ENHANCED ARTIFICIAL BEE COLONY-LEAST SQUARES SUPPORT VECTOR MACHINES ALGORITHM FOR TIME SERIES PREDICTION

ZURIANI BINTI MUSTAFFA

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2014

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

ZURIANI MUSTAFFA

calon untuk Ijazah **PhD**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"ENHANCED ARTIFICIAL BEE COLONY-LEAST SQUARES SUPPORT VECTOR MACHINES ALGORITHM FOR TIME SERIES PREDICTION"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***14 Mei 2014.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: ***May 14, 2014.***

Pengerusi Viva: *(Chairman for VIVA)*	**Assoc. Prof. Dr. Huda Haji Ibrahim**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Prof. Dr. Puteh Saad**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Assoc. Prof. Dr. Faudziah Ahmad**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Yuhanis Yusof**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Siti Sakira Kamaruddin**	Tandatangan *(Signature)*

Tarikh:
(Date) ***May 14, 2014***

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to :

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUMCollege of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Sejak beberapa dekad yang lalu, Mesin Sokongan Vektor Kuasa Dua Terkecil (LSSVM) telah digunakan secara meluas dalam masalah ramalan di pelbagai domain aplikasi. Walaubagaimanapun, literatur sedia ada menunjukkan keupayaan LSSVM bergantung kepada nilai parameter hiper, iaitu parameter regularisasi dan parameter kernel, di mana ianya akan mempengaruhi generalisasi LSSVM dalam tugasan ramalan. Kajian ini mencadangkan algoritma hibrid, berdasarkan Koloni Lebah Buatan (ABC) dan LSSVM yang terdiri dari tiga algoritma; ABC-LSSVM, *lv*ABC-LSSVM and *cm*ABC-LSSVM. Algoritma *lv*ABC diperkenalkan untuk mengatasi masalah minimum setempat dengan menambah baik proses carian menggunakan mutasi Levy. Dalam pada itu, algoritma *cm*ABC yang menggunakan mutasi konvensional dapat mengatasi masalah penyesuaian terlebih atau penyesuaian terkurang. Kombinasi algoritma *lv*ABC dan *cm*ABC, yang kemudiannya dikenali sebagai algoritma Koloni Lebah Buatan Dipertingkat-Mesin Sokongan Vektor Kuasa Dua Terkecil (*e*ABC-LSSVM) telah direalisasikan pada ramalan harga komoditi sumber asli yang tidak boleh diperbaharui. Setelah tugas pengumpulan data dan pra pemprosesan data siap dilakukan, algoritma *e*ABC-LSSVM direkabentuk dan dibangunkan. Keupayaan *e*ABC-LSSVM dinilai berdasarkan lima metrik statistik, iaitu Min Peratusan Ralat Mutlak (MAPE), ramalan ketepatan, simetri Min Peratusan Ralat Mutlak (sMAPE), Peratusan Ralat Punca Kuasa Min (RMSPE) dan Theil's U. Keputusan menunjukkan *e*ABC-LSSVM mempunyai kadar ralat ramalan yang lebih rendah berbanding dengan lapan model hibrid antara LSSVM dan algoritma Evolusi Pengkomputan (EC). Tambahan pula, algoritma yang dicadangkan juga telah dibandingkan dengan teknik ramalan tunggal iaitu Mesin Sokongan Vektor (SVM) dan Rangkaian Neural dengan Rambatan ke Belakang (BPNN). Secara umumnya, *e*ABC-LSSVM telah menghasilkan ramalan ketepatan melebihi 90%. Ini menunjukkan *e*ABC-LSSVM berkeupayaan dalam menyelesaikan masalah optimisasi terutamanya dalam bidang ramalan. Algoritma *e*ABC-LSSVM dijangka dapat memberi manfaat kepada para pelabur dan pedagang komoditi dalam perancangan pelaburan dan pengunjuran keuntungan.

Kata kunci: Koloni Lebah Buatan, Mesin Sokongan Vektor Kuasa Dua Terkecil, Ramalan Siri Masa

