprise of other crude oil price benchmarks (WTI and Dubai crude oil price), in order to improve its generalization application. The limitations of our proposed model are as follows: The model projection is restricted to monthly crude oil price. Thus, it cannot project daily, weekly and yearly crude oil price. Also, the model is restricted to Brent crude oil price benchmark, thereby it may not be beneficial to WTI and Dubai crude oil price benchmarks. The genetic algorithm used in the present study requires many parameter settings such as mutation, crossover, population size, generations, etc. As such, in the future study, we intend to apply the

cuckoo search algorithm that requires the settings of only two parameters which include probability of worse nest to be abandoned (P_a) and the step size (α).

7.6 Contributions of the study

7.6.1 Contributions to the Machine Learning

This research indicates that it is possible to build a model based on the hybridization of ANN and GA through retraining in order to capture the impact of uncertainties and project the price of crude oil. This research may trigger other researchers to propose new methods based on our methodology for the projection of crude oil price while considering the impact of uncertain events.

The major contributions to the machine learning environment:

i. It was found that the impact of uncertainties such as Gulf war, world financial recession, Asian financial crisis, the Iraq war and Venezuela unrest on crude oil price can be modelled by Neuro-Genetic through retraining in order to project crude oil price while incorporating the impact of the uncertainties.

ii. The crude oil price projected by Neuro-Genetic was statistically equal to the actual crude oil price. The Neuro-Genetic model has the potentials for representing the real world system because it was able to produce results that are the same as the actual real world system.

iii. The GA significantly performs better than the correlation coefficient, PCA, wavelet transform and trial-and-error in selecting WCOP, OECDES, OECDCOC, OPECCP and NOPECCP as the optimal inputs attributes among the ten (10) attributes available for used.

iv. Comparison of the Neuro-Genetic model with backprogation ANN and SVM shows that the Neuro-Genetic model was significantly better than the ANN and SVM in both accuracy and convergence speed.

v. The comparions conducted between raw and normalized data shows that the normalized data is significantly more accurate than the raw data, whereas convergence speed was found to be statistically equal.

vi. The comparison between ANN and SVM shows ANN is more accurate and faster than the SVM in the projection of crude oil price in retraining and without retraining.

vii. The Neuro-Genetic model methodology documented in this research is an alternative to the methods already in existence. Therefore, the method in the research has added to the number of methods in the literature which can further be modified by future researchers for application in other domains.

7.6.2 Contributions to the Energy Economy

i. The Neuro-Genetic model was able to improve the projection accuracy of crude oil price which can be used by investors to design a risk management framework that can reduce investment risks and increase profits.

ii. The Neuro-Genetic model is on a small experimental scale and investigation, but, it can be scaled up into a model such that, the output could be used for pricing purposes that serving as an advisory tool to investors and government for proper planning.

iii. The Neuro-Genetic model could potentially be used to advise governments in oil producing countries on the formulation of international policy related to crude oil price as well as food subsidy, stock market, risk management framework, budget planning, and creating a framework for reducing the negative effect of crude oil price volatility.

iv. Accurate projection of crude oil price is required for energy demand and supply projection, which could possibly create stability in the oil market. In this manner, we could improve economic activities and reduce or eliminate suffering typically caused by volatile crude oil price in communities.

References

Abbas, Y.A., & Aqel, M.M. (2003). Pattern recognition using multilayer neural-genetic algorithm. Neurocomputing, 51, 237–247.

Abdel-Aal, R.E. (2008). Univarate modeling and forecasting of monthly energy demand time using abductive and neural networks. Computers & Industrial Engineering, 54, 903 – 917.

Abdi, J., Moshiri, B., Abdulhai B., & Khaki A.S. (2012). Short-term traffic flow forecasting: parametric and nonparametric approaches via emotional temporal difference learning. Neural Computing & Applications, DOI 10.1007/s00521-012-0977-3.

Abdullah, S.N., & Zeng, X. (2010). Machine learning approach for crude oil price prediction with artificial neural networks-Quantitative (ANN-Q) model. Proceedings of the International Joint Conference on Neural Networks Art. No. 5596602.

Abramovich, F., Bailey, T. C., & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. Journal of the Royal Statistical Society: Series D (The Statistician), 49(1), 1-29.

Abramson, B., & Finizza, A. (1991). Using belief networks to forecast oil prices. International Journal of Forecasting, 7(3), 299–315.

Abrishami, H., & Varahrami, V. (2012). Survey of a rule based expert system for gas price forecasting. World Applied Sciences Journal, 16 (4), 493 – 500.

Acampora, G., Lee, C., Vitiello, A., & Wang, M. (2012). Evaluating cardiac health through semantic soft computing techniques. Soft Computing, 16, 1165 – 1181.

Ahmadi, M. A., & Shadizadeh, S. R. (2013). Intelligent approach for prediction of minimum miscible pressure by evolving genetic algorithm and neural network. Neural Computing & Applications, 1-8.

Ai-ping, J., & Feng-wen, H. (2009). Methods for optimizing weights of wavelet neural network based on adaptive annealing genetic algorithm. In: Industrial Engineering and Engineering Management, 2009. IE&EM'09. 16th International Conference on (pp. 1744-1748). IEEE.

Alexandre, E., Cuadra, L., Rosa, M., & López-Ferreras, F. (2007). Feature selection for sound classification in hearing aids through restricted search driven by genetic algorithms. Audio, Speech, and Language Processing, IEEE Transactions on, 15(8), 2249-2256.

Alexandridis, A., & Livanis, E. (2008). Forecasting crude oil prices using wavelet neural networks. University of Macidonia web. http://users.uom.gr/~aalex/files/Download/FSDET%202008%20.pdf. Accessed 10 July, 2012.

Alizadeh, A., & Mafinezhad, K. (2010). Monthly brent oil price forecasting using artificial neural networks and a crisis index. Proceedings of International Conference on Electronics and Information Engineering, 465 – 468.

Ahmadi, M. A., & Shadizadeh, S. R. (2013). Intelligent approach for prediction of minimum miscible pressure by evolving genetic algorithm and neural network. Neural Computing & Applications, 1-8.

Ammar, H.H., & Tao, Y. (2000). Fingerprint registration using genetic algorithms, Application-Specific Systems and Software Engineering Technology. Proceedings. 3rd IEEE Symposium on, pp. 148 – 154.

Amin, A. E. (2013). A novel classification model for cotton yarn quality based on trained neural network using genetic algorithm. Knowledge-Based Systems, 39, 124-132.

Anand, R., Mehrotra, K., Mohan, C. K., & Ranka, S. (1995). Efficient classification for multiclass problems using modular neural networks. Neural Networks, IEEE Transactions on, 6(1), 117-124.

Assareh, E., Behrang, M.A., Assari, M.R., & Ghanbarzadeh, A. (2010). Application of particle swarm optimization and genetic algorithms techniques on demand estimation of oil in Iran. Energy, 35, 5223– 5229.

Awad, M. (2010). Input Variable Selection using parallel processing of RBF Neural Networks. The International Arab Journal of Information Technology, 7(1), 6–13.

Azad, H. B., Mekhilef, S., & Ganapathy, V. G. (2014). Long-Term Wind Speed Forecasting and General Pattern Recognition Using Neural Networks. *IEEE* Transactions on Sustainable Energy, 5(2), 546-553.

Azadeh, A., Moghaddam, M., Khakzad, M., & Ebraimipour, V. (2012). A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting. Computers & Industrial Engineering, 62, 421 – 430.

Azar, T.Z. (2012). Fast neural network learning algorithms for medical applications. Neural Computing & Applications. DOI 10.1007/s00521-012-1026-y.

Bahrammirzaee, A.A. (2010). comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. Neural Computing & Applications, 19, 1165–1195.

Balasundaram, S., & Tanveer, M. (2013). On Lagrangian twin support vector regression. Neural Computing and Applications, 22(1), 257-267.

Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. In: Proceedings of an International Conference on Genetic Algorithms and their applications (pp. 101-111).

Balas, C.E., Koc, L., & Tür, R. (2010). A rtificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters. Applied Ocean Research, 32, 425–33.

Bao, Y., Zhang, X., Yu, L., Keung, K.L., & Wang, S. (2007). Hybridizing wavelet and least squares support vector machines for crude oil price forecasting. School of management and economics of UESTC web. http://www.mgmt.uestc.edu.cn/prc/papers/IWIF-II%20Wang%20Shouyang%202.pdf. Accessed 12 June, 2012.

Barone-Adesi, G., Bourgoin, F., & Giannopoulos, K. (1998). Don't look back. Risk 11, 100–103.

Bartschi, M. W. (1996). A genetic algorithms for resource – constrained scheduling. PhD thesis submitted to department of mechanical engineering, Massachusetts Institute of Technology.

Barton, A.J. (2009). Learning the neuron functions within neural networks based on genetic programming. Msc thesis submitted to faculty of graduate studies and research, Carleton University.

Basheer, J.A., & Hajeer, M. (2000). Artificial neuron.: fundamentals, computing, design and application. Journal of microbiology methods, 43, 3 – 31.

Beale, M.H., Hagan, M.T., & Demuth, H.B. (2013). Neural network toolbox™ user's guide. The MathWorks, Inc., Natick.

Belevičius, R., Ivanikovas, S., Šešok, D., Valentinavičius, S., & Žilinskas, J. (2011). Optimal placement of piles in real grillages: experimental comparison of optimization algorithms, information technology and control, .40(2), 123 – 132.

Belohlavek, R. (1999). Feed forward networks with fuzzy signals. Soft Computing, 3, 37 – 43.

Behzadian, K., Kapelan, Z., Savic, D., & Ardeshir, A. (2009). Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. Environmental Modelling & Software, 24(4), 530-541.

Bishop, M.C. (2006). Pattern recognition and machine learning. Springer: New York.

Blanco, A., Delgado, M., & Pegalajar, M.C. (2001). A real-coded genetic algorithm for training recurrent neural networks. Neural Networks, 14, 93–105.

Blanz, V., Schölkopf, B., Bülthoff, H. C. B. V. V., Burges, C., Vapnik, V., & Vetter, T. (1996). Comparison of view-based object recognition algorithms using realistic 3D models. In: Artificial Neural Networks—ICANN 96 (pp. 251-256). Springer Berlin Heidelberg.

Bunke, H., & Kandel, A. (2000). Neuro-Fuzzy Pattern Recognition. World Scientific Publishing CO: Singapore.

Burges, C.J.C., & Sch¨olkopf, B. (1997). Improving the accuracy and speed of support vector learning machines. In: Mozer, M., Jordan, M. and Petsche, T. (Eds), Advances in Neural Information Processing Systems 9, 375–381, Cambridge, MA, MIT Press.

Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. Journal of Chemical Information and Computer Sciences, 43(6), 1882-1889.

Carcenac, M. (2008). A modular neural network applied to image transformation and mental images. Neural Computing & Applications, 17, 549–568.

Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. Neural Computing & Applications, 10(2), 184-192.

Capko, D., Erdeljan, A., Vukmirovic S., & Lendak, I. (2011). A hybrid algorithms for partitioning of data model in distribution management systems. Information Technology & Control, 4 (4), 316 – 321.

Capraro, C. T., Bradaric, I., Capraro, G. T., & Lue, T. K. (2008, May). Using genetic algorithms for radar waveform selection. In: Radar Conference, 2008. RADAR'08. IEEE (pp. 1-6). IEEE.

Chaiyaratana, N., & Zalzala, A. M. (1999). Hybridisation of neural networks and genetic algorithms for time-optimal control. In: Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on (Vol. 1). IEEE.

Charles, A., & Darne´, O. (2009). The efficiency of the crude oil markets: Evidence from variance ratio tests. Energy Policy, 37, 4267–4272.

Chen, A., Leung, M.T., & Wang, L. (2012). Application of polynomial projection ensembles to hedging crude oil commodity risk. Expert Systems with Applications, 39, 7864 -7873.

Chen, S. T., Yu, P. S., & Liu, B. W. (2011). Comparison of neural network architectures and inputs for radar rainfall adjustment for typhoon events. Journal of Hydrology, 405(1), 150-160.

Chen, X., & Qu, Y. (2011). A prediction method of crude oil output based on artificial neural networks. Proceeding of IEEE International Conference on Computation and Information sciences, 702 – 704.

Chen, Y., Su, C., & Yang, T. (2012). Rule extraction from support vector machines by genetic algorithms. Neural Computing & Applications. DOI 10.1007/s00521-012-0985-3.

Chen, S. T., Kuo, H. I., & Chen, C. C. (2010). Modeling the relationship between the oil price and global food prices. Applied Energy, 87(8), 2517-2525.

Chevillon, G., & Rifflart, C. (2009). Physical market determinants of the price of crude oil and the market premium. Energy Economics 31, 537 – 549.

Clementine. (2007). User's Guide for Neural Network. IBM Corperation, Integral Solutions Limited: USA.

