

**NEURO – GENETIC MODEL FOR THE PROJECTION OF CRUDE  
OIL PRICE CAPABLE OF HANDLING OF UNCERTAINTY**

**HARUNA CHIROMA**

**THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY**

**2015**

# UNIVERSITI MALAYA

## ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Haruna Chiroma (I.C./Passport No.: A03813576)

Registration/Matrix No.: WHA110056

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

Neuro-Genetic Model for the Projection of Crude Oil Price Capable of Handling  
Uncertainty

Field of Study: Neural Networks and Genetic Algorithm

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copy-right whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature

Date

Subscribed and solemnly declared before,

Witness’s Signature

Date

Name:

Designation:

## ABSTRACT

Some events occur sometimes without any warning, such as war, revolution, financial crises, terrorist attacks, political conflicts, false news, natural disasters, earthquakes, and extreme weather conditions. These types of events which we termed as uncertain events, when related to crude oil have significant effects on the price and will contribute to oil price volatility. Volatility in crude oil market 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. The purpose of the research is to propose a model that can predict the price of crude oil in the real world scenario. This study presents an alternative model based on Neural Network and Genetic Algorithm (Neuro-Genetic) for the projection of crude oil price while considering the impact of uncertainties. The difference between the crude oil price projected by the Neuro-Genetic model and the actual price was not statistically significant. The results obtained by the Neuro-Genetic model performs significantly better than the backpropagation neural network and support vector machine in both accuracy and CPU processing time. The model was able to learn patterns from volatile crude oil price datasets during the 1991 Gulf War, the 1997 Asian financial crisis, the 2002 Venezuelan unrest, the second Gulf War of 2003, and the 2007 global financial recession. The retraining applied in the modeling process possibly allow the Neuro-Genetic model to learn and capture new data patterns during the uncertain events. Thus, the model can effectively be applied as an alternative mechanism by policy makers in the formulation of policies related to energy demand and supply, bio-fuel, fuel subsidy, global food price subsidy, the stock market as well as national planning and budget. Intergovernmental organizations such as the Organization of Petroleum Exporting Countries (OPEC) can use our proposed model to serve as a guide for the

formulation of policies related to international crude oil price. The model has the potential for realistic, practical application in the real world.

## ABSTRAK

Beberapa peristiwa berlaku kadang-kadang tanpa sebarang amaran , seperti peperangan, revolusi , krisis kewangan , serangan penganas , konflik politik , berita palsu , bencana alam , gempa bumi , dan keadaan cuaca yang melampau . Jenis-jenis acara yang kita digelar sebagai peristiwa yang tidak menentu , apabila berkaitan dengan minyak mentah mempunyai kesan yang besar ke atas harga dan akan menyumbang kepada harga minyak turun naik . Turun naik dalam pasaran minyak mentah mempunyai kesan negatif secara langsung dan tidak langsung kepada ekonomi global dan ditimpakan penderitaan pada masyarakat di seluruh dunia. Kesan turun naik minyak mentah tidak mempunyai sempadan geografi kerana tidak ada sekatan kepada negara tertentu atau rantau di dunia. Tujuan kajian ini adalah untuk mencadangkan satu model yang boleh meramalkan harga minyak mentah dalam senario dunia sebenar . Kajian ini membentangkan satu model alternatif berdasarkan Neural Network dan Algoritma Genetik ( Neuro - genetik ) untuk unjuran harga minyak mentah manakala mengingati kesan ketidakpastian . Harga minyak mentah diunjurkan oleh model Neuro - Genetik dan harga sebenar didapati secara statistik sama . Keputusan yang diperolehi oleh model Neuro - genetik yang melakukan lebih baik daripada itu rangkaian neural rambatan balik dan sokongan mesin vektor dalam ketepatan dan masa pemprosesan CPU . Model ini dapat mempelajari corak dari set data harga minyak mentah yang tidak menentu semasa Perang Teluk 1991, 1997 krisis kewangan Asia 2002 rusuhan Venezuela , Perang Teluk kedua 2003 , dan 2007 kemelesetan kewangan global . Latihan semula yang digunakan dalam proses pemodelan mungkin membenarkan model Neuro - genetik untuk mempelajari dan menguasai corak data baru dalam acara yang tidak menentu . Oleh itu , model yang boleh berkesan digunakan sebagai mekanisme alternatif oleh pembuat dasar dalam penggubalan dasar-dasar yang berkaitan