Abstract

Over the past decades, the Least Squares Support Vector Machines (LSSVM) has been widely utilized in prediction task of various application domains. Nevertheless, existing literature showed that the capability of LSSVM is highly dependent on the value of its hyper-parameters, namely regularization parameter and kernel parameter, where this would greatly affect the generalization of LSSVM in prediction task. This study proposed a hybrid algorithm, based on Artificial Bee Colony (ABC) and LSSVM, that consists of three algorithms; ABC-LSSVM, *lv*ABC-LSSVM and *cm*ABC-LSSVM. The *lv*ABC algorithm is introduced to overcome the local optima problem by enriching the searching behaviour using Levy mutation. On the other hand, the *cm*ABC algorithm that incorporates conventional mutation addresses the over-fitting or under-fitting problem. The combination of *lv*ABC and *cm*ABC algorithm, which is later introduced as Enhanced Artificial Bee Colony – Least Squares Support Vector Machine (*e*ABC-LSSVM), is realized in prediction of non renewable natural resources commodity price. Upon the completion of data collection and data pre processing, the *e*ABC-LSSVM algorithm is designed and developed. The predictability of *e*ABC-LSSVM is measured based on five statistical metrics which include Mean Absolute Percentage Error (MAPE), prediction accuracy, symmetric MAPE (sMAPE), Root Mean Square Percentage Error (RMSPE) and Theils' U. Results showed that the *e*ABC-LSSVM possess lower prediction error rate as compared to eight hybridization models of LSSVM and Evolutionary Computation (EC) algorithms. In addition, the proposed algorithm is compared to single prediction techniques, namely, Support Vector Machines (SVM) and Back Propagation Neural Network (BPNN). In general, the *e*ABC-LSSVM produced more than 90% prediction accuracy. This indicates that the proposed *e*ABC-LSSVM is capable of solving optimization problem, specifically in the prediction task. The *e*ABC-LSSVM is hoped to be useful to investors and commodities traders in planning their investment and projecting their profit.

Keywords: Artificial Bee Colony, Least Squares Support Vector Machines, Time series prediction

Acknowledgement

Alhamdulillah…praise to Allah S.W.T for every breath I take, thankful for his mercy giving me a good health, strength and spirit throughout this journey.

First and foremost, I am heartily thankful to my main supervisor, Dr. Yuhanis Yusof for her valuable guidance and continuous support from the initial to the final level which enabled me to accomplish this research. She showed me different ways to approach a research problem and the need to be courage, persistent and patience to achieve my goal. Thank you for always reminds me for the thing that I overlooked, and thank you for being a great supervisor with a friendly touch. Your understanding is much indeed appreciated.

I also like to express my deepest appreciation to my co-supervisor, Dr. Siti Sakira Kamaruddin. The patience supervision and enthusiastic encouragement that she gave truly help the progression of this research. It is an honour for me to have both of you as my supervisors.

I would also like to dedicate this work to my lovely husband, Dr. Mohd Herwan. His continuous support and *du'a* means a lot me. Thank you for being with me, through ups and downs, and sharing the tears and laughs together. Not forgotten, my wonderful daughter, Zahra An Naurah who never fails to amaze me. She is such a glowing flower in my heart. Thank you for being my side throughout this meaningful journey. I owe both of them a lot. Last, I offer my regards and blessings to all families and friends of those who supported me in any respect during the completion of this research. Thank you for praying for my success.

May this little achievement will keep me grounded. Alhamdulillah for the end of this quest and for a new beginning. May Allah bless us all…

To the other half…the love of my life, Mohd Herwan
&
my sweet honey bee, Zahra An Naurah

Table of Contents

Permission to Use ii
Abstrak iii
Abstract iv
Acknowledgement v
Table of Contents vii
List of Tables xii
List of Figures xvi
List of Appendices xx
List of Abbreviations xxi

CHAPTER ONE INTRODUCTION 1

1.1 Background Study 1
1.2 Time Series Prediction 3
1.3 Problem Statement 10
1.4 Research Question 12
1.5 Objectives of Study 13
1.6 Limitation of Study 13
1.7 Scope of Study 14
1.8 Significance of Study 15
1.9 Thesis Outline 15

CHAPTER TWO LITERATURE REVIEW 17

2.1 Optimization of LSSVM Hyper Parameters 17
2.1.1 Optimization of LSSVM using Cross Validation 17
2.1.2 Optimization of LSSVM using Evolutionary Computation Algorithm 24
2.2 Methods to Enhance Searching Behavior in ABC Algorithm 39
2.3 Methods to Prevent Over fitting and Under fitting in LSSVM and SVM 44
2.4 Evaluation in Time Series Predictions 50
2.4.1 Metrics 50
2.4.2 Methods 56