Chin, S.Y., Ganapathy, V., & Ooi, L.M. (2010). Improved genetic algorithms based optimization path planning for mobile robot. In: International Conference on Control, Automation, Robotics and Vision, 1565 – 1570.

Cho, S. B. (1997). Combining modular neural networks developed by evolutionary algorithm. In: Evolutionary Computation, 1997., IEEE International Conference on (pp. 647-650). IEEE.

Choi, H., & Lee, C. (2011). Motion adaptive deinterlacing with modular neural networks. Circuits and Systems for Video Technology, IEEE Transactions on, 21(6), 844-849.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev, Lawrence Erlbaum Associates, Inc.

Coley, D. A. (1999). An introduction to genetic algorithms for scientist and engineers. World scientific publishing Co. Pte. Ltd: Singapore.

Davis, T. E., & Principe, J. C. (1991). A simulated annealing−like convergence theory for the simple genetic algorithm. In R. K. Belew and L. B. Booker, eds., Proceedings of the Fourth International Conference on Genetic Algorithms. Morgan Kaufmann.

De Falco, I., Del Balio, R., Della Cioppa, A., & Tarantino, E. (1996). Breeder genetic algorithms for airfoil design optimisation. In Evolutionary Computation, Proceedings of IEEE International Conference on (pp. 71-75). IEEE.

Dehuri, S., & Cho, S. (2010). A hybrid genetic based functional link artificial neural network with a statistical comparison of classifiers over multiple datasets. Neural Computation &Applications, 19, 317–328.

De Jong, K. (1988). Learning with genetic algorithms: An overview. Machine Learning, 3, 121 – 138.

Demetgul, M., Unal, M., Tansel, I. N., & Yazıcıoğlu, O. (2011). Fault diagnosis on bottle filling plant using genetic-based neural network. Advances in Engineering Software, 42(12), 1051-1058.

Demˇsar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research, 7, 1–30.

Deng, W., Chen, R., He, B., Liu, Y., Yin, L., & Guo, J. (2012). A novel two-stage hybrid swarm intelligence optimization algorithm and application. Soft Computing DOI: 10.1007/s00500-012-0855-z

Dreyfus, G. (2005). Neural networks: methodology and applications. Springer-Verlag: Berlin Heidelberg.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. Advances in Neural Processing system. 9, 155–161.

Dunis, C.L., Rosillo, R., Fuente, D., & Pino, R. (2012). Forecasting IBEX-35 moves using support vector machines. Neural Computing & Applications, DOI 10.1007/s00521-012-0821-9.

Editorial. (2009). Hybrid learning machine. Neurocomputing 72: 2729 – 2730

Ekonomou, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. Energy, 35(2), 512-517.

El – Sharif, I., Brown, D. and Burton, B. (2005). Evidence on the nature and extend of the relationship between oil prices and equity values in the UK. Energy Economics, 27, 819 – 830.

Estivill-Castro, V. (1993). The design of competitive algorithms via genetic algorithms Computing and Information, 1993. Proceedings ICCI, Fifth International Conference on, pp. 305 – 309.

Fadare, D. A. (2009). Modelling of solar energy potential in Nigeria using an artificial neural network model. Applied Energy, 86(9), 1410-1422.

Fan, Y., Liang, Q., & Wei, Y. (2008). A generalized pattern matching approach for multi – step prediction of crude oil price. Energy Economics 30, 889 – 904.

Fausett, L.V. (1994). Fundamentals of Neural Networks. Architectures, Algorithms, and Applications. Prentice-Hall: Englewood Cliffs.

Fattouh, B. (2010).The dynamics of crude oil price differentials. Energy Economics, 32, 334 – 342.

Fazel, M.H.Z., & Gamasaee, R. (2012). Type-2 fuzzy hybrid expert system for prediction of tardiness in scheduling of steel continuous casting process. Soft Computing DOI 10.1007/s00500-012-0812-x.

Fernandez, V. (2006). Forecasting crude oil and natural gas spot prices by classification methods. IDEAS web. http://www.webmanager.cl/prontus_cea/cea_2006/site/asocfile/ASOCFILE120061 1281 05820.pdf. Access 4 June 2012.

Fernandez-Lozano, C., Canto, C., Gestal, M., Andrade-Garda, J.M., Rabuñal, J.M., Dorado, J., & Pazos, A. (2013). Hybrid Model Based on Genetic Algorithms and SVM Applied to Variable Selection within Fruit Juice Classification. The Scientific World Journal, http://dx.doi.org/10.1155/2013/982438, Article ID 982438, 13 pages, 2013.

Flach, P.A. (2001) On the state of the art in machine learning: A personal review. Artificial Intelligence, 131,199–222.

Foster, A. E. (2002). Commodity futures price prediction, an artificial intelligence approach. Msc thesis sumitted to graduate faculty, University of Georgia.

Frechling D.C. (1996). Practical Tourism Forecasting. Elsvier.

Ganatra, A., Kosta, Y.P., Panchal, G., & Gajjar, C. (2011). Initial classification through backpropagation in a neural network following optimization through GA to evaluate the fitness of an algorthms. International journal of computer science and information technology, 3(1), 98 – 116.

Ghai, V.V. (2006). Knowledge based approach using neural networks for predicting corrosion rate. Msc thesis submitted to college of engineering and technology, University of Ohio.

Gheyas, I.A & Smith, L.S. (2010). A neural network-based framework for the reconstruction of incomplete data sets. Neurocomputing, 73, 3039 – 3065.

Ghaffari, A., & Zare, S. (2009). A novel algorithms for prediction of crude oil price variation based on soft computing. Energy Economics, 31, 531 – 536.

Gholamian, M. R., Fatemi Ghomi, S. M. T., & Ghazanfari, M. (2005). A hybrid systematic design for multiobjective market problems: a case study in crude oil markets. Engineering Applications of Artificial Intelligence, 18(4), 495-509.

Goldberg, D.E. (1989). Genetic algorithms in search, optimization, machine learning. Addison Wesley: Reading.

Goldberg, D. E. (1987). Simple genetic algorithms and the minimal deceptive problem. In L. D. Davis, ed.,Genetic Algorithms and Simulated Annealing. Morgan Kaufmann.

Gradojevic, N., Gençay, R., & Kukolj, D. (2009). Option pricing with modular neural networks. IEEE Transactions on Neural Network, 20 (4), 629 – 637.

Granitto, P.M.,Verdes, P.F., & Ceccatto, H.A. (2005). Neural network ensembles: evaluation of aggregation algorithms. Artificial Intelligence, 163, 139–162.

Grefenstette, J. J. (1986). Optimization of control parameters for genetic algorithms. IEEE Transactions on Systems, Man, and Cybernetics, 16(1), 122–128.

Grimm, L.G., & Yarnold, P.R. (2002). Reading and understanding multivariate statistics. Washington DC, American Psychological Association.

Guo, F., & Qi, M. (2011). Application of Elman Neural Network based on Improved Niche Adaptive Genetic Algorithm. The 2nd international Conference on Intelligent Control and Information Processing, July 25 – 28, 660 – 664 .

Guo, P., Wang, X., & Han, Y. (2010, October). The enhanced genetic algorithms for the optimization design. In Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on (Vol. 7, pp. 2990-2994). IEEE.

Gurney, K. (1999). An introduction to neural networks. UCL Press Limited: New Fetter.

Guvenc, U., Duman, S., Saracoglu, B., & Ozturk, A. (2011). A hybrid GA–PSO approach based on similarity for various types of economic dispatch problems. Electron Electrical Engineering, Kaunas: Technologija 2(108), 109–114.

Haidar, I., & Wolff, C. R. (2011) Forecasting crude oil price (revisted). United State association of energy economics web. http://www.usaee.org/usaee2011. Accessed 10 June, 2012.

Haidar, I., Kulkarni, S., & Pan, H. (2008). Forecasting model for crude oil prices based on artificial neural networks. Symposium on Machine Learning and Applications. The Proceeding of fourth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 103-108.

Hair, F.J., Black, W.C., Babin, J.B. & Anderson, R.E. (2010). Multivariate data analysis. Pearson Prentice Hall: New Jersey.

Harpham, C., Dawson, C. W., & Brown, M. R. (2004). A review of genetic algorithms applied to training radial basis function networks. Neural Computing & Application, 13, 193–201.

Hamdan, M. (2008). A heterogeneous framework for the global parallelization of genetic algorithms. International Arab Journal of Information Technology, 5 (2), 192 – 199.

Hamel, L. (2009). Knowledge discovery with support vector machines. John Wiley & Sons, Inc. New Jersey.

Hamilton, J.D. (2011). Historical oil schocks. http://dss.ucsd.edu/~jhamilto/oil_history.pdf Prepared for the Handbook of Major Events in Economic History. Accessed 12 August, 2012.

Hanke, J.K., & Wichern, D.W. 2009. Business forecasting. Pearson international edition, 9th edn, 399 – 512.

Haupt, R. L., & Haupt, S.E. (2004). Practical genetic algorithms. John Wiley & Sons Inc.: New Jersey.

He, K., Chi, X., Chen, S., & Keung, K.L. (2009). Estimating VaR in crude oil market: A novel multi – scale non-linear ensemble approach incorporating wavelet analysis and neural network. Neurocomputing, 72, 3428 – 3438.

He, K., Yu, L., & Lai, K. K. (2012). Crude oil price analysis and forecasting using wavelet decomposed ensemble Model. Energy, 46, 564–574.

Hernandez, L., Baladrón, C., Aguiar, J.M., Carro, B., Sanchez-Esguevillas A.J., & Lloret, J. (2013). Short-Term Load Forecasting for Microgrids Based on Artificial Neural Networks. Energies, 6, 1385 – 1408.

Hasan, B. H. F., & Saleh, M. S. M. (2011). Evaluating the Effectiveness of Mutation Operators on the Behavior of Genetic Algorithms Applied to Non-deterministic Polynomial Problems. Informatica: An International Journal of Computing and Informatics, 35(4), 513-518.

Haykin, S. (2009). Neural networks and learning machines (3rd edn). Pearson Education, Inc. New Jersey.

Hinterding, R. (1995). Gaussian mutation and self-adaption for numeric genetic algorithms. In: Evolutionary Computation, IEEE International Conference on (Vol. 1, p. 384). IEEE.

Holland, J.H. (1975). Adaptation in natural and artificial systems. MIT Press: Massachusetts. Reprinted in 1998.

Hornick, K., Stinchcombe, M., & White, H. (1989). Multilayer feed forward networks are universal approximators. Neural Networks, 2, 359–366.

Huang, Y., & Shi, K. (1996). Genetic algorithms in the identification of fuzzy compensation system. System, man, and cybernetics, IEEE international conference on. 2, 1090– 1095.

Huang, C., Chen, L., Chen, Y., & Chang, F.M. (2009). Evaluating the process of a genetic algorithm to improve the back-propagation network: A Monte Carlo study. Expert Systems with Applications, 36, 1459–1465.

Hou, A., & Suardi, S. (2012). A nonparametric GARCH model of crude oil price return volatility. Energy Economics, 34, 618–626.

Husken, M., Jin, Y., & Sendhoff, B. (2005). Structure optimization of neural networks for evolutionary design optimization. Soft Computing, 9, 21–28.

International Monetary Fund. (2000). The impact of higher oil prices on the global economy. A report prepared by the research department and approved by Michael Mussa.

Ishikawa, M. (1995). Learning of modular structured networks. Artificial Intelligence, 75, 51-62.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural computation, 3(1), 79-87.

Jacobs, R. A., & Jordan, M. I. (1993). Learning piecewise control strategies in a modular neural network architecture. Systems, Man and Cybernetics, IEEE Transactions on, 23(2), 337- 345.

Jammazi, R., & Aloui, C. (2012). Crude oil forecasting: experimental evidence from wavelet decomposition and neural network modeling. Energy Economics, 34, 828 – 841.

Jareanpon, C., Pensuwon, W., Frank, R. J., & Davey, N. (2004). An Adaptive RBF Network optimised using a genetic algorithm applied to rainfall forecasting. In: Communications and Information Technology. IEEE International Symposium on, 2,1005-1010.

Jia, D., Zhang, G., Qu, B., & Khurram, M. (2011). A hybrid particle swarm optimization algorithm for high-dimensional problems. Computers & Industrial Engineering, 61,1117– 1122.

Jin, Y. (2005). A comprehensive survey of fitness approximation in evolutionary computation', Soft Computing, 9, 3 – 12.

Jinliang, Z., Mingming, T., Tao, M. (2009a). Effect simulation of international gold prices on crude oil prices. Proceedings of International Colloquium on Computing, Communication, Control and management 459 – 463.

Jinliang, Z., Mingming, T., & Tao, M. (2009b). Based on wavelet-boltzman neural network and kernel density estimation model predict international crude oil prices. Proceedings of IEEE International Conference on future Computer and Communication 150 – 153.