dengan permintaan dan bekal tenaga , bio - bahan api , subsidi bahan api , global subsidi harga makanan , pasaran saham dan juga perancangan negara dan bajet . Pertubuhan antara kerajaan seperti Pertubuhan Negara-negara Pengeksport Petroleum (OPEC) boleh menggunakan model yang dicadangkan kami untuk berkhidmat sebagai panduan untuk menggubal dasar yang berkaitan dengan harga minyak mentah antarabangsa . Model ini mempunyai potensi untuk aplikasi praktikal realistik dalam dunia sebenar .

## **ACKNOWLEDGEMENTS**

I wish to give my gratitude to the almighty Allah for giving me the opportunity to complete the thesis. My sincere appreciation goes to my supervisor, Associate Professor Sameem Abdul-kareem, deputy Dean (Postgraduate research), Faculty of Computer Science and Information Technology for taking her time to guide and thoroughly go through each and every line of the thesis despite her tight schedules. I believed the constructive comments of my supervisor have significantly improved the quality of the thesis which could have not being so without her inputs.

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	ii
<b>Abstrak</b> .....	iv
<b>ACKNOWLEDGEMENTS</b> .....	vi
<b>List of Tables</b> .....	xiii
<b>List of Figures</b> .....	xvi
<b>List of Abbreviations</b> .....	xix
<b>Chapter One</b>	
1.1 Background.....	1
1.2 Neuro-Genetic model.....	5
1.2.1 Genetic Algorithm.....	5
1.2.2 Artificial Neural Network.....	5
1.2.3 Neuro-Genetic.....	6
1.3 Problem Statement.....	7
1.4 Aim of the Research.....	9
1.5 Objectives.....	9
1.6 Research Questions.....	10
1.7 Mapping the Objectives and Research Questions.....	11
1.8 Motivation.....	11

1.9 Methodology .....	12
1.10 Scope of the Research.....	14
1.11 Significance of the Study .....	14
1.11.1 To the Machine Learning Community .....	14
1.11.2 To the Energy and Economy .....	15
1.12 Thesis Outline .....	15

## **Chapter Two**

2.1 Introduction.....	18
2.2 Unexpected Events .....	18
2.3 The Negative Effects of Crude Oil Price Volatility .....	21
2.4 Impact on Developing and Transitional Economies .....	24
2.5 Crude Oil Market.....	25
2.5.1 International Crude Oil Price Benchmarks .....	26
2.5.2 Regular Attributes Affecting Crude Oil Price .....	29
2.6 Summary .....	33

## **Chapter Three**

3.1 Introduction.....	34
3.2 Method Employed in Literature Search.....	34
3.3 Hybrid Intelligent Systems vs. Individual Intelligent Systems .....	35

3.4 Applications of the Artificial Intelligent Techniques in the Projection of Crude Oil Price.....	36
3.4.1 Conventional Methods .....	36
3.4.2 Neural Network .....	38
3.4.3 Wavelet Transforms and Artificial Neural Network .....	47
3.4.4 Support Vector Machine .....	49
3.4.5 Fuzzy .....	50
3.4.6 Expert System.....	51
3.4.2 Hybrid Intelligent Systems .....	55
3.5 Hybrid Intelligent System Vs. Individual Intelligent System, Attribute Selection, and Model Evaluation in Crude Oil Price Projection .....	65
3.6 Frequency of Data Collection .....	72
3.7 Data Normalization .....	74
3.8 The Trend of the Researches in the Domain of Crude Oil Price Projection.....	75
3.9 Justification for Alternative Approach to Crude oil price Projection .....	75
3.10 Summary .....	76