2.4.2.1 Statistical Approach....56
2.4.2.2 Computational Intelligence Approach....59
2.5 Least Squares Support Vector Machines....68
2.6 Artificial Bee Colony....72
2.7 Levy Probability Distribution....77
2.7.1 Levy Probability Distribution in Evolutionary Computation....80
2.8 Summary....83
CHAPTER THREE METHODOLOGY....84
3.1 Introduction....84
3.2 Data Collection and Preparation....85
3.2.1 Test for Correlation....88
3.2.2 Test for Non linearity....90
3.2.3 Data Normalization....91
3.3 Algorithm Design....92
3.3.1 ABC-LSSVM....93
3.3.1.1Flow of ABC-LSSVM....93
3.3.1.1.1 Initialization Phase....93
3.3.1.1.2 EB Phase....95
3.3.1.1.3 OB Phase....96
3.3.1.1.4 SB Phase....97
3.3.2 Enrich the Exploitation Process in Search Space (*lv*ABC)....98
3.3.3 Preventing Over fitting and Under fitting (*cm*ABC)....100
3.3.4 *e*ABC-LSSVM....101
3.3.4.1 Flow of *e*ABC-LSSVM....102
3.3.4.1.1 Initialization Phase....103
3.3.4.1.2 EB Phase....104
3.3.4.1.3 OB Phase....106
3.3.4.1.4 SB Phase....107
3.4 Algorithm Development....107
3.5 Evaluation....108
3.5.1 Experimental Setup....108

3.5.1.1 Input and Output Variables.....109
3.5.1.2 Training, Validation and Testing.....111
3.5.1.3 Properties Setting… 112
3.5.2 Performance Evaluation Metrics.....113
3.5.3 Benchmarking Techniques.....115
3.6 Summary.....116

CHAPTER FOUR ARTIFICIAL BEE COLONY-LEAST SQUARES SUPPORT VECTOR MACHINES.....117

4.1 Introduction.....117
4.2 Summary.....121

CHAPTER FIVEENHANCED ARTIFICIALBEE COLONY-LEAST SQUARES SUPPORT VECTOR MACHINES.....122

5.1 Introduction.....122
5.2 Summary.....128

CHAPTER SIX RESULTS AND DISCUSSION.....129

6.1 Introduction.....129
6.2 *e*ABC-LSSVM on Energy Fuels.....129
6.2.1 Data Set A.....129
6.2.1.1 *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....129
6.2.1.1.1Crude Oil Price.....130
6.2.1.1.2 Heating Oil Price Prices.....135
6.2.1.1.3 Gasoline Price….....139
6.2.1.1.4 Propane Prices.....143
6.2.1.1.5 Results Analysis for Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....147
6.2.1.2 *e*ABC-LSSVM vs Other Techniques.....148
6.2.1.2.1 Crude Oil Price.....149
6.2.1.2.2 Heating Oil Prices.....152
6.2.1.2.3 Gasoline Prices.....…….....155
6.2.1.2.4 Propane Prices...…….....158

6.2.1.2.5 Results Analysis for Data Set B: *e*ABC-LSSVM vs. Other Techniques….....161
6.2.2 Data Set B….....162
6.2.2.1 *e*ABC-LSSVM vs ABC-LSSVM and Its Variants …..…...162
6.2.2.1.1 Crude Oil Prices….....163
6.2.2.1.2 Heating Oil Prices…….....….....……..167
6.2.2.1.3 Gasoline Prices….....171
6.2.2.1.4 Propane Prices….....175
6.2.2.1.5 Results Analysis for Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....179
6.2.2.2*e*ABC-LSSVM vs Other Techniques…….....….....180
6.2.2.2.1 Crude Oil Prices….....181
6.2.2.2.2 Heating Oil Prices….....184
6.2.2.2.3 Gasoline Prices….....187
6.2.2.2.4 Propane Prices….....190
6.2.2.2.5 Results Analysis for Data Set B: *e*ABC-LSSVM vs. Other Techniques….....193
6.3*e*ABC-LSSVM on Metal….....194
6.3.1 Data Set C….....194
6.3.1.1 *e*ABC-LSSVM vsABC-LSSVM and Its Variants….....194
6.3.1.1.1 Gold….....195
6.3.1.1.2 Results Analysis for Data Set C: *e*ABC-LSSVM vs ABC-LSSVM and Its Variants.....199
6.3.1.2 *e*ABC-LSSVM vs Other Techniques.....199
6.3.1.2.1 Gold.....200
6.3.1.2.2 Results Analysis for Data Set C: *e*ABC-LSSVM vs Other Techniques.....203
6.4 Summary……..…….....203

CHAPTER SEVEN CONCLUCIONS AND RECOMMENDATION FOR FUTURE WORKS.....…205

7.1 Conclusion.....205

7.2 Contribution.....205
7.2.1 Knowledge Contribution.....207
7.2.2 Practical Contribution.....208
7.3 Recommendation for Future Works.....209
REFERENCES.....211
APPENDICES.....225