Ji, Q. (2011). System analysis approach for the identification of factors driving crude oil prices. Computers & Industrial Engineering. doi:10.1016/j.cie.2011.07.021.

Jie, L., Ding, F., & Bo, G. (2009). Parameter optimization of aeroengine PID controller based on genetic algorithms, 418 – 421.

Jo, S. (2011). The Effects of Oil Price Uncertainty on the Macroeconomy. Bank of Canada working paper, 2012 – 2040

Joachims, T. (1998). Text categorization with support vector machines: learning with many relevant features. In: Proceedings of ECML-98, 10th European Conference on Machine Learning, 137–142.

Ka, Y.W. & Chi, L.Y. (2008). Positioning Weather Systems from Remote Sensing Data Using Genetic Algorithms. Computational Intelligence for Remote Sensing,133, 217 – 243.

Kaboudan, M.A. (2001). Compumetric forecasting of crude oil prices. The Proceedings of IEEE Congress on Evolutionary Computation, 283–287.

Kafaee, M. & Saramad, S. (2009). Pile-up correction by Genetic Algorithmand Artificial Neural Network. Nuclear Instrument, Methematics and Physics Research A, 607, 652–658.

Kalyvas, E. (2001). Using neural networks and genetic algorithms to predict stock market returns. Msc thesis submitted to faculty of science and engineering, University of Manchester.

Karova, M. (2004). Solving timetabling problems using genetic algorithms. In Electronics Technology: Meeting the Challenges of Electronics Technology Progress, 2004. 27th International Spring Seminar on, 1, 96-98.

Kaufmann, R.K. (2011). The role of market fundamentals and speculation in recent price changes for crude oil. Energy Policy, 39, 105 – 115.

Kaynar, O., Yilmaz, I., & Demirkoparan, F. (2011). Forecasting of natural gas consumption with neural networks and neuro fuzzy system. Energy Education Science & Technology Part B, 26(2), 221 – 238.

Keung, K.L., He, K., & Yen, J. (2007). Modeling VaR in crude oil market: A multi scale nonlinear ensemble approach incorporating wavelet analysis and ANN. Springer, 554 – 561.

Khan, E. (1998). Neural fuzzy based intelligent systems and applications. In: Lakhmi, C.J.; Martin, N.M. (eds) Fusion of neural networks, fuzzy systems and genetic algorithms: industrial applications, Chapter 5.

Khashei, M., Bijari, M., & Reza, S.H. (2012). Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting. Soft Computing, 16, 1091 – 1105.

Khashman, A., & Nwulu, I.N. (2011a). Intelligent prediction of crude oil using support vector machines. 9th IEEE International Symposium on Applied Machine Intelligence Informatics, 165 – 169.

Khashman, A., & Nwulu, I.N. (2011b). Support vector machine versus back propagation algorithms for oil prediction. Springer – Verlag, 530 – 538.

Khazem, H.A. (2007). Using artificial neural networks to forecast the futures prices of crude oil. A PhD thesis submitted to school of business and entrepreneurship, Nova Southeastern University.

Khoukhi, A., Oloso, M., Elshafei, M., Abdulraheem, A., & Al-majed, A. (2011). Support vector regression and functional networks for viscosity and gas/oil ratio curves estimation. International Journal of Computational Intelligence and Applications, 10(03), 269-293.

Kilian, L. (2006). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. University of Michigan and CEPR.

Kim, K., & Boo, W. L. (2004). Stock market prediction using artificial neural networks with optimal feature transformation. Neural network and applications, 13, 255 – 260.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological cybernetics, 43, 59 – 69.

Kotsiantis, S.B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica, 31, 249-268.

Koza, J. (992). Genetic programming. MIT press: Cambridge.

Krawiec, K. (2007). Generative learning of visual concept using multi objective genetic programming. Pattern Recognition Letters. 28, 2385 – 2400.

Kromer, P., Platos, J, Snasel, & V. Snasel. (2009). Modeling permutation for genetic algorithms. In: Proceedings of the International Conference on soft computing and pattern recognition, 100 – 105.

Kulkar, S., & Haidar, I. (2009). Forecasting model for crude oil price using artificial neural networks and commodity future prices. International Journal of Computer Science & Information Security, 2(1), 81 – 88.

Lacks, R., Borgermann, C., & Dirkmorfeld, M. (2009). Forecasting the price development of crude oil with artificial neural networks. Proceedings of International Conference on Artificial Neural Networks, 248 – 255

Laboudi, Z., & Chikhi, S. (2012). Comparison of genetic algorithms and quantum genetic algorithms. International Arab Journal of Information Technology, 9 (3), 243 – 249.

Leardi, R., & Lupianez, A. G. (1998). Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemometrics and Intelligence laboratory systems, 41, 195– 207.

Lertpalangsunti, N. (1997). An implemented framework for the construction of hybrid intelligent forecasting systems. PhD thesis, University of Regina.

Lin, C. (2009). An adaptive genetic algorithms based on population diversity strategy. In: Proceedings of the 3rd International Conference on genetic and evolutionary computation, 93 – 96.

Lisboa, P. J., Edisbury, B., & Vellido, A. (2000). Business applications of neural networks: the state-of-the-art of real-world applications (Vol. 13). OECD Publishing.

Liu, J., Bai, Y., & Li, B. (2007). A new approach to forecast crude oil price based on fuzzy neural network. Proceedings of IEEE Computer Society Fourth International Conference on Fuzzy System and Knowledge Discovery art. No. 4406243 273 – 277

Liu, J. (2009). Application of fuzzy neural networks based on Genetic Algorithms in Integrated Navigation system, IEEE second conference on Intelligent Computation Technology and Automation, 656 – 659.

Lizhe, Y., Bo, X., & Xianjie, W. (2010). BP network model optimized by adaptive genetic algorithms and the application on quality evaluation for class teaching. In Future Computer and Communication (ICFCC), 2010 2nd International Conference on, 3, V3- 273.

Lopes, D.C, Magalhães, M.R.M., Melo, J.D., & Dória, A.D.N. (2009). Implementation and Evaluation of Modular Neural Networks in a Multiple Processor System on Chip to Classify Electric Disturbance. In: proceedings of IEEE International Conference on Computational Science and Engineering, 412 – 417.

Ma, H., & Wu, Y. (2010). A grey forecasting model based on BP neural network for crude oil production and consumption in China. Proceedings of IEEE Computer Society Third International Symposium on Information Processing, 105 – 109

Ma, X., (2009). Fuel oil price forecasting using symbiotic evolutionary immune clustering neural network. Proceedings of Second International Conference on Intelligent Computation technology and Automation, 322 – 325.

Malliaris, M.E., & Malliaris, S.G. (2008). Forecasting energy product prices. European Journal of Finance, 14(6), 453 –468.

Malik, F., & Nasereddin, M. (2006). Forecasting output using oil prices: A cascaded artificial neural network approach. Journal of Economics and Business, 58(2), 168-180.

Man, F. K., Tang, S. K. & Kwong S. (1999). Genetic algorithms: Concepts and designs. Heidelberg: Springer-Verlang.

Mantzaris, D., Anastassopoulos, G., & Adamopoulos, A. (2011). Genetic algorithm pruning of probabilistic neural networks in medical disease estimation. Neural Networks, 24,831–835.

Martinez, G., Melin, P., & Castillo, O. (2005). Optimization of Modular Neural Networks Using Hierarchical Genetic Algorithms Applied to Speech Recognition. In: Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, 4, 1400 – 1405.

Mehdi, M.Z. (2009). Forecasting momentary price of crude oil and gas fossile energies in the world markets through fuzzy based modeling. International gas union web. http://www.igu.org/html/wgc2009/papers/docs/wgcFinal00056.pdf. Accessed 20 May 2012.

Mehrara, M., Moeini, A., & Ahrari, M. (2011). Comparison of two approach of neural network for forecasting of oil future prices. Asia Journal of Business Management Science 1(3), 54-64.

Meng, X., & Song, B. (2007). Fast genetic algorithms used for PID parameter optimization. In: Proceedings of the International Conference on automation and logistics, 2144 – 2148.

McInerney, M., & Dhawan, A.P. (1993). Use of genetic algorithms with backpropagation in training of feedforward neural networks, Neural Networks, IEEE International Conference on,1, 203 – 208.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 1943, pp.115–133. Reprinted in Anderson and Rosenfeld, 1988.

Michalewicz, Z. (1999). Genetic algorithms + data structures = evolution program. Springer –Verlag: New York.

Mingming, T., & Jinliang, Z. (2012). A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices. Journal of Economics and Business 64, 275 – 286.

Mingming, T., Jinliang, Z., & Mingxin, T. (2010). Effect simulation of international natural gas prices on crude oil prices based on WBNNK model. IEEE computer society Six International Conference on Natural Computation, 1643 – 1648.

Mitton, T. (2002). A cross-firm analysis of the impact of corporate governance on the East Asian financial crisis . Journal of Financial Economics, 64, 215–241.

Mitchell, M. (1998). An introduction to genetic algorithms. MIT press: Massachusetts.

Montana, D. J., & Davis, L.D. (1989). Training feedforward networks using genetic algorithms. In: Proceedings of the International Joint Conference on Artificial Intelligence. Morgan Kaufmann.

Morana, C.A. (2001). Semiparametric approach to short-term oil price forecasting. Energy Economics, 23 (3), 325–338.

Moshiri, S., & Foroutan, F. (2006). Forecasting nonlinear crude oil futures prices. Energy Journal, 27, 81–95.

Mostafa, M.M. (2010). Clustering the ecological footprint of nations using Kohonen's self –organizing maps. Expert Systems with Applications, 37, 2747 – 2755.

Movagharnejad, K., Mehdizadeh, B., Banihashemi, M., Sheikhi, M.K. (2011). Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural networks. Energy, 36, 3979 – 3984.

Murata, T., & Ishibuchi, H. (1995). MOGA: multi-objective genetic algorithms. Evolutionary Computation, IEEE International Conference on, doi:10.1109/ICEC.1995.489161.

Nagata, Y., & Hoong, C.K. (2003). Optimization of a fermentation medium using neural networks and genetic algorithms. Biotechnology Letters 25, 1837–1842.

Narayan, P. K., & Narayan, S. (2010). Modelling the impact of oil prices on Vietnam's stock prices. Applied Energy, 87(1), 356-361.

NeuroSoluitions. (2012). Getting started: Network performance. NeuroDimension, Inc., Gainesville USA.

Niculescu, S. P. (2003). Artificial neural networks and genetic algorithms in QSAR. Journal of Molecular Structure: THEOCHEM, 622(1), 71-83.

Nissanke, M. (2009). The Global Financial Crisis and the Developing World: Transmission Channels and Fall-outs for Industrial Development. Working paper available at http://www.unido.org/fileadmin/user_media/Publications/Research_and_statistics/Branch _publications/Research_and_Policy/Files/Working_Papers/2009/WP%2006%20Th e%20 Global%20Financial%20Crisis%20and%20the%20Developing%20Wor.pdf

Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. Expert Systems with Applications, 39, 12605–12617.

Organization of the Petroleum Exporting Countries (OPEC). (2012a). Brief history. Retrieved from http://www.opec.org/opec_web/en/about_us/24.htm. Accessed, 12 October, 2012.

Organization of Petroleum Exporting Countries (OPEC) . (2012b) Background information 161 ordinary meeting of the conference. Retrieved from http://www.opec.org/opec_web/static_files_project/media/downloads/press_room/PressB ack.pdf Accessed, 12 October, 2012.

Ortiz – Cruz, A., Rodriguez, E., Ibarra-Valdez, C., & Alvarez-Ramirez, J. (2012). Efficiency of crude oil markets: evidence from informational entropy analysis. Energy Policy 41, 365 – 373.

Osuna, E., Freund, R., & Girosi, F. (1997). Training support vector machines: an application to face detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pages 130–136.

Pacifici, F., Del Frate, F., Solimini, C., & Emery, W. J. (2008). Neural networks for land cover applications. In: Computational Intelligence for Remote Sensing, 267-293. Springer Berlin Heidelberg.

Pallant, J. (2010). SPSS survival manual: A step by step guide to data analysis using SPSS, McGraw-Hill International 2010.

Pan, H., Haidar, I., & Kulkarni, S. (2009). Daily prediction of short-term trend of crude oil prices using neural networks exploiting multimarket dynamics. Frontiers in Computer Science in China, 3(2), 177 -191.

Panella, M., Barcellona, F., Santucci, V., & D'Ecclesia, R.L. (2011). Neural networks to model energy commodity price dynamics. United State association of energy economics web. http://www.usaee.org/usaee2011/submissions/OnlineProceedings/Panella.pdf. Accessed 13 June 2012.

Panella, M., D'Ecclesia, R. L., Stack, D. G., & Barcellona, F. (2014). Crude oil prices and kernel–based models. International Journal of Financial Engineering and Risk Management, 1(3), 214-238.