## **Chapter Four**

4.1 Introduction.....	79
4.2 Artificial Neural Network.....	79

4.2.1 Mathematical Model of the Artificial Neural Network.....	83
4.3 Genetic Algorithm.....	84
4.4 Genetic Algorithms Terminologies .....	85
4.4.1 Fitness.....	85
4.4.2 Gene, Chromosome, Allele, Phenotype, Genotype and Breeding.....	85
4.4.3 Encoding and Decoding .....	86
4.4.4 Crossover (Single Point) .....	87
4.4.5 Mutation .....	88
4.4.6 Search Space.....	89
4.4.7 Population.....	90
4.4.8 Reproduction .....	90
4.4.9 Selection .....	91
4.4.10 Elitism .....	93
4.4.11Convergence and Search Termination .....	93
4.5 Strengths of Genetic Algorithm .....	95
4.6 Neuro-Genetic Model.....	96
4.7 Summary.....	97

## **Chapter Five**

5.1 Introduction.....	98
5.2 Data Collection.....	98

5.3 Data Analysis and Cleaning .....	94
5.4 Transformation of the Experimental Data .....	101
5.5 Attributes .....	104
5.6 The Selection of Input Attributes Using Genetic Algorithm .....	107
5.7 The State of the Art Methods of Input Attribute Selection.....	109
5. 8 The Proposed Neuro-Genetic Model.....	110
5.9 Summary.....	116

## **Chapter Six**

6.1 Introduction.....	117
6.2 Preliminary Results .....	117
6.2.1 Comparison between Raw and Normalized Data Standardizations methods .....	117
6.2.2 Results of the ANOVA .....	120
6.2.3 Regression Analysis.....	122
6.2.4 Selection of Input Attributes using Genetic Algorithm .....	127
6.2.5 BPNN Parameters .....	128
6.2.6 Correlation.....	128
6.2.7 Principal Component Analysis .....	129
6.2.8 Trial and Error .....	130
6.2.9 Haar Wavelet Transformation .....	130
6.2.10 Comparing Performance of the GA with the State of the Methods.....	132
6.3 Final Results: Neuro-Genetic Model.....	138

6.3.1 Training Performance Results .....	138
6.3.2 Testing Performance Results .....	142
6.4 Statistical Test .....	148
6.5 Summary .....	153
<b>Chapter Seven</b>	
7.1 Introduction .....	154
7.2 Data Standardization .....	155
7.3 Selection of Inputs Attributes by Genetic Algorithm .....	156
7.4 Neuro-Genetic Model .....	156
7.5 Further Work .....	158
7.6 Contributions of the study .....	159
7.6.1 Contributions to the Machine Learning .....	159
7.6.2 Contributions to the Energy Economy .....	160
References .....	162
Publications .....	188
Appendix A .....	192
Appendix B .....	204

## List of Tables

<b>Table 1.1:</b> Mapping the Objectives and Research Questions-----	11
<b>Table 2.1:</b> Unexpected events-----	20
<b>Table 2.2:</b> Factors responsible for the Interruption of crude oil supply-----	21
<b>Table 2.3:</b> Regular factors affecting crude oil price-----	29
<b>Table 3.1:</b> Summary of published papers by strengths-----	66
<b>Table 3.2:</b> Overview of published papers by methods-----	67
<b>Table 3.3:</b> Summary of published papers by weaknesses-----	70
<b>Table 3. 4:</b> Overview of published papers by the frequency of the data collection-----	73
<b>Table 4.1:</b> Population-----	90
<b>Table 5.1:</b> Sources of the data from different repositories in <a href="http://www.eia.doe.gov">http://www.eia.doe.gov</a> ----	99
<b>Table 5.2:</b> Sample of raw data before normalization-----	103
<b>Table 5.3:</b> Sample of normalized data-----	104
<b>Table 5.4:</b> The parameters values used in the Neuro-Genetic model-----	111
<b>Table 6.1:</b> Experimental results for the normalized and raw data methods-----	119
<b>Table 6.2:</b> ANOVA results (F and P-value) at a 95% confidence interval for mean MSE----- -----	120
<b>Table 6.3:</b> Regression results-----	123