List of Tables

Table 1.1: Correlation among Energy Fuels from 22 December 1992–30 November 2007 ...6
Table 2.1: Summaries of Time Series Prediction Technique using LSSVM Optimized by CV.....22
Table 2.2: Summaries of Time Series Prediction Technique using LSSVM Optimized by EC Algorithm.....36
Table 2.3:Summaries of Methods to Overcome the Searching Behavior in ABC Algorithm 43
Table 2.4: Summaries of Methods to Prevent Over fitting and Under fitting in LSSVM and SVM.....48
Table 2.5: Summaries of metrics utilized in Evaluating of Time Series Prediction.....54
Table 2.6: Summaries of Time Series Prediction Technique using Statistical Approach.....58
Table 2.7: Summaries of Time Series Prediction Technique using CI Approach.....66
Table 2.8: Data Representation of *e*ABC-LSSVM Algorithm.....77
Table 2.9: Summaries of Levy Probability Distribution in Evolutionary Computation Algorithm.....82
Table 3.1: Data Set A.....85
Table 3.2: Sample of Data Set A.....85
Table 3.3: Data Set B.....86
Table 3.4: Sample Data Set B.....86
Table 3.5: Data Set C.....87
Table 3.6: Sample of Data Set C.....87
Table 3.7: Price Correlation among Input Data Set A.....89
Table 3.8: Price Correlation among Input Data Set B.....89
Table 3.9: Price Correlation among Input Data Set C.....89
Table 3.10: Samples of Normalized Input for Data Set A.....92
Table 3.11: Samples of Normalized Input for Data Set B.....92
Table 3.12: Samples of Normalized Input for Data Set C.....92
Table 3.13: Initial Fitness Function.....94
Table 3.14: EB Phase.....95
Table 3.15: Probability.....96
Table 3.16: OB Phase 97

Table 3.17: Initial Fitness Functions.....104
Table 3.18: EB Phase.....105
Table 3.19: Probability.....105
Table 3.20: OB Phase.....107
Table 3.21: Assigning Input and Output Variable for Data Set A.....109
Table 3.22: Assigning Input and Output Variable for Data Set B.....110
Table 3.23: Assigning Input and Output Variable for Data Set C.....110
Table 3.24: Training, Validation and Testing.....111
Table 3.25 Example of data arrangement for Data Set A.....112
Table 3.26: Properties of Prediction Techniques Utilized.....113
Table 3.27: Range of Percentage by MAPE.....114
Table 6.1: CL Price Prediction Data Set A: *eABC*-LSSVM vs. ABC-LSSVM and Its Variants.....131
Table 6.2: Significant Test for CL Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....131
Table 6.3: HO Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....136
Table 6.4: Significant Test for HO Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....136
Table 6.5: HU Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....140
Table 6.6: Significant Test for HU Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....140
Table 6.7: PN Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....144
Table 6.8: Significant Test for PN Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....144
Table 6.9: CL Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....150
Table 6.10: Significant Test for CL Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....150
Table 6.11: HO Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....153
Table 6.12: Significant Test for HO Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....153

Table 6.13: HU Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....156
Table 6.14: Significant Test for HU Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....156
Table 6.15: PN Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....159
Table 6.16: Significant Test for PN Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....159
Table 6.17: CL Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....164
Table 6.18: Significant Test for CL Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....164
Table 6.19: HO Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....168
Table 6.20: Significant Test for HO Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....168
Table 6.21: HU Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....172
Table 6.22: Significant Test for HU Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....172
Table 6.23: PN Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....176
Table 6.24: Significant Test for PN Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....176
Table 6.25: CL Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....182
Table 6.26: Significant Test for CL Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....182
Table 6.27: HO Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....185
Table 6.28: HO Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....185
Table 6.29: HU Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....188
Table 6.30: Significant Test for HU Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....188

Table 6.31: PN Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....191
Table 6.32: Significant Test for PN Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....191
Table 6.33 Data Set C Price Prediction: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....196
Table 6.34: Significant Test for Data Set C Price Prediction: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....196
Table 6.35: Data Set C Price Prediction: *e*ABC-LSSVM vs. Other Techniques.....201
Table 6.36: Significant Test for Data Set C Price Prediction: *e*ABC-LSSVM vs. Other Techniques.....201