Pang, Y., Xu, W., Yu, L., Ma, J., Lai, K. K., Wang, S., & Xu, S. (2011). Forecasting the crude oil spot price by wavelet neural networks using OECD petroleum inventory levels. New Mathematics and Natural Computation, 7(02), 281-297.

Park, J. H., Huh, S. H., Kim, S. H., Seo, S. J., & Park, G. T. (2005). Direct adaptive controller for nonaffine nonlinear systems using self-structuring neural networks. Neural Networks, IEEE Transactions on, 16(2), 414-422.

Patruikic, M., Milan, R.R., Jakovljevic, B., & Dapic, V. (2008). Electric energy forecasting in crude oil processing using support vector machines and particle swamp optimization. Proceedings of the 9th Symposium on Neural Network

Application in Electrical Engineering. Faculty of Electrical Engineering, university of Belgrade, Serbia, September 25 – 27

Peter, G.Z. (2000). neural networks for classification: a survey. IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, 30(4), 451-462.

Peter, G.Z., Patuwo, B.E., & Hu, M.Y. (2001). A simulation study of artificial neural networks for nonlinear time-series forecasting. Computers & Operation Research, 28, 381-396.

Plourde, A., & Watkins, G.C.(1998).Crude oil prices between 1985 and 1994: how volatile in relation to other commodities?. Resource and Energy Economics, 20, 245–262.

Phichhang, O.U., & Wang, H. (2011). Application of neural networks in modeling and forecasting volatility of crude oil markets: Evidence from US and China. Advance materials research, 230 – 232, 953 – 957.

Pil, J.H., Choi, B., & Wha, I.H. (2011). Euntai KimMulticlass Lagrangian support vector machine, Neural Computing & Applications DOI 10.1007/s00521-011-0755-7.

Qi, Y., & Zhang, W. (2009). The improved SVM for forecasting the fluctuation of international crude oil prices. Proceedings of IEEE International Conference on Electronic Commerce and Business Intelligence, 269 – 271.

Quek, C., Pasquier, M., & Kumar, N. (2008). A novel recurrent neural network-based prediction system for option trading and hedging. Applied Intelligence,29(2), 138-151.

Qunli, W., Ge, H., & Xiaodong, C. (2009). Crude oil forecasting with an improved model based on wavelet transform and RBF neural network. IEEE International Forum on Information Technology and Applications, 231 – 234.

Rafiq, S., Salim, R., & Bloch, H. (2009). Impact of crude oil price volatility on economic activities: An empirical investigation in the Thai economy. Resource Policy 34, 121– 132.

Rafiul, M.H., Nath, B., & Michael, K. (2007). A fussion of HMM, ANN and GA for stock market forecasting. Expert system with applications, 33, 171 – 180.

Raja, M. A. Z., & Samar, R. (2013). Neural network optimized with evolutionary computing technique for solving the 2-dimensional Bratu problem. Neural Computing and Applications, 23(7-8), 2199-2210.

Rast, M. (2001). Fuzzy neural networks for modeling commodity markets. Proceedings of IFSA World Congress and 20th NAFIPS International Conference 2: 952— 955.

Raudys, A. (2005). Accuracy of MLP based data visualization used in oil prices forecasting task. Lecture Notes Computer Science, 3617, 761 – 769.

Renders, J. M., & Flasse, S. P. (1996). Hybrid methods using genetic algorithms for global optimization. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 26(2), 243-258.

Reza, M.A., & Ahmadi, E.G. (2007). A hybrid artificial intelligence approach to monthly forecasting of crude oil prices time series. The Proceedings of the 10th International Conference on Engineering Application of Neural Networks, CEUR-WS284 160 – 167.

Reza, M., S., Teh, Y. W., Amineh, A., Saeed, R. A. S. Y., & Adel, L. (2011). Applications of support vector machines in oil refineries: A survey.International Journal of Physical Sciences, 6(27), 6295-6302.

Riechmann, T. (2001). Learning in economics: analysis and application of genetic algorithms. Physica – Verlag Heidelberg: New York.

Rooij, A.J.F.V., Jain, L.C., Johnson, R.P. (1998). neural networks training using genetic algorithms. World scientific publishing Co. Pte. Ltd : Singapore, 1998.

Rothlauf, F. (2002). Representations for genetic and evolutionary algorithms. Physica – Verlag Heidelberg: New York.

Runge, C. F., & Senauer, B. (2007). How biofuels could starve the poor. Foreign affairs, 41-53.

Saenko, I., & Kotenko, I. (2011). Genetic algorithms for role mining problem. In: Proceedings of the IEEE 19th International Euromicro Conference on Parallel, distributed and network based processing, 646 – 650.

Sakawa, M. (2002). Genetic algorithms and fuzzy multiobjective optimization. Kluwer Academic Publishers: Dordrecht.

Saishanmuga, V. R., & Rajagopalan, S. P. (2012). A neuro – genetic system for face recognition. International Journal of computer science Issues, 9(3), 263 – 267.

Samanta, B., Al-Balushi, K.R., & Al-Araimi, S.A. (2003). Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. Engineering Application of Artificial Intelligence,16 ,657–665.

Season, D. (2005). Non – linear PLS using genetic programming. PhD thesis submitted to school of chemical engineering and advance materials, University of Nwecastle.

Sedki, A., Ouazar, D., & El Mazoudi, E. (2009). Evolving neural network using real coded genetic algorithm for daily rainfall–runoff forecasting. Expert Systems with Applications, 36(3), 4523-4527.

Serter, M.U, Inan, O., & Yılmaz, N. (2012). A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA. Neural Computing and Applications DOI 10.1007/s00521-012-0982-6, 2012.

Shambora, W.E., & Rossiter, R. (2007). Are there exploitable inefficiencies in the futures market for oil?. Energy Economy, 29,18 – 27.

Shanker, M., Hu, M.Y., & Hung, M.S. (1996). Effect of data standardization on neural network training. Omega, 24, 385 – 397.

Shapiro, A. F. (2002). The merging of neural networks, fuzzy logic, and genetic algorithms. Insurance: Mathematics and Economics, 31(1), 115-131.

Sorsa, A., Peltokangas, R., & Leiviska, K. (2008, September). Real-coded genetic algorithms and nonlinear parameter identification. Intelligent Systems, 4th IEEE International Conference on, 2, 10-42.

Sorsa, A., Peltokangas, R., & Leiviska, K. (2009). Real – Coded genetic algorithms and nonlinear parameter identification, 4th International IEEE Conference on Intelligent System, 10 – 47.

Shouyang, W., Lean, Y., & Lai, K.K. (2005). **C**rude oil price forecasting with tei@i methodology**.** Journal of Systems Science and Complexity, 18(2), 145 – 166.

Sivanandam, S.N., & Deepa, S.N. (2008). Introduction to genetic algorithms. Springer – Verlag: New York.

Srivastava, V., Bipin, K.T., & Vinay, K.P. (2012). Evolutionary fuzzy clustering and functional modular neural network-based human recognition, Neural Computing & Applications. DOI 10.1007/s00521-012-0973-7.

Sorsa, A., Peltokangas, R., & Leiviska, K. (2008). Real coded genetic algorithms and nonlinear parameter identification. In: Proceedings of the IEEE 4th International Conference on Intelligent Systems, 42 – 46.

Sotoudeh, M., & Farshad, E. (2012). Application of neural network for forecasting gas price in America. Journal of Mathematics and Computer Science, 4(2), 216–226.

Soytas, U., Sari, R., Hammoudeh, S., & Hacihasanoglu, E. (2009). World oil prices, precious metal prices and macro-economy in Turkey. Energy Policy, 37(12), 5557–5566

Specht, D.F. (1988). Probabilistic neural networks for classification, mapping or associative memory. In: Proceedings of the IEEE International Conference on Neural Networks, 1, 525 – 532.

Su, F. C., & Wu, W. L. (2000). Design and testing of a genetic algorithm neural network in the assessment of gait patterns. Medical Engineering & Physics, 22(1), 67-74.

Tan, J., Zhang, Z., Zhen, L., Zhang, C., & Deng, N. (2012). Adaptive feature selection via a new version of support vector machine. Neural Computing & Applications. DOI 10.1007/s00521-012-1018-y.

Tan, M., He, G., Li, X., Liu, Y., Dong, C., & Feng, J. (2012). Prediction of the effects of preparation conditions on pervaporation performances of polydimethylsiloxane (PDMS)/ceramic composite membranes by backpropagation neural network and genetic algorithm. Separation and Purification Technology, 89, 142-146.

Tasdemir, S., Saritas, I., Ciniviz, M., Allahverdi, N. (2011). Artificial neural networks and fuzzy expert system comparison for prediction of performance and emission parameters on a gasoline engine. Expert Systems with Applications 38, 13912 – 13923.

Tay, F.E.H., & Cao, L. (2001) Application of support vector machines in financial time series forecasting. Omega, 29, 309 – 317.

Tehrani, R., & Khodayar, F. (2011). A hybrid optimization artificial intelligent model to forecast crude oil using genetic algorithms. African Journal of Business Management, 5 (34), 13130 – 13135.

Torban, F.B. (2010). Models for oil price prediction and forecasting .Thesis, San Diego State University.

Vavak, F., & Fogarty, T.C. (1996). Comparisonof steady state and generational genetic algorithms for use in nonstationaryenvironments, Evolutionary Computation, Proceedings of IEEE International Conference on, 192 – 195.

Verleger, P.K. (1993). Adjusting to volatility energy prices. Working paper, Institute for International Economics, Washing DC.

Versace, M., Bhatt, R., Hinds, O., & Shiffer, M. (2004). Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks. Expert Systems with Applications, 27, 417–425.

Vose, D.M., & Liepins, G.E. (1991). Punctuated equilibria in genetic search. Complex system 5, 31 – 44.

Vose, M. D. (1993). Modeling simple genetic algorithms. In L. D. Whitley, (ed.), Foundations of Genetic Algorithms 2. Morgan Kaufmann.

Wang, B., Huang, H., Wang, A.X. (2011). Support vector machine based MSM model for financial short-term volatility forecasting. Neural Computing & Applications. DOI 10.1007/s00521-011-0742-z.

Wang, J., Huopo, P., & Fajiang, L. (2012). Forecasting crude oil price and stock price by jump stochastic time effective neural network model. Journal of Applied Mathematics. Volume 2012, Article ID 646475, 15 pages, doi:10.1155/2012/646475.

Wang, T., & Yang, J. (2010). Nonlinearity and intraday efficiency test on energy future markets. Energy Economics, 32, 496 – 503.

Wang, S.Y., Yu, L., & Lai, K.K. (2004). A novel hybrid AI system framework for crude oil price forecasting. Lecture Notes in Computer Science, 3327, 233–242.

Wang, X. (1994). Modern power system planning. New York, USA: McGraw Hill.

Wang, S., Dong, X., Sun, R. (2010). Predicting saturates of sour vacuum gas oil using artificial neural networks and genetic algorithms. Expert Systems with Applications, 37, 4768–4771.

Whitley, L. D. (1993). An executable model of a simple genetic algorithm. In L. D. Whitley, ed., Foundations of Genetic Algorithms 2. Morgan Kaufmann.

Witten, I.H., Frank, E., & Hall, A.M. (2011). Data Mining: Practical Machine Learning Tools and Techniques (3rdEdn.). San Mateo: Morgan Kaufmann.

Wilson, J.H., & Keating, B. (2009). Business forecasting. McGraw Hill 6th edn, 343 – 346.

Woll, S. L., & Cooper, D. J. (1997). Pattern-based closed-loop quality control for the injection molding process. Polymer Engineering & Science, 37(5), 801-812.

Xie, W., Yu, L., Xu, S., & Wang, S. (2006). A new method for crude oil price forecasting based on support vector machines. Lecture Notes in Computer Science, 3994, 441–451.

Xu, W., Wang, J., Zhang, X., Zhang, W., & Wang, S. (2007). A new hybrid approach for analysis of factors affecting crude oil price. Proceedings of the 7th international conference on Computational Science, Part III: 964-971.

Xu, Y., & Liu, C. (2012). A rough margin-based one class support vector machine. Neural Computing & Applications. DOI 10.1007/s00521-012-0869-6.

Yacci, P. (2009). Feature selection of microarray data using genetic algorithms and artificial neural networks. Msc thesis submitted to department of biological science, Rochester Institute of Technology.

Ye, M., Zyren, J., & Shore, J. (2006). Forecasting short-run crude oil price using high and low- inventory variables. Energy Policy, 34, 2736–2743.

Yi, Y., & Qin, N. (2009). Oil price forecasting based on self – organizing data mining. Proceedings of IEEE International Conference on Grey systems and Intelligence services1386 – 139.

Yu, L., Wang, S., & Wen, B. (2008a). An AI-agent trapezoidal fuzzy ensemble forecasting model for crude oil prediction. Proceeding of the IEEE 3rd International Conference on Innovative Computing Information and Control, 327.