<b>Table 6.4:</b> ANOVA results (F and P-value) for regression at a 95% confidence interval for the mean R-----	124
<b>Table 6.5:</b> 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-----	126
<b>Table 6.6</b> The ten best ten genetically evolved ANN with input features selection by GA-----	127
<b>Table 6.7:</b> The correlation coefficient relationship between the input attributes and the Brent crude oil price-----	129
<b>Table 6.8:</b> ANOVA post hoc results at 95% confidence interval for comparing training MSE performance among the models-----	134
<b>Table 6.9:</b> ANOVA post hoc results at 95% confidence interval for comparing validation MSE performance among the models-----	135
<b>Table 6.10</b> ANOVA post hoc results at 95% confidence interval for comparing test MSE performance among the models-----	136
<b>Table 6.11:</b> ANOVA post hoc results at 95% confidence interval for comparing convergence speed among the models-----	138
<b>Table 6.12:</b> Inter correlation matrix-----	139
<b>Table 6.13:</b> Performances of the Neuro-Genetic model compared with those of the state of the art approach based on SVM, and ANN-----	147
<b>Table 6. 14:</b> The t-test results of the Neuro-Genetic model-----	148

<b>Table 6.15</b> ANOVA Result for comparing the performance accuracy of Neuro-Genetic model, ANN, and SVM-----	150
<b>Table 6.16:</b> Tukey post hoc multiple comparison test for Neuro-Genetic, ANN, and SVM-----	151
<b>Table 6.17:</b> ANOVA test results for comparing CPU processing time of Neuro-Genetic, ANN, and SVM-----	151
<b>Table 6.18:</b> Tukey post hoc multiple comparison test of CPU processing time for the Neuro-Genetic model, ANN, and SVM-----	152

## List of Figures

<b>Figure 1.1:</b> Proposed methodology of the research-----	13
<b>Figure 2. 1:</b> WTI crude oil price over a period covering 1 January 1986 to 15 March 2011 -----	19
<b>Figure 2.2:</b> WTI crude oil price from January 1946 to May-----	20
<b>Fig. 2.3</b> Plots of WTI, Brent and Dubai -----	27
<b>Figure 3.1:</b> Visualization of the projection results by ANN-----	39
<b>Figure 3.2:</b> Crude oil price projection publications by year-----	75
<b>Figure 4.1:</b> Typical artificial neural network structure-----	81
<b>Figure 4.2:</b> Representation of Allele, Gene, Chromosome, Genotype and Phenotype----	86
<b>Figure 4.3:</b> Binary Encoding-----	87
<b>Figure 4.4:</b> Crossover (single point)-----	88
<b>Figure 4.5:</b> Mutation (single point)-----	89
<b>Figure 4.6:</b> Basic selection process-----	92
<b>Figure 4.7:</b> Rank Selection-----	92
<b>Figure 4.8:</b> Roulette Wheel-----	93
<b>Figure 4.9:</b> Genetic algorithms flowchart-----	94
<b>Figure 4.10:</b> Genetic algorithms cycle-----	95
<b>Figure 5.1:</b> Graphical representation of the data collected for the research -----	100