List of Figures

Figure 1.1: Classes of Commodities 4
Figure 2.1: Class of Evolutionary Computation.....24
Figure 2.2: Mapping of the Input Space to a High Dimensional Feature Space.....69
Figure 2.3: LPD with Different Value of α.....78
Figure 3.1: Methodology.....84
Figure 3.2: Representation of Food Source Position as Possible Solution in *X*.....94
Figure 3.3: Representation of Food Source Position as Possible Solution in *X*.....103
Figure 4.1: Flow of ABC-LSSVM in Initialization and EB Phase.....118
Figure 4.2: Flow of ABC-LSSVM in OB and SB Phase.....119
Algorithm 4.1: ABC-LSSVM algorithm.....120
Algorithm 5.1: *lv*ABC-LSSVM.....123
Algorithm 5.2: *cm*ABC-LSSVM.....124
Figure 5.1: Flow of *e*ABC-LSSVM in Initialization and EB Phase.....125
Figure 5.2: Flow of *e*ABC-LSSVM in OB and SB Phase.....126
Algorithm 5.1: *e*ABC-LSSVM algorithm.....127
Figure 6.1: Comparison of CL Price Predictions Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....133
Figure 6.2: Comparison of Convergence for CL Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....133
Figure 6.3: Exploitation of Search Space by *e*ABC-LSSVM in CL Price Prediction Data Set A.....134
Figure 6.4: Exploitation of Search Space by ABC-LSSVM in CL Price Prediction Data Set A.....134
Figure 6.5: Comparison of HO Price Predictions Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....137
Figure 6.6: Comparison of Convergence for HO Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....137
Figure 6.7: Exploitation of search space by *e*ABC-LSSVM in HO Price Prediction Data Set A.....138
Figure 6.8: Exploitation of search space by ABC-LSSVM in HO Price Prediction Data Set A.....138

Figure 6.9: Comparison of HU Price Predictions Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....141
Figure 6.10: Comparison of Convergence for HU Price Prediction Data Set A: *e*ABC-LSSVM vs. Variants ABC-LSSVM.....141
Figure 6.11: Exploitation of Search Space by *e*ABC-LSSVM in HU Price Prediction Data Set A.....142
Figure 6.12: Exploitation of Search Space by ABC-LSSVM in HU Price Prediction Data Set A.....142
Figure 6.13: Comparison of PN Price Predictions Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....145
Figure 6.14: Comparison of Convergence for PN Price Prediction Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....145
Figure 6.15: Exploitation of Search Space by *e*ABC-LSSVM in PN Price Prediction Data Set A.....146
Figure 6.16: Exploitation of Search Space by ABC-LSSVM in PN Price Prediction Data Set A.....146
Figure 6.17: Comparison of CL price predictions Data Set A: *e*ABC-LSSVM vs. Other Techniques.....151
Figure 6.18: Comparison of Convergence for CL Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....151
Figure 6.19: Comparison of HO price predictions Data Set A: *e*ABC-LSSVM vs. Other Techniques.....154
Figure 6.20: Comparison of Convergence for HO Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....154
Figure 6.21 Comparison of HU Price Predictions Data Set A: *e*ABC-LSSVM vs. Other Techniques.....157
Figure 6.22 Comparison of Convergence for HU Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....157
Figure 6.23: Comparison of PN Price Predictions Data Set A: *e*ABC-LSSVM vs. Other Techniques.....160
Figure 6.24: Comparison of convergence rate for PN Price Prediction Data Set A: *e*ABC-LSSVM vs. Other Techniques.....160
Figure 6.25: Comparison of CL Price Predictions Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....165

Figure 6.26: Comparison of Convergence for CL Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....165
Figure 6.27: Exploitation of Search Space by *e*ABC-LSSVM in CL Price Prediction Data Set B.....166
Figure 6.28: Exploitation of Search Space by ABC-LSSVM in CL Price Prediction Data Set B.....166
Figure 6.29: Comparison of HO Price Predictions Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....169
Figure 6.30: Comparison of Convergence for HO Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....169
Figure 6.31: Exploitation of Search Space by *e*ABC-LSSVM in HO Price Prediction Data Set B.....170
Figure 6.32 Exploitation of Search Space by ABC-LSSVM in HO Price Prediction Data Set B.....170
Figure 6.33: Comparison of HU Price Predictions Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....173
Figure 6.34: Comparison of Convergence for HU Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....173
Figure 6.35: Exploitation of Search Space by *e*ABC-LSSVM in HU Price Prediction Data Set B.....174
Figure 6.36: Exploitation of Search Space by ABC-LSSVM in HU Price Prediction Data Set B.....174
Figure 6.37: Comparison of PN Price Predictions Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....177
Figure 6.38 Comparison of Convergence Rate for PN Price Prediction Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....177
Figure 6.39: Exploitation of Search Space by *e*ABC-LSSVM in PN Price Prediction Data Set B.....178
Figure 6.40: Exploitation of Search Space by ABC-LSSVM in PN Price Prediction Data Set B.....178
Figure 6.41: Comparison of CL Price Predictions Data Set B: *e*ABC-LSSVM vs. Other Techniques.....183
Figure 6.42: Comparison of Convergence for CL Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....183