Yu, L., Wang, S., Wen, B., & Lai, K. K.(2008b) Forecasting crude oil price with an EMD- based neural networks ensemble learning paradigm. Energy Economics, 30, 2623 – 2635.

Yu, L., Wang, S., Lai, K.K. (2008c). A generalized intelligent – agent – based fuzzy group forecasting model for oil price prediction. Proceedings of IEEE International Conference on Systems, man and cybernetics 489 – 493.

Yu, L., Wang, S., Lai, K.K. (2005). A rough set refined text mining approach for crude oil market tendency forecasting. International Journal of Knowledge and System Sciences, 2(1),33 –46.

Yu, X., & Meng, B. (2006). Research on dynamics in group decision support systems based on multi-objective genetic algorithms. In Service Systems and Service Management, 2006 International Conference on, 2, 895-900.

Yu, W. (2009). A novel fuzzy-neural-network modeling approach to crude-oil blending. Control Systems Technology, IEEE Transactions on, 17(6), 1424-1431.

Yun, W., & Jae, H.K. (2010). Hedging strategy for crude oil trading and the factors influencing hedging effectiveness. Energy Policy, 38, 2404–2408.

Zalloi, M. M. (2009). Forcecasting momentary price of crude oil and gas fossile energies in the world markets through fuzzy based modeling. proceedings of the 24th World Gas Conference, Argentina, 5-9 October.

Zhang, L., & Zhang, X. (2011). Measurement of the Optical Properties Using Genetic Algorithm Optimized Neural Networks. In Photonics and Optoelectronics (SOPO), 2011 Symposium on, 1-4).

Zhang, X., Lai, K.K., & Wang, S. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. Energy economics, 30, 905 – 918.

Zhang, X., Wu, Q., & Zhang, J. (2010). Crude oil price forecasting using fuzzy time series. Proceedings of IEEE 3rd International Symposium on Knowledge Acquisition and Modeling, 213 – 216.

Zhang, X., Yu, L., Wang, S., & Lai, K.K. (2009). Estimating the impact of extreme events on crude oil prices: An EMD – based event analysis method. Energy economics, 31, 768 – 778.

Zhang, S., Chengqi, Z.Z., & Yang, Q. (2003). Data preparation for data mining. Applied Artificial Intelligence: International Journal, 17(5 – 6), pp. 375 – 381.

Zhang, G., Patuwo, B.E., & Hu, M.Y. (1998). Forecasting with artificial neural networks: the state of the art', International Journal o Forecasting, 14, 35-62.

Zhang, G., Y Hu, M., Eddy Patuwo, B., & C Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. European Journal of Operational Research, 116(1), 16-32.

Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159 – 175.

Zhang, D., & Zhou, L. (2004). Discovering Golden Nuggets: Data Mining in Financial Application. zhou ieee transactions on systems, man, and cybernetics—part c: applications and reviews, 34(4), 513 – 522.

Zimberg, B. (2008). Crude oil price forecasting with ANFIS. University of Vaasa web. http://lipas.uwasa.fi/~phelo/ICIL2008TelAviv/24.pdf. Accessed 3 July 2012

Zweiri, Y. H., Whidborne, J. F., & Seneviratne, L. D. (2003). A three-term backpropagation algorithm. Neurocomputing, 50, 305-318.

PUBLICATIONS

ISI INDEX JOURNALS

1. Chiroma, H. & Abdulkareem, S. (2015). Neuro–Genetic Model for Crude Oil Price Projection while Considering the Impact of Uncertainties. *International Journal of Oil, Gas and Coal Technology* (*ISI/SCOPUS Cited Publication*). **Accepted for publication (IN PRESS). Q3**
2. Chiroma, H., Abdulkareem, S., & Herawan, T. (2015). Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction. *Applied Energy*, *142*, 266-273. (*ISI/SCOPUS Cited Publication*) **Published**. **Q1**
3. Chiroma, H., Abdulkareem, S., Abubakar, A., & Usman, M.J. (2013). Computational Intelligence Techniques with Application to Crude Oil Price Projection: A Literature Survey from 2001–2012. *Neural Network World*, Vol. 23(6), 523-551. (*ISI/SCOPUS Cited Publication*) **Published**.**Q4**
4. Chiroma, H., Abdul-kareem, S., Herawan, T., Fister Jr., I. & Fister, I. (2015). Advances on the Modifications of Cuckoo Search Algorithm. *Neural Computing & Applications*. (*ISI/SCOPUS Cited Publication*). **Accepted with minor revisions (presently doing the revisions). Q2**
5. Chiroma, H. & Abdulkareem, S. Prediction of crude oil prices through a genetically evolved neural network with input features selection by genetic algorithm. *Applied Mathematics and Information Science*. (*ISI/SCOPUS Cited Publication*) **Under Review. Q1.**

CHAPTERS IN EDITED BOOK

Chiroma, H., Abdul-Kareem, S., Abubakar, A., Zeki, A. M., & Usman, M. J. (2014). Orthogonal Wavelet Support Vector Machine for Predicting Crude Oil Prices. Springer Singapore. (*ISI/SCOPUS Cited Publication*).

Chiroma, H., Abdulkareem, S., Abubakar, A. I., & Herawan, T. (2014). Kernel Functions for the Support Vector Machine: Comparing Performances on Crude Oil Price Data. In: *Recent Advances on Soft Computing and Data Mining* (pp. 273-281). Springer International Publishing. (*ISI/SCOPUS Cited Publication*).

Chiroma, H., Abdulkareem, S. *et al*. (2014). Soft Computing Approach in Modeling Energy Consumption. In: *Computational Science and Its Applications–ICCSA 2014* (pp.770-782). Springer International Publishing. Lecture notes in computer science (*ISI/SCOPUS Cited Publication*).

Chiroma, H., Abdulkareem, S., *et al*. (2015). An Intelligent Modeling of Oil Consumption. In: *Advances in Intelligent Informatics* (pp. 557-568). Springer International Publishing.

Chiroma, H., Abdul-Kareem, S., *et al*., (2014). Neural Network Intelligent Learning Algorithm for Inter-related Energy Products Applications. In: *Advances in Swarm Intelligence* (pp. 284-293). Springer International Publishing. (*ISI/SCOPUS Cited Publication*).

Chiroma, H., Abdul-Kareem, S., & Abubakar, A. (2014). A Framework for Selecting the Optimal Technique Suitable for Application in a Data Mining Task. In *Future Information Technology* (pp. 163-169). Springer Berlin Heidelberg. (*ISI/SCOPUS Cited Publication*).

Chiroma, H., Abdulkareem, S., Abubakar, A. I., Sari, E. N., Herawan, T., & Gital, A. Y. U. (2014). Hybridization of Haar Wavelet Decomposition and Computational Intelligent Algorithms for the Estimation of Climate Change Behavior. In *Information and Communication Technology* (pp. 238-247). Springer Berlin Heidelberg. (*ISI/SCOPUS Cited Publication*).

Chiroma, H., Abdulkareem, S., Abubakar, A. I., Sari, E. N., & Herawan, T. (2014). A Novel Approach to Gasoline Price Forecasting Based on Karhunen-Loève Transform and Network for Vector Quantization with Voronoid Polyhedral. In *Information and Communication Technology* (pp. 257-266). Springer Berlin Heidelberg. (*ISI/SCOPUS Cited Publication*).

CONFERENCE PROCEEDINGS

Chiroma, H., Abdulkareem, S., Abubakar, A., Zeki, A., Gital, A. Y. U., & Usman, M. J. (2013). Co—Active neuro-fuzzy inference systems model for predicting crude oil price based on OECD inventories. In *Research and Innovation in Information Systems (ICRIIS), 2013 International Conference on* (pp. 232-235). IEEE.

Chiroma, H., Abdulkareem, S., & Gital, A. Y. U. (2014). An Intelligent Model Framework for Handling Effects of Uncertainty Events for Crude Oil Price Projection: Conceptual Paper. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 1).

Chiroma, H., Abdulkareem, S., Abubakar, A., Zeki, A., & Ya'u Gital, A. (2013). Intelligent system for predicting the price of natural gas based on non-oil commodities. In *Industrial Electronics and Applications (ISIEA), 2013 IEEE Symposium on* (pp. 200-205). IEEE.

APPENDIX A

RESEARCH DATASET

Date	Brent ($/barrel)	OECDES (mb/d)	OECDCOC (tb/b)	USCOP (tb/d)	OPECCP (tb/d)	USCOS (tb/d)	USCOSR (tb/d)	USCOI (tb/d)	NOPECCP (tb/d)	WCOP (tb/d)	USESTG (tb/d)
May-1987	18.58	1562	36417.91	258426	0	1310	100773	134549	38884.04	55674.96	0
Jun-1987	18.86	3454	38864.65	248356	0	1087	101755	144197	38055.35	55375.71	0
Jul-1987	19.86	1580	39522.88	255782	0	981	103457	164131	38996.7	57939.92	0
Aug-1987	18.98	1616	38038.39	254508	0	954	105394	170802	38851.39	58737.6	0
Sep-1987	18.31	3619	39388.79	246163	0	832	105794	153312	39117.58	58130.08	0
Oct-1987	18.76	1637	39437.97	259272	0	777	105521	159390	39147.31	58325.25	0
Nov-1987	17.78	1662	39729.91	251915	0	758	105296	150389	39168.45	57862.85	0
Dec-1987	17.05	3643	39428.72	257844	0	964	95851	143825	39215.22	57936.91	0
Jan-1988	16.75	3637	39661.77	255743	0	1383	98805	144515	39285.66	57137.66	0
Feb-1988	15.73	3571	42468.69	242848	0	1504	97616	134863	39373.71	57217.71	0
Mar-1988	14.73	3496	42379.83	259587	0	1609	101042	150912	39571.92	57578.92	0
Apr-1988	16.6	3520	38609.68	248629	0	1259	107276	155012	39358.1	57890.1	0
May-1988	16.31	3605	37528.04	255089	0	1042	105691	165510	39112.93	57606.93	0
Jun-1988	15.54	3608	40004.52	245102	0	967	107954	159667	38465.74	57271.74	0
Jul-1988	14.91	3690	39143.87	249229	0	890	103275	158100	38799.99	57695.99	0
Aug-1988	14.89	3691	40136.57	250459	0	917	101807	157766	38706.67	58852.67	0
Sep-1988	13.18	3681	40202.82	236847	0	1121	102094	156355	38385.88	59206.88	0
Oct-1988	12.41	3692	40605.9	248704	0	1308	103149	172080	38868.87	60890.87	0
Nov-1988	13.02	3656	43416.84	240700	0	1319	96092	152093	38732.14	61350.14	0
Dec-1988	15.31	3588	40745.44	246189	0	1378	92129	162132	38813.29	61597.29	0
Jan-1989	17.17	3646	41106.85	246056	20704	1463	97056	175497	38761.55	58706.81	0
Feb-1989	16.89	3696	43349.78	218076	0	1344	100064	148539	38471.31	58227.58	0

Mar-1989	18.7	3537	42956.21	234822	0	1380	100856	156100	38550.51	58629.78	0
Apr-1989	20.32	3575	39519.13	233147	0	689	106668	172496	38354.68	59059.99	0
May-1989	18.63	3629	39091.25	242303	0	577	104770	177604	38204.64	58980.95	0
Jun-1989	17.67	3603	40722.96	228731	0	614	99200	179269	37638.74	59017.97	0
Jul-1989	17.62	3681	39036.04	230766	0	575	101924	192628	38099.91	59536.22	0
Aug-1989	16.77	3723	40991.57	233878	0	524	104758	203517	38524.43	60428.66	0
Sep-1989	17.77	3725	40500.17	226436	0	548	97697	180834	38466.56	60511.7	0
Oct-1989	18.91	3708	41722.77	231036	0	641	98753	191783	38686.88	61081.98	0
Nov-1989	18.73	3726	43123.21	226082	0	744	101485	185133	38770.13	61842.35	0
Dec-1989	19.84	3634	41440.45	227439	0	1021	99711	169361	38205.99	61485.18	0
Jan-1990	21.25	3671	42177.79	233931	21940	1241	105370	192572	38444.45	60921.39	60921.39
Feb-1990	19.81	3688	43049.65	209923	0	1000	98047	165046	37999.62	61178.88	61178.88
Mar-1990	18.39	3712	42113.63	230434	0	755	110567	189625	38698.65	62082.42	62082.42
Apr-1990	16.61	3744	40187.88	222215	0	714	105908	174393	38480.65	61805.93	61805.93
May-1990	16.35	3826	40306.57	227179	0	937	111473	200067	38081.38	61238.41	61238.41
Jun-1990	15.1	3824	40968.74	213175	0	868	111503	192704	37496.57	60409.37	60409.37
Jul-1990	17.17	3822	41803.57	222371	0	966	109670	212514	37352.08	60513.8	60513.8
Aug-1990	27.17	3812	43089.5	225886	0	557	105882	200000	37399.93	56965.63	56965.63
Sep-1990	34.9	3811	40312.97	216715	0	412	102107	169907	37757.09	59513.98	59513.98
Oct-1990	36.02	3805	41093.12	233799	0	459	102146	159080	38101.83	59854.23	59854.23
Nov-1990	33.07	3761	41710.54	221598	0	404	100526	152549	38186.58	60672.39	60672.39
Dec-1990	28.27	3739	41585.05	227463	0	460	94496	142930	37982.55	60883.96	60883.96
Jan-1991	23.57	3639	44249.66	232498	21895	707	93010	164167	38500.53	60637.25	60637.25
Feb-1991	19.54	3676	42949.44	213836	0	470	95317	153588	38256.49	60326.57	60326.57
Mar-1991	19.08	3653	40884.57	233926	0	573	100975	160149	38636.29	60581.69	60581.69
Apr-1991	19.18	3690	40824.23	225256	0	623	101380	165877	37777.77	59183.33	59183.33
May-1991	19.19	3669	40601.69	229670	0	460	111484	197257	37663.21	59006.36	59006.36