<b>Figure 5.2:</b> Attributes of the dataset -----	101
<b>Figure 5.3:</b> Chromosomes representing ANN-----	108
<b>Figure 5.4:</b> The propose framework for the Neuro-Genetic model-----	112
<b>Figure 5.5:</b> Indicating the impact of uncertainties on crude oil price over a period of 1987 to 2011-----	113
<b>Figure 5.6:</b> Pseudo-code of the Neuro-Genetic algorithms-----	115
<b>Figure 6.1:</b> Regression plots for raw and normalized data-----	124
<b>Figure 6.2:</b> KLT of the dataset with variance of each component-----	130
<b>Figure 6.3:</b> HWT scale coefficients of Brent crude oil price time series data-----	131
<b>Figure 6.4:</b> Comparing training MSE of the models-----	133
<b>Figure 6.5:</b> Validation MSE of the compared models-----	134
<b>Figure 6.6:</b> Comparison of testing MSE of the models -----	135
<b>Figure 6.7:</b> Comparing convergence speed of the models-----	137
<b>Figure 6.8:</b> Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Initial training prior to uncertainties)----- -----	140
<b>Figure 6.9:</b> Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (First Retraining)-----	140
<b>Figure 6.10:</b> Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Second Retraining)-----	141

<b>Figure 6.11:</b> Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Third Retraining)-----	141
<b>Figure 6.12:</b> Search results for training the Neuro-Genetic for crude oil price projection based on GA showing convergence performance (Fourth Retraining )-----	142
<b>Figure 6.13:</b> Comparing crude oil price projected by Neuro-Genetic model with an actual crude oil price (Initial training prior to uncertainties)-----	143
<b>Figure 6.14:</b> Comparison of crude oil price projected by Neuro-Genetic Model with an actual crude oil price (First Retraining)-----	143
<b>Figure 6.15:</b> Comparing crude oil price projected by Neuro-Genetic model with actual an crude oil price (Second Retraining)-----	144
<b>Figure 6.16:</b> Comparing crude oil price projected by Neuro-Genetic model with an actual crude oil price (Third Retraining)-----	144
<b>Figure 6.17:</b> Comparing crude oil price projected by Neuro-Genetic model with an actual crude oil price (Fourth Retraining) -----	145

## List of Abbreviations

<i>ac</i>	Percentage of accuracy
<b>AI</b>	Artificial Intelligence
<b>AIT</b>	Artificial Intelligent Techniques
<b>ANFIS</b>	Adaptive Neuro Fuzzy Inference Systems
<b>ANN</b>	Artificial Neural Network
<b>ANN-Q</b>	ANN–Quantitative
<b>ANOVA</b>	One-way Analysis of Variance
<b>ARIMA</b>	Autoregressive Integrated Moving Average
<b>B</b>	Batch weight/bias learning rules
<b>BFG</b>	BFGS Quasi-Newton
<b>BR</b>	Bayesian Regulation
<b>CL</b>	Crude oil
<b>Daqing</b>	Daqing oil field
<b>DOAJ</b>	Direct Open Access Journals
<b>EIAUSDE</b>	Energy Information Administration of the US Department of Energy
<b>EMD–ARIMA–ALNN</b>	ARIMA, empirical mode decomposition–feed forward ANNs– adaptive linear NNs

<b>EMD–ARIMA–Averaging</b>	Empirical Mode Decomposition- Autoregressive Integrates Moving Average
<b>EMD–FNN–ALNN</b>	Empirical Mode Decomposition FFNN Adaptive Linear ANN
<b>EMD–FNN–ALNN</b>	Empirical Mode Decomposition-Based ANN Ensemble Learning Pattern
<b>ENN</b>	Elman Neural Networks
<b>FNN</b>	Fuzzy Neural Network
<b>FOB</b>	Freight on Board
<b>FR</b>	Fuzzy regression
<b>GA</b>	Genetic Algorithm
<b>GARCH</b>	Autoregressive Conditional Heteroskedasticity
<b>GDP</b>	Gross Domestic Product
<b>GDX</b>	Gradient Descent BEP
<b>ROPF</b>	Regular Oil Price Fluctuation
<b>GMDHNN</b>	Group Method of Data Handling ANN
<b>GP</b>	Genetic Programming
<b>GPMGA</b>	Generalize Pattern Matching based on Genetic Algorithms
<b>GRNN</b>	Generalized Regression ANN