Figure 6.43: Comparison of HO Price Predictions Data Set B: *e*ABC-LSSVM vs. Other Techniques.....186
Figure 6.44: Comparison of Convergence HO Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....186
Figure 6.45: Comparison of HU Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....189
Figure 6.46: Comparison of Convergence for HU Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....189
Figure 6.47: Comparison of PN Price Predictions Data Set B: *e*ABC-LSSVM vs. Other Techniques.....192
Figure 6.48: Comparison of Convergence for PN Price Prediction Data Set B: *e*ABC-LSSVM vs. Other Techniques.....192
Figure 6.49: Comparison of Data Set C price prediction: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....197
Figure 6.50: Comparison of convergence for Data Set C Price Prediction: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants.....197
Figure 6.51: Exploitation of Search Space by *e*ABC-LSSVM in Data Set C Price Prediction 198
Figure 6.52: Exploitation of search space by ABC-LSSVM in Data Set C Price Prediction 198
Figure 6.53: Comparison of Data Set C price predictions: *e*ABC-LSSVM vs. Other Techniques.....202
Figure 6.54: Comparison of Convergence for Data Set C Price Prediction: *e*ABC-LSSVM vs. Other Techniques.....202

List of Appendices

Appendix A Correlation Among Input.....225
Appendix B BDS Test Results.....228
Appendix C Testing Prediction Results.....229
Appendix D Sample of Raw Data.....233
Appendix E Publications during the Doctorate Study.....234

List of Abbreviations

ABC	-	Artificial Bee Colony Algorithm
ACO	-	Ant Colony Optimization
AFSA	-	Artificial Fish Swarm Algorithm
AI	-	Artificial Intelligent
ANFIS	-	Adaptive Neuro Fuzzy Inference
ANN	-	Artificial Neural Network
ARIMA	-	Autoregressive Integrated Moving Average
ARE	-	Average Relative Error
BCC	-	Bacteria Colony Chemotaxis
BPNN	-	Back Propagation Neural Network
CI	-	Computational Intelligence
CV	-	Cross Validation
DE	-	Differential Evolution
EB	-	Employed Bee
EC	-	Evolutionary Computation
EMD	-	Empirical Mode Decomposition
EP	-	Evolutionary Programming
ERM	-	Empirical Risk Minimization
GA	-	Genetic Algorithm
GDP	-	Gross Domestic Product
GMSE	-	Generalization Mean Square Error
GSA	-	Gravitational Search Algorithm
IMF	-	Intrinsic Mode Function
LMSE	-	Learning Mean Square Error
LSSVM	-	Least Squares Support Vector Machines
MAE	-	Mean Absolute Error
MAPE	-	Mean Absoluter Percentage Error
MATLAB	-	Matrix Laboratory Software
MLP	-	Multilayer Perceptron

MRE	-	Mean Relative Error
MSE	-	Mean Square Error
NLL	-	Negative Log Likehood
OB	-	Onlooker Bee
OECD	-	Organization for Economic Co-operation and Development
PCA	-	Principal Component Analysis
PNN	-	Probabilistic Neural Network
PSO	-	Particle Swarm Optimization
QP	-	Quadratic Programming
RBF	-	Radial Basis Function
RE	-	Relative Error
RMSE	-	Root Mean Square Error
SB	-	Scout Bee
SOM	-	Self Organizing Maps
SRM	-	Structural Risk Minimization
SVM	-	Support Vector Machines
SVR	-	Support Vector Regression
WTI	-	West Texas Intermediate

CHAPTER ONE
INTRODUCTION

1.1 Background Study

Since its emergence in the past decades, Least Squares Support Vector Machines (LSSVM) (Suykens, Van Gestel, De Brabanter, De Moor, & Vandewalle, 2002)which is an extension of Support Vector Machines (SVM) (Vapnik,1995) has contributed significant impact in machine learning community. As a powerful algorithm, it has been recognized as one of the standard tools in solving various data mining tasks which include prediction, classification and many others (Cheng, Guo, & Wu, 2010, Li, 2009).Nonetheless, besides its diversity in application, it is worth noting that the capability of LSSVM is highly dependent on the value of its hyper-parameters, namely regularization parameter, γ and kernel parameter, σ^2 (Jiang & Zhao, 2013).

In this regard, this study proposes a hybridization of LSSVM with a relatively new optimization algorithm namely Artificial Bee Colony (ABC) (Karaboga, 2005). The ABC algorithm which has been introduced by Dervis Karaboga is enlightened from the intelligent foraging behavior of honey bees swarm (Karaboga, 2005). In 2008, an extensive review and comparative analysis regarding its performance efficiency was examined which concluded that the ABC algorithm is comparable to the other existing optimization algorithms including Differential Evolution (DE), Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) (El-Abd, 2012; Karaboga & Basturk, 2008).