Jun-1991	18.17	3700	41055.26	219610	0	471	106599	190019	37084.33	59198.91	59198.91
Jul-1991	19.4	3715	41212.61	227756	0	468	105095	184597	37501.93	60191.61	60191.61
Aug-1991	19.77	3753	41430.8	226801	0	402	102769	205988	36736.59	59495.34	59495.34
Sep-1991	20.5	3794	41055.24	221029	0	481	97165	174347	37792.43	60534.06	60534.06
Oct-1991	22.21	3836	43024.71	230557	0	683	101185	176182	37515.58	60489.45	60489.45
Nov-1991	21.11	3802	43038.49	219833	0	671	98846	165853	37510.65	60735.22	60735.22
Dec-1991	18.41	3793	42061.61	226267	0	706	91203	172508	37523.71	61143.3	61143.3
Jan-1992	18.16	3742	44203.69	228178	23079	795	98335	184640	37237.56	61259.8	61259.8
Feb-1992	18.05	3720	44216.21	214269	0	499	99666	147293	36620.42	60422.66	60422.66
Mar-1992	17.63	3707	43320.06	227798	0	551	100343	164957	36725.96	59773.37	59773.37
Apr-1992	18.92	3661	42494.04	218775	0	318	101634	183798	37006.19	60133.34	60133.34
May-1992	19.89	3677	40181.11	222227	0	308	100632	187873	35762.88	59005.23	59005.23
Jun-1992	21.16	3731	41879.4	215019	0	363	95581	185143	35845.74	59247.74	59247.74
Jul-1992	20.24	3741	42746.15	221064	0	285	101348	210673	36055.59	59712.63	59712.63
Aug-1992	19.74	3746	41413.45	214575	0	263	99646	200165	35677.26	59699.38	59699.38
Sep-1992	20.27	3797	43077.03	210904	0	341	96867	186539	35703.8	59960.92	59960.92
Oct-1992	20.26	3805	43192.79	220898	0	318	97393	207565	36061.55	60758.67	60758.67
Nov-1992	19.21	3809	43857.63	210724	0	312	95641	183635	35712.31	60469.68	60469.68
Dec-1992	18.14	3790	43020.44	220200	0	365	87748	184060	35858.01	60790.38	60790.38
Jan-1993	17.39	3748	41878.91	215791	23830	298	99879	195059	35500.82	60623.27	60623.27
Feb-1993	18.47	3820	45169.01	194393	0	290	97874	172358	35711.85	60978.6	60978.6
Mar-1993	18.79	3764	45422.9	216198	0	354	98190	201113	35637.93	60263.6	60263.6
Apr-1993	18.67	3731	42559.88	206439	0	269	102928	207851	35455.89	59560.26	59560.26
May-1993	18.51	3751	40128.97	212255	0	325	104773	211074	35428.69	59746.64	59746.64
Jun-1993	17.65	3813	42693.95	203842	0	234	106378	216035	35038.55	59462.45	59462.45
Jul-1993	16.78	3837	42629.37	207339	0	278	100646	225954	35230.49	60068.67	60068.67
Aug-1993	16.7	3908	41974.46	209492	0	237	94911	205858	35072.8	59890.98	59890.98

Sep-1993	16.01	3932	43554.27	201347	0	241	92607	197433	34964.24	59902.29	59902.29
Oct-1993	16.61	3901	42915.57	212003	0	308	96993	222624	35467.44	60366.06	60366.06
Nov-1993	15.2	3822	45475.4	207349	0	314	97983	209905	35821.18	60429.91	60429.91
Dec-1993	13.73	3829	43438.59	212585	0	499	93261	211966	35893.63	60817.46	60817.46
Jan-1994	14.29	3736	44224.66	211320	24160	307	96577	184295	36294.94	61137.55	61137.55
Feb-1994	13.8	3706	46957.61	189569	0	336	98716	176774	36146.68	60959.68	60959.68
Mar-1994	13.82	3740	45413.92	209114	0	311	103724	197531	36017.01	60866.98	60866.98
Apr-1994	15.23	3813	43549.14	198363	0	274	102685	208661	35730.35	60409.54	60409.54
May-1994	16.19	3838	41951.9	207334	0	289	101088	223135	36038.98	60851.5	60851.5
Jun-1994	16.76	3885	43953.88	198318	0	217	100774	220729	36220.48	61161.43	61161.43
Jul-1994	17.6	3908	43361.55	201527	0	252	103984	243574	35941.79	60767.31	60767.31
Aug-1994	16.89	3932	44555.38	202869	0	204	100596	232117	36114.68	60661.91	60661.91
Sep-1994	15.9	3941	44616.66	198264	0	261	98110	236041	36249.72	61270.47	61270.47
Oct-1994	16.49	3902	44331.9	206398	0	234	95681	221230	36745.86	61768.27	61768.27
Nov-1994	17.19	3901	45572.28	198850	0	224	100930	211012	36850.97	61943.6	61943.6
Dec-1994	15.93	3843	44643.33	209550	0	312	97929	222973	37198.85	62265.65	62265.65
Jan-1995	16.55	3785	44849.73	207148	24790	230	102167	201657	36771.56	61841.11	61841.11
Feb-1995	17.11	3823	47406.51	190245	0	229	101749	183282	37076.45	62343.97	62343.97
Mar-1995	17.01	3839	46399.58	204613	0	223	101302	229132	36708.85	61605.45	61605.45
Apr-1995	18.65	3888	43841.03	198111	0	208	107365	211126	36986.29	62402.88	62402.88
May-1995	18.35	3938	43166	205509	0	207	101758	227067	36472.3	62413.88	62413.88
Jun-1995	17.31	3898	44205.92	197359	0	153	98065	237810	36278.61	61555.18	61555.18
Jul-1995	15.85	3919	42877.28	199912	0	223	98406	225215	36915.81	62507.51	62507.51
Aug-1995	16.1	3900	44760.49	199846	0	183	104607	230537	36824.65	62641.35	62641.35
Sep-1995	16.7	3879	45012.97	192484	0	183	98306	240210	37404.93	63036.63	63036.63
Oct-1995	16.11	3785	44267.04	199055	0	252	97953	219316	36806.85	62673.54	62673.54
Nov-1995	16.86	3754	46989.33	197559	0	220	100885	219061	37351.52	62892.32	62892.32

Dec-1995	17.93	3692	45122.54	202428	0	195	85731	214397	37634.35	63286.53	63286.53
Jan-1996	17.85	3658	46831.32	201354	25278	336	92442	226397	37494.65	63236.8	63236.8
Feb-1996	18	3702	48877.95	190727	0	244	90758	191754	37798.39	63633.93	63633.93
Mar-1996	19.85	3739	47099.81	203715	0	204	92993	223671	37558.5	63486.32	63486.32
Apr-1996	20.9	3760	45005.28	193317	0	186	98701	221116	37613.4	63338.64	63338.64
May-1996	19.15	3779	44070.92	198208	0	214	95315	248909	37551.58	63338.6	63338.6
Jun-1996	18.46	3790	44151.05	193746	0	183	103729	238747	37753.56	63660.57	63660.57
Jul-1996	19.57	3803	45477.48	196471	0	159	102546	241797	37715.14	63737.67	63737.67
Aug-1996	20.51	3794	45645.42	197162	0	173	106264	249286	37383.07	63386.74	63386.74
Sep-1996	22.63	3781	45333.13	194456	0	183	96195	220599	37801.69	63858.81	63858.81
Oct-1996	24.16	3785	47421.14	200907	0	153	95389	238717	38088.06	64222.4	64222.4
Nov-1996	22.76	3797	47271.31	194268	0	150	92068	220314	38508.69	64670.32	64670.32
Dec-1996	23.78	3759	46247.64	201685	0	182	83917	226532	38522.9	65244.27	65244.27
Jan-1997	23.54	3822	48032.46	198466	26552	140	96987	232255	38550.03	65182.45	65182.45
Feb-1997	20.85	3817	48039.86	182400	0	154	95016	208159	38655.55	65541.03	65541.03
Mar-1997	19.13	3861	45750.44	200013	0	146	97777	240380	38466.91	65528.6	65528.6
Apr-1997	17.56	3850	46675.2	193239	0	90	99712	239601	38731.4	66047.78	66047.78
May-1997	19.02	3842	44738.93	200702	0	131	103826	268234	38339.84	65398.23	65398.23
Jun-1997	17.58	3876	45971.17	193252	0	65	100863	262763	37934.27	64625.69	64625.69
Jul-1997	18.46	3900	47286.46	198691	0	69	98530	253506	38368.45	65070.24	65070.24
Aug-1997	18.6	3919	45652.86	196769	0	1	93208	267244	38112.99	65950.21	65950.21
Sep-1997	18.46	3948	47221.52	194575	0	1	98028	265199	38439.77	66312.32	66312.32
Oct-1997	19.87	3898	47953.21	200482	0	0	99540	276743	38817.59	66826.9	66826.9
Nov-1997	19.17	3925	47213.28	193778	0	0	101970	250987	38898.22	66680.22	66680.22
Dec-1997	17.18	3911	47012.69	202464	0	0	90663	237228	39075.37	66496.43	66496.43
Jan-1998	15.19	3891	46690.63	202756	27631	0	97791	258506	39166.39	67706.5	67706.5
Feb-1998	14.07	3947	48115.57	181321	0	0	100457	225255	39089.2	68081.35	68081.35

Mar-1998	13.1	4065	47959.32	198639	0	0	104200	251835	38989.1	67965.25	67965.25
Apr-1998	13.53	4048	46201.78	194483	0	0	107158	269552	38954.53	67828.25	67828.25
May-1998	14.36	4050	44218.27	196754	0	0	108207	278595	38488.23	67293.6	67293.6
Jun-1998	12.21	4108	46864.06	188018	0	0	106937	263842	38723.75	67014.12	67014.12
Jul-1998	12.08	4090	47500.54	192026	0	0	109684	294715	38771.94	66877.51	66877.51
Aug-1998	11.91	4096	46507.53	192282	0	0	103298	284481	38028.63	65903.83	65903.83
Sep-1998	13.34	4085	46936.12	173676	0	0	99572	255012	38019.15	65984.28	65984.28
Oct-1998	12.7	4029	47302.25	190420	0	0	105095	268678	38258.3	66098.41	66098.41
Nov-1998	11.04	4084	48079.18	184198	0	0	100411	268204	38841.42	66946.7	66946.7
Dec-1998	9.82	4014	47206.38	187347	0	0	99108	258909	38926.25	66771.57	66771.57
Jan-1999	11.11	3967	47371.34	184864	26479	0	105624	260189	39048.88	66986.56	66986.56
Feb-1999	10.27	4003	49942.42	167034	0	0	107750	237109	38844.06	67311.78	67311.78
Mar-1999	12.51	4049	50462.43	182362	0	0	108567	270921	38724.85	66988.91	66988.91
Apr-1999	15.29	4008	46798.07	176616	0	0	107334	277686	38537.01	65545.34	65545.34
May-1999	15.23	4020	44721.13	182122	0	0	115090	282050	38406.77	65351.49	65351.49
Jun-1999	15.86	4017	47555.48	172808	0	0	106727	266645	37903.97	64307.1	64307.1
Jul-1999	19.08	3990	47305.38	179738	0	0	110321	291123	38848.05	65818.11	65818.11
Aug-1999	20.22	3958	47568.67	179187	0	0	104502	276154	38484.27	65704.19	65704.19
Sep-1999	22.54	3914	47988.33	174113	0	0	100804	255799	38492.74	65747.39	65747.39
Oct-1999	22	3756	47935.25	184361	0	0	100063	266999	38934.43	66253.79	66253.79
Nov-1999	24.58	3766	48852.6	178811	0	0	99431	246722	39461.76	66236.07	66236.07
Dec-1999	25.47	3745	48185.5	184716	0	0	93024	255266	39513.65	65422.5	65422.5
Jan-2000	25.51	3733	47052.05	179316	28190	0	97824	242698	39390.79	66451.01	66451.01
Feb-2000	27.78	3774	49880.49	169703	0	0	101767	241236	39336.94	67070.15	67070.15
Mar-2000	27.49	3793	49004.75	183464	0	0	102119	272503	39434.92	67101.39	67101.39
Apr-2000	22.76	3823	46016.01	175625	0	0	105340	280228	39188.93	67758.29	67758.29
May-2000	27.74	3887	47040.09	181242	0	0	103648	281621	39158.36	68274.64	68274.64