<b>HIS</b>	Hybrid intelligent systems
<b><i>hn</i></b>	Hidden layer nodes
<b>HO</b>	Heating oil
<b>HTW-MBPNN</b>	Harr a Trous Wavelet multilayer back- propagation ANN
<b>HU</b>	Gasoline
<b>HWT</b>	Haar Wavelet Transform
<b>HWTBPNN</b>	Haar Wavelet Transform backpropagation ANN
<b>IIS</b>	Individual Intelligent System
<b>IMF</b>	International Monetary Fund
<b>KLT</b>	Karhunen-Loève Transform
<b>KLTBNN</b>	Karhunen-Loève Transform backpropagation neural network
<b>LM</b>	Levenberg–Marquardt
<b>LMBP</b>	Levenberg – Marquardt backpropagation
<b>LRM</b>	Linear Regression Model
<b>LSSVM</b>	Least-Square-Support Vectors
<b>MAE</b>	Mean Absolute Error
<b>MAPE</b>	Mean Absolute Percentage Error
<b>MLFFNN</b>	Multilayer FFNN
<b>MLNN</b>	Multi-layer backpropagation ANN
<b>MLRM</b>	Multiple Linear Regression Model

<b>MPE</b>	Mean Percentage Errors
<b>MSE</b>	Mean Square Error
<b>NESWM</b>	Hybrid of NN, expert system, and web mining
<b>NG</b>	Natural gas
<b>NM</b>	Normalized Method
<b>NMSE</b>	Normalized Mean Square Error
<b>NOPECCP</b>	Non OPEC crude oil production
<b>NYMEX</b>	New York Mercantile Exchange
<b>OECD</b>	Organization for Economic Co – operation and Development
<b>OECDCOG</b>	OECD crude oil consumption
<b>OECDDES</b>	OECD crude oil ending stocks
<b>OPEC</b>	Organization of Petroleum Exporting Countries
<b>OPECCP</b>	OPEC crude oil production
<b>PCA</b>	Principal Component Analysis
<b>PMRS</b>	Pattern Modeling and Recognition System
<b>PN</b>	Propane
<b>R</b>	Regression
<b>RBFNN</b>	Radial Basis Function ANN
<b>rBPNN</b>	Correlation BPNN
<b>RM</b>	Raw Method
<b>RMSE</b>	Root Mean Square Error

<b>RNN</b>	Recurrent ANN
<b>RW</b>	Random Walk
<b>SHCI</b>	Shanghai Composite Index
<b>ARIMA</b>	Autoregressive Integrates Moving Average
<b>FNN</b>	Feed-forward ANNs
<b>SINOPEC</b>	Stock Price of China's largest oil company: China Petroleum & Chemical Corporation
<b>STEO</b>	Short Term Energy Outlook
<b>SVM</b>	Support Vector Machine
<b>SZCI</b>	Shenzhen Compositional Index
<b>SZPI</b>	Shenzhen Petrochemical Index
<b>TE</b>	The trial and error (TE)
<b>TEBPNN</b>	Trial and error backpropagation neural network
<b>USCOI</b>	US crude oil imports
<b>USCOP</b>	US crude oil production
<b>USCOS</b>	US crude oil supplied
<b>USCOSR</b>	US crude oil stocks at refineries
<b>USESTG</b>	US gasoline ending stocks
<b>WBNNK</b>	Wavelet Boltzmann cooperative ANNs and kernel density estimation
<b>WCOP</b>	World crude oil production
<b>WDNEVaR</b>	Wave Decomposition Network Value at Risk
<b>WDVaR</b>	Wavelet Decomposition Value at Risk

<b>W-LSSVM</b>	Wavelet and least-square-support vectors
<b>WRBFNN</b>	Wavelet transform and RBFNN
<b>WTI</b>	West Texas Intermediate