As to address the limitation of standard ABC and increase the capability of LSSVM in prediction task as well, an enhanced ABC, later known as *e*ABC, is proposed. The motivation of enhancement is due to the demerits owned in standard ABC which is expose with local minimum trapping (Gao & Liu, 2012; Lee & Cai, 2011). Besides, the ABC also tends to fall into boundary values (Karaboga & Akay, 2009) which will cause LSSVM to face with over fitting and under fitting situation (Fu, Liu, & Sun, 2010; Lendasse, Ji, Reyhani, & Verleysen, 2005; Wu, Feng, & He, 2007). This situation is discussed in Section 1.3 and the detail of enhancement is described in Chapter 3. Later, the proposed *e*ABC-LSSVM is realized in time series prediction of non-renewable natural resources commodities price.

Non-renewable natural resources commodities price is chosen due to its important role in human life, specifically energy fuels. This is proven with the avalanche studies in literature which witnessed the emerging of various prediction techniques, covering from statistical to Computational Intelligence (CI) techniques (Labys, 2006; Xiao, He & Wang, 2012). Nevertheless, despite the considerable amount of studies, there are still some scarceness and gaps that can be filled in as a remedy to treat the limitation, which is discussed in section 1.2. As been concluded by Makridakis and Hibon (2000), there is no prediction technique that is universally good, but rather each technique is suitable and exclusively applied for specific situation. Thus, finding an appropriate approach for specific data set is important in order to obtain better prediction accuracy. A detail review on related prediction techniques are presented in Chapter 2, while the substance is featured in this chapter.

1.2 Time Series Prediction

Time series prediction is of interest in various fields. It is important due to prediction of future values could be a vital input for current planning and decision making (Montgomery, Jennings, & Kulachi, 2008). To realize the time series prediction, historical data is employed to generate future values, i.e. to make prediction (Ismail, Shabri, & Samsudin, 2011). Due to the importance of time series prediction, numerous approaches have been presented, and this includes in non renewable natural resources commodities price prediction (Khasman & Nwulu, 2011, Khazem, 2008, Varahrami, 2011). In literature, the prediction model for such commodities price are presented in various horizon or time scales; whether in short term, medium term or long term (Alizadeh, Moghaddam, Khakzad & Ebrahimipour, 2012). Thus, different data frequency has been employed depending on the suitability viz. daily, weekly, monthly or annually (Bao, Zhang, Yu, Lai & Wang, 2011; Malliaris & Malliaris, 2008, Yusof, A. Rashid, & Mohamed, 2010).

Prediction of non-renewable natural resources commodities price is considered as time series prediction due to utilization of historical data in prediction task which is essential in understanding and extrapolating the future (Ismail, et al., 2011).As for commodity, it is defined as raw materials, hard assets, and visible goods that underpin civilization in almost every single way of human life (Frush, 2008). In the international market, basically there are six major classes of commodities namely energy fuels, metals, agricultures, livestock, exotics and financials (see Figure 1.1.) (Frush, 2008). Among them, energy fuels and metals are classified into non-

renewable natural resources commodities. Both commodities have been utilized since ancient time till present age. For instance, energy fuels, such as crude oil is used to support human lives where it influences not only to the industries but also political relationship among countries (Khazem, 2008; Yuan, Zhuang, Liu, & Huang, 2014). On the other hand, metals, such as gold and silver have primarily served as a medium of exchange before the utilization of fiat money (Khalifa, Hong, & Ramchander, 2010; Wang & Lee, 2010).

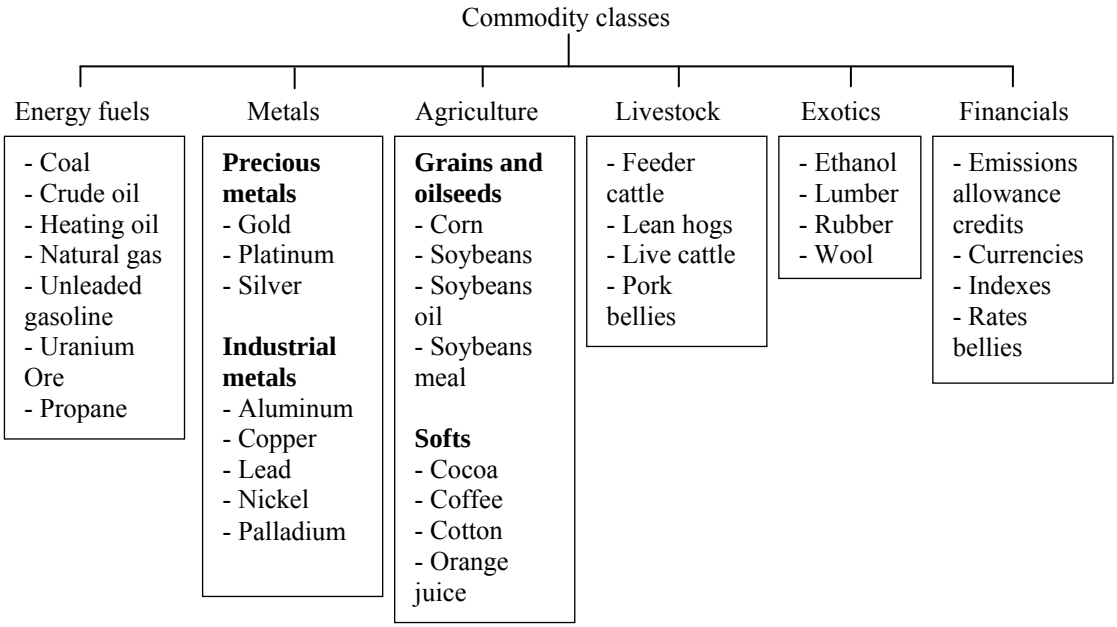


Figure 1.1: Classes of Commodities. Adapted from Commodities Demystified (p. 5), by S. Frush, 2008, New York, NY: McGraw-Hill. Copyright 2008 by the McGraw-Hill. Adapted with permission.

Unfortunately, these crucial resources are very limited in production and irreplaceable (Kemp 2004). With the limitation in resources and continuously increasing demand, this situation leads to only one result; higher prices. As for

investors, this means opportunity, however, the public, this indicates inflation (Zhang, Wu, & Zhang, 2010). Due to that matter, the importance of price prediction for such data has resulted to a large growing body of literature and research among the community is continuously carried out (Bao, et al., 2011;Jammazi & Aloui, 2012).

When energy fuel issue is highlighted, it is inevitable that crude oil is more dominant that other materials of the same class (Jammazi & Aloui, 2012; Kulkarni & Haidar, 2009). It is identified as the largest and most actively traded commodity in the world (Hammoudeh, Sari, & Ewing, 2008; Radetzki, 2008) and can be refined into varying amount of heating oil, gasoline and propane (Malliaris & Malliaris, 2008). As a major kind of energy fuel resource, the dependent of crude oil involved many sectors of global economy (Khazem, 2008) and the fluctuation of crude oil price may contribute significant impact not only to consumer goods but also industrial sectors (Jammazi & Aloui, 2012). Even though there is a claim that the period of crude oil is nearly to be over; however, it is revealed that the global demand of crude oil will continue to grow in the long run. Despite a drop in demand from Organization for Economic Co-operation and Development (OECD) countries, there is positive demand of crude oil from non OECD countries such as China and India (Kulkarni & Haidar, 2009; U.S. Energy Information Administration [EIA], 2013).

As a refinement product of crude oil, heating oil also play its role to the community. It represents 25% of the production of a typical barrel, which make it serves as second biggest component of crude oil (Frush, 2008). It can be used as a substitute

for natural gas in power generation (Dunsby, Eckstein, Gaspar, & Mulholland, 2008). Meanwhile, gasoline which is primarily used for internal combustion engines is the main product produced from refining of crude oil. For each barrel of crude oil is refined; it produces about 3.05 billion barrels of finished motor gasoline (Dunsby, et al., 2008). As for propane, it is beneficial for home heating and also as an alternative fuel for vehicles (Fontanills, 2007). Since heating oil, gasoline and propane are produced by the same raw materials, it is not surprising when the price correlation among them and crude oil are found to be positive (see Table 1.1).

Table 1.1: Correlation among Energy Fuels Price from 22 December 1992 – 30 November 2007

	Crude oil	**Heating oil**	**Gasoline**	**Propane**
Crude oil	1	-	-	-
Heating oil	0.9831	1	-	-
Gasoline	0.9927	0.9804	1	-
Propane	0.9746	0.9668	0.9729	1

From literature, it is observed that the uncertainty pattern of rise and fall of energy fuels price reflects the importance of development of price prediction technique. The obvious high nonlinear pattern of crude oil makes the task in price prediction regarded as challenging (Jammazi & Aloui, 2012; Khashman & Nwulu, 2011; Khazem, 2008). Similar situation is expected for heating oil, gasoline and