Jun-2000	29.8	3848	47583.94	174686	0	0	101323	286000	39351.22	68070.84	68070.84
Jul-2000	28.68	3848	46904.51	177920	0	0	101429	291346	39651.04	68689.57	68689.57
Aug-2000	30.2	3858	49521.01	179451	0	0	103079	308103	39500.51	69518.59	69518.59
Sep-2000	33.14	3848	48575.09	172731	0	0	99565	284528	39525.44	69533.28	69533.28
Oct-2000	30.96	3866	48075.49	180080	0	0	95302	278052	39678.82	69968.79	69968.79
Nov-2000	32.55	3819	48577.63	174980	0	0	100576	267387	40274.55	70524.7	70524.7
Dec-2000	25.66	3801	48215.83	181508	0	0	97702	286114	40483.69	69268.66	69268.66
Jan-2001	25.62	3794	49612.27	179767	28339	0	101270	276911	40085.04	69183.65	69183.65
Feb-2001	27.5	3820	49557.93	161843	27900	0	96600	241046	40045.06	68704.93	68704.93
Mar-2001	24.5	3856	49033.41	182290	28575	0	103407	297683	40031.99	69367.05	69367.05
Apr-2001	25.66	3886	47083.39	175879	27872	0	109386	303320	39800.89	68437.47	68437.47
May-2001	28.31	3890	47050.05	180712	27622	0	105328	306450	39359.65	67747	67747
Jun-2001	27.85	3899	47109.21	172974	26139	0	102734	273160	39288.95	66193.23	66193.23
Jul-2001	24.61	3899	47931.7	178208	27334	0	104073	296127	40033.44	68132.38	68132.38
Aug-2001	25.68	3951	48795.2	177488	27820	0	103763	290870	39721.14	68306.07	68306.07
Sep-2001	25.62	3965	47371.28	171270	26972	0	102146	280182	40096.92	67834.03	67834.03
Oct-2001	20.54	3949	48049.85	178129	26919	0	103905	285531	40047.67	67731.28	67731.28
Nov-2001	18.8	3949	48798.46	176441	26807	0	102057	279594	40510.02	68081.66	68081.66
Dec-2001	18.71	3933	48263.73	182511	25903	0	102630	274020	41006.24	67673.96	67673.96
Jan-2002	19.42	3965	48669.43	182076	25276	0	94497	269965	40799.59	66927.16	66927.16
Feb-2002	20.28	3946	49186.9	164666	25306	0	106419	245094	40841.76	66998.49	66998.49
Mar-2002	23.7	3929	48046.39	182460	25526	0	108477	272781	40415.43	66794.85	66794.85
Apr-2002	25.73	3964	47257.07	175333	24344	0	108983	279024	41059.73	66258.45	66258.45
May-2002	25.35	3987	46125.6	183057	25208	0	112638	289026	40762.84	66829.48	66829.48
Jun-2002	24.08	3975	47038.15	176532	24920	0	104101	279718	40896	66675.97	66675.97
Jul-2002	25.74	3974	48375.23	178275	25497	0	105667	284708	40797.07	67156.75	67156.75
Aug-2002	26.65	3916	47958.32	179670	25356	0	103468	295874	40664.78	66884.73	66884.73

Sep-2002	28.4	3928	47919.18	162336	26183	0	92279	263909	40384.29	67435.25	67435.25
Oct-2002	27.54	3896	48546.4	166105	26961	0	100278	295480	41010.78	68846.21	68846.21
Nov-2002	24.34	3832	49415.11	168706	27042	0	93394	289623	41061.56	68976.41	68976.41
Dec-2002	28.33	3798	48224.46	177372	25261	0	91070	270973	41210.86	67322.85	67322.85
Jan-2003	31.18	3723	49476.63	178209	25791	0	98119	267623	41162.71	67648.85	67648.85
Feb-2003	32.77	3796	51350.79	161726	27112	0	100276	237278	41443.94	69266.06	69266.06
Mar-2003	30.61	3820	48715.79	179764	27730	0	102018	285994	41341.8	69794.5	69794.5
Apr-2003	25	3871	48051.64	171629	26934	0	106882	297835	41022.98	68679.95	68679.95
May-2003	25.86	3922	47144.47	175438	27080	0	102847	314748	40960.7	68654.07	68654.07
Jun-2003	27.65	3955	47734.87	169643	26346	0	104328	301151	40908.79	67867.94	67867.94
Jul-2003	28.35	3971	48325.89	170308	26480	0	101151	311061	41411.74	68504.6	68504.6
Aug-2003	29.89	3990	47887.98	172630	26964	0	100886	310705	41331.72	68968.61	68968.61
Sep-2003	27.11	3970	48873.58	168118	27361	0	100109	308615	41496.37	69580.68	69580.68
Oct-2003	29.61	3980	49508.93	173860	27936	0	103786	311950	41884.23	70542.9	70542.9
Nov-2003	28.75	3933	48594.51	166229	27971	0	96953	280530	42085.88	70779.63	70779.63
Dec-2003	29.81	3922	48897.85	172530	28640	0	92051	300206	42668.72	72046.8	72046.8
Jan-2004	31.28	3904	49638.49	172922	28868	0	96168	289762	42126.54	71752.37	71752.37
Feb-2004	30.86	3893	50985.41	161508	28824	0	95699	270188	42156.54	71738.92	71738.92
Mar-2004	33.63	3909	50590.64	173905	28709	0	103340	312739	42196.98	71674.36	71674.36
Apr-2004	33.59	3956	49114.05	166640	28677	0	101813	303443	42213.28	71668.66	71668.66
May-2004	37.57	3978	47153.46	172028	28494	0	104376	324012	42034.87	71317.26	71317.26
Jun-2004	35.18	4008	49013.71	162034	29667	0	98040	315989	42452.33	72917.56	72917.56
Jul-2004	38.22	4024	49456.96	169830	30071	0	98002	319253	42499.5	73388.23	73388.23
Aug-2004	42.74	4016	49248	164884	29878	0	96524	324258	41602.27	72318.23	72318.23
Sep-2004	43.2	4020	49702.74	152260	30398	0	93874	290923	41732.64	72968.6	72968.6
Oct-2004	49.78	4073	49723.62	160094	30423	0	98387	321222	42311.17	73579.62	73579.62
Nov-2004	43.11	3997	50477.1	162510	29900	0	94607	307154	42492.34	73252.8	73252.8

Dec-2004	39.6	4031	49755.62	170647	30103	0	93172	313120	41983.63	72956.82	72956.82
Jan-2005	44.51	4017	50275.5	168862	30347	0	92117	309901	41911.64	73118.43	73118.43
Feb-2005	45.48	4006	51897.89	154034	30467	0	98376	286124	42069.59	73416.9	73416.9
Mar-2005	53.1	4048	51531.39	173531	30662	0	101660	317497	42164.33	73712.3	73712.3
Apr-2005	51.88	4133	48984.99	166929	30770	0	101052	306734	42341.59	74019.25	74019.25
May-2005	48.65	4115	48267.78	173364	30646	0	105189	323400	42569.1	74135.32	74135.32
Jun-2005	54.35	4160	50003.79	164078	30792	0	104470	322953	42042.49	73779.35	73779.35
Jul-2005	57.52	4129	48996.05	162808	31034	0	101570	321697	41626.43	73615.07	73615.07
Aug-2005	63.98	4129	50720.58	161835	30983	0	99363	322526	41699.41	73652.46	73652.46
Sep-2005	62.91	4167	49367.28	126329	31382	0	101841	274650	40831.08	73197.78	73197.78
Oct-2005	58.54	4154	48282.24	141056	31034	0	102219	292751	41275.69	73264.99	73264.99
Nov-2005	55.24	4081	50739.54	145507	30975	0	103112	307852	41823.63	73753.91	73753.91
Dec-2005	56.86	4117	50120.39	154462	30905	0	100329	309886	42185.83	74045.93	73899.92
Jan-2006	62.99	4121	50327.17	157638	30613	0	97335	302761	41895.84	73469.18	73469.18
Feb-2006	60.21	4079	51265.74	140878	30661	0	101371	279530	41804.79	73426.21	73426.21
Mar-2006	62.06	4108	51103.4	155848	30493	0	99460	302241	41789.57	73252.91	73252.91
Apr-2006	70.26	4157	48009.27	152401	30521	0	104491	295774	41778.01	73269.32	73269.32
May-2006	69.78	4149	48184.82	159652	30192	0	102063	319400	41709.96	72871.65	72871.65
Jun-2006	68.56	4202	49750.11	154795	30431	0	101588	321369	41341.97	72817.72	72817.72
Jul-2006	73.67	4232	49057.52	157882	30793	0	98922	317114	41971.19	73829.09	73829.09
Aug-2006	73.23	4262	50041.74	156201	30924	0	97032	327476	41487.55	73476.75	73476.75
Sep-2006	61.96	4245	49307.93	150910	30571	0	98739	321298	41542.87	73179.2	73179.2
Oct-2006	57.81	4214	49711.27	158340	30345	0	97725	313272	42069.89	73505.07	73505.07
Nov-2006	58.76	4171	50533.56	151958	29943	0	95948	296627	42088.3	73131.75	73131.75
Dec-2006	62.47	4178	49814.58	160818	29870	0	83895	296219	41998.1	72968.35	72968.35
Jan-2007	53.68	4119	49420.33	158281	29526	0	94775	316538	41898.58	72560.95	72560.95
Feb-2007	57.56	4094	51110.49	143625	29577	0	92545	252238	42274.35	72850.76	72850.76

Mar-2007	62.05	4125	49935.36	158211	29632	0	100362	321781	42144.92	72771.76	72771.76
Apr-2007	67.49	4167	48553.95	155264	29861	0	99732	304815	42151.51	73011.46	73011.46
May-2007	67.21	4158	48434.45	161376	29722	0	101179	320168	41789.75	72514.43	72514.43
Jun-2007	71.05	4186	49174.27	152194	29615	0	101649	300437	41451.61	72073.37	72073.37
Jul-2007	76.93	4168	49275.28	156166	29905	0	100191	308120	41721.94	72642.93	72642.93
Aug-2007	70.76	4155	49609.02	154541	29869	0	95607	319805	41164.28	72058.45	72058.45
Sep-2007	77.17	4125	49213.97	147058	30507	0	101553	309200	41249.8	72788.87	72788.87
Oct-2007	82.34	4082	50217.06	156694	30673	0	99742	303317	41805.91	73526.75	73526.75
Nov-2007	92.41	4082	50558.02	151270	30538	0	94320	300115	41674.11	73244.84	73244.84
Dec-2007	90.93	4132	49633.3	158405	31207	0	89070	304870	41441.52	73703.03	73703.03
Jan-2008	92.18	4073	49720.06	158405	31172	0	93858	312536	41544.52	73774.74	73774.74
Feb-2008	94.99	4087	50316.51	149504	31266	0	94310	279439	41642.38	73971.1	73971.1
Mar-2008	103.64	4076	48104.36	160923	31481	0	98327	298712	41608.38	74157.45	74157.45
Apr-2008	109.07	4099	48843.83	154602	31169	0	99615	299373	41338.34	73578.87	73578.87
May-2008	122.8	4115	47404.18	159511	31509	0	96216	299584	41368.69	73903.83	73903.83
Jun-2008	132.32	4161	47164.29	154002	31573	0	92656	300528	41250.47	73910.79	73910.79
Jul-2008	132.72	4180	47929.51	160579	31918	0	93754	314092	41618.41	74636.58	74636.58
Aug-2008	113.24	4164	46677.22	155186	31712	0	96782	320049	40683.66	73509.72	73509.72
Sep-2008	97.23	4175	46535.78	119477	31394	0	94416	253396	39990.32	72511.42	72511.42
Oct-2008	71.58	4200	48114.67	146926	31397	0	92876	312681	41030.22	73554.08	73554.08
Nov-2008	52.45	4207	46682.22	152591	30632	0	95928	298309	41614.12	73390.17	73390.17
Dec-2008	39.95	4247	47946.22	158429	29978	0	86598	291995	41449.7	72589.7	72589.7
Jan-2009	43.44	4259	46896.88	159363	28974	0	96326	303136	41325.79	71480.03	71480.03
Feb-2009	43.32	4279	47195.32	146742	28893	0	99191	254061	41870.2	72001.35	72001.35
Mar-2009	46.54	4282	46564.26	161714	28767	0	101686	290728	41741.96	71808.26	71808.26
Apr-2009	50.18	4292	45550.11	158540	28815	0	97762	281218	41774.27	71960.97	71960.97
May-2009	57.3	4305	43915.43	166850	28836	0	96963	272716	41313.55	71555.99	71555.99

Jun-2009	68.61	4311	45644.38	158270	28926	0	95023	274054	41405.11	71762.23	71762.23
Jul-2009	64.44	4323	45483.28	167373	29248	0	94169	281911	41992.4	72612.61	72612.61
Aug-2009	72.51	4327	45079.56	166589	29357	0	94701	273235	41349.12	72104.89	72104.89
Sep-2009	67.65	4277	45796.46	166812	29365	0	94883	277619	41873.33	72579.4	72579.4
Oct-2009	72.77	4285	46191.98	171031	29405	0	93404	265536	42278.81	73035.43	73035.43
Nov-2009	76.66	4204	45751.9	161579	29337	0	93937	262201	42307.12	73010.89	73010.89
Dec-2009	74.46	4273	45944.9	168937	29210	0	90944	253260	42111.33	72709.83	72709.83
Jan-2010	76.17	4259	44975.82	167356	29413	0	93270	263249	42160.93	72860.22	72860.22
Feb-2010	73.75	4241	47095.97	155292	29449	0	93358	245302	42573.81	73568.61	73568.61
Mar-2010	78.83	4282	46825.58	170898	29329	0	98708	289582	42759.83	73763.81	73763.81
Apr-2010	84.82	4313	45825.33	161318	29378	0	97688	291766	42554.21	73735.67	73735.67
May-2010	75.95	4318	44737.01	167343	29419	0	96706	299304	42628.1	73766.56	73766.56
Jun-2010	74.76	4324	46528.55	161508	29632	0	98060	297814	42092.28	73872.01	73872.01
Jul-2010	75.58	4353	46520.21	164709	29519	0	96434	307894	42338.54	74139.34	74139.34
Aug-2010	77.04	4296	46822.04	168780	29536	0	94216	295837	42213.77	74062.56	74062.56
Sep-2010	77.84	4308	47550.63	168248	29430	0	93467	276863	42488.7	74368.65	74368.65
Oct-2010	82.67	4284	46093.75	173485	29467	0	95956	264726	42792.21	74238.69	74238.69
Nov-2010	85.28	4231	47107.6	166749	0	0	89527	260966	43134.29	74847.23	74847.23
Dec-2010	91.45	4281	46494.7	174045	0	0	88982	269553	43080.6	74822.94	74822.94
Jan-2011	96.52	4210	45814.04	170672	0	0	88892	284678	43017.1	75308.35	75308.35
Feb-2011	103.72	4185	47244.81	151807	0	0	85808	229140	42631.15	74519.04	74519.04
Mar-2011	114.64	4218	46700.36	173318	0	0	93978	284677	42696.95	73434.94	73434.94
Apr-2011	123.26	4243	44583.54	165819	0	0	93848	265167	42482.24	73340.92	73340.92
May-2011	114.99	4236	44267.43	173711	0	0	96011	280840	41715.01	72587.34	72587.34
Jun-2011	113.83	4242	45886.37	167109	0	0	95156	277050	41774.85	73294.73	73294.73
Jul-2011	116.97	4226	45723.97	167990	0	0	93695	287563	41982.37	73748.89	73748.89
Aug-2011	110.22	4186	47013.27	174645	0	0	91322	277006	42281.61	74182.09	74182.09

Sep-2011	112.83	4165	46615.25	167263	0	0	88791	267410	41662.86	73613.69	73613.69
Oct-2011	109.55	4182	45804.3	182035	0	0	93400	276125	42562.8	74219.32	74219.32
Nov-2011	110.77	4132	46352.89	179769	0	0	91921	261729	42686.17	75145.29	75145.29
Dec-2011	107.87	4127	46325.86	186002	0	0	90640	270037	42976.52	75582.12	75582.12

APPENDIX B

MATLAB CODE FOR THE NEURO-GENETIC MODEL

```
function ga()
clear all
close all
clc
warning('off','all');
warning;
tic;
tStart = tic;
record_Generation = [];
record_cputime = [];
record_accuracy = [];
ff='nnbptrain'; % objective function
npar=300; % number of optimization variables
varhi=1;
varlo=0; % variable limits
II Stopping criteria
maxit=50; % max number of iterations
mincost=0; % minimum cost
III GA parameters
popsize=20; % set population size
mutrate=0.01; % set mutation rate
CrossoverFraction=0.5; %crossover function
selection=1; % fraction of population kept
Nt=npar; % continuous parameter GA Nt=#variables
keep=floor(selection*popsize); % #population
members that survive
nmut=ceil((popsize-1)*Nt*mutrate); % total number of % mutations
M=ceil((popsize-keep)/2); % number of matings
Create the initial po_pulation
Generation=0;              % generation counter initialized
par=(varhi-varlo)*rand(popsize,npar)+varlo; % random
Lb=-1*ones(1,npar);
Upper bounds
Ub=1*ones(1,npar);
Random initial solutions
for i=1:popsize
par(i,:)=Lb+(Ub-Lb).*rand(size(Lb));
end
cost=feval(ff,par)*.1;  % calculates population cost
[cost,ind]=sort(cost);  % min cost in element 1
par=par(ind,:);        % sort continuous
minc(1)=min(cost)*.1;   % minc contains min of
meanc(1)=mean(cost);    % meanc contains mean of population
_____________________________________________________
Iterate through generations
loop=0;
while ((Generation<maxit)||(cost<mincost))
loop=loop+1;
increments generation counter
```

```
Pair and mate
M=ceil((popsize-keep)/2); % number of matings
prob=flipud([1:keep]/sum([1:keep])); % weights
chromosomes
odds=[2 cumsum(prob(1:keep))]; % probability
distribution
function
pick1=rand(1,M); % mate #1
pick2=rand(1,M); % mate #2
216 MATLAB CODE% ma and pa contain the indicies of the chromosomes
that will mate
ic=1;
while ic<=M
for id=2:keep+1
if pick1(ic)<=odds(id) && pick1(ic)>odds(id-1)
ma(ic)=id-1;
end
if pick2(ic)<=odds(id) && pick2(ic)>odds(id-1)
pa(ic)=id-1;
end
end
ic=ic+1;
end
```

```
Performs mating using single point crossover
ix=1:2:keep; % index of mate #1
xp=ceil(rand(1,M)*Nt); % crossover point
r=rand(1,M); % mixing parameter
```

```
Mutate the population
mrow=sort(ceil(rand(1,nmut)*(popsize-1))+1);
mcol=ceil(rand(1,nmut)*Nt);
for ii=1:nmut
par(mrow(ii),mcol(ii))=(varhi-varlo)*rand+varlo;
mutation
end % ii
```

```
The new offspring and mutated chromosomes are
evaluated
cost=feval(ff,par)*.1;
```

```
Sort the costs and associated parameters
[cost,ind]=sort(cost);
par=par(ind,:)*.1;
```

```
 Do statistics for a single nonaveraging run
 minc(Generation+1)=min(cost);
 meanc(Generation+1)=mean(cost);
 Generation=Generation+1;
```

```
if Generation>maxit|| cost(1)<mincost
break
end
[Generation cost(1)];
```

```
Upper_Limit = 1; % TOOK FROM CLEMENTINE'S NN MANUAL
Lower_Limit = -1; % TOOK FROM CLEMENTINE'S NN MANUAL
AC = ((1.0 - abs(cost) / (Upper_Limit - Lower_Limit)) * 100.0);
Err=[];
Err =[Err; cost];
tElapsed=toc(tStart);
fprintf('time=%g\t',tElapsed)
fprintf('Generation=%d  MSE=%g  AC=%g\n',Generation, Err, AC)
Err1(Generation)=Err;
record_Generation = [record_Generation Generation];
record_cputime =[record_cputime tElapsed];
record_accuracy = [record_accuracy AC];
records = [record_Generation' record_cputime' record_accuracy' Err1'];
end
figure(1)
plot(Err1);
xlabel('Generation '), ylabel('Fitness'), title('Neuro-Genetic Fitness')
save('GANN.mat','records')
end

function objval = nnbptrain(par)
[Nind Nvar]=size(par);


load PhDNormalizedDATA
X=PhDNormalizedDATA;
size(X);
size(X,1)*.8;
m=size(X,1)*.8;
ceil(m);
%
%    % training data %%
%    %%%%%%%%%%%%%%%%%%%%%
train=X(1:ceil(m),:);
x=train(:,1:10)';
t=train(:,11)';

% % testing data %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

test=X(ceil(m)+1:296,:);
x=test(:,1:10)';
t=test(:,11:11)';
%
%DATASTART
load DATASTART
X=DATASTART;
size(X);
size(X,1)*.8;
m=size(X,1)*.8;
ceil(m);
% %
% %    % training data %%
% %    %%%%%%%%%%%%%%%%%%%%%
```

```
   train=X(1:ceil(m),:);
   x=train(:,1:5)';
   t=train(:,6:6)';

% % % testing data %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

test=X(ceil(m)+1:51,:);
x=test(:,1:5)';
t=test(:,6:6)';

%
%DATAONE
 load DATAONE
 X=DATAONE;
 size(X);
   size(X,1)*.8;
   m=size(X,1)*.8;
    ceil(m);
% %
% %    % training data %%
% %    %%%%%%%%%%%%%%%%%%%%
   train=X(1:ceil(m),:);
   x=train(:,1:5)';
   t=train(:,6:6)';

% % testing data %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

test=X(ceil(m)+1:51,:);
x=test(:,1:5)';
t=test(:,6:6)';
%
DATATWO
 load DATATWO
 X=DATATWO;
 size(X);
   size(X,1)*.8;
    m=size(X,1)*.8;
    ceil(m);

% % %    % training data %%
% % %    %%%%%%%%%%%%%%%%%%%%
  train=X(1:ceil(m),:);
   x=train(:,1:5)';
   t=train(:,6:6)';

% % % testing data %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

 test=X(ceil(m)+1:51,:);
 x=test(:,1:5)';
 t=test(:,6:6)';
```

```
% DATATHREE
 load DATATHREE
 X=DATATHREE;
 size(X);
   size(X,1)*.8;
   m=size(X,1)*.8;
    ceil(m);
% % %
% % %    % training data %%
% % %    %%%%%%%%%%%%%%%%%%%%
   train=X(1:ceil(m),:);
   x=train(:,1:5)';
   t=train(:,6:6)';

% % % testing data %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

test=X(ceil(m)+1:51,:);
x=test(:,1:5)';
t=test(:,6:6)';
% DATAFOUR
load DATAFOUR
X=DATAFOUR;
size(X);
  size(X,1)*.5;
  m=size(X,1)*.5;
   ceil(m);
% %
% %    % training data %%
% %    %%%%%%%%%%%%%%%%%%%%
  train=X(1:ceil(m),:);
   x=train(:,1:5)';
   t=train(:,6:6)';
% % % testing data %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

test=X(ceil(m)+1:51,:);
x=test(:,1:5)';
t=test(:,6:6)';


            [ni N] = size(x);
             [no N] = size(t);
             nh = 5;
             inp=ni;
             for i=1:Nind
             z=par(i,:);
             wih = reshape(z(1:nh*inp),nh,inp);
             b1 = reshape(z(nh*inp+1:nh*inp+nh),nh,1);
             who = reshape(z(nh*inp+nh+1:nh*inp+nh+(nh*no)),no,nh);
             b2 = reshape(z(nh*inp+nh+(nh*no)+1:nh*inp+nh+(nh*no)+no),no,1);
             netj = wih*x+b1*ones(1,N);
```

```
    outj =tansig(netj);
    outj=logsig(netj);
      outj=compet(netj);
      netk = who*outj + b2*ones(1,N);
   outk=tansig(netk);
     y=logsig(netk)*.1; %% linear transfer function%%
  delk = ((outk*(1-outk)')*(t-outk));
  s=0;
  s = s + who'*delk;
  delj =outj*(1-outj)'*s;
  who = who+.2*delk*outj';
  wih = wih+.2*delj*x';
  h = tansig(wih*x+b1*ones(1,N));
  y = tansig(who*h+b2*ones(1,N))*.1;
    error=(t-y);
    objval = mse(error)'*.1;
   objval = mse(t-y)'*.1;
    end
    figure(2)
   predict =error';
   pred = [predict t'];
   plot(pred)
   title('Target data via  predict data')
  xlabel('Time (sec)','fontsize',10); ylabel('x(t)','fontsize',10);
 hleg1=legend('predict data', 'Actual data');
   save('predGANN.mat','pred')
end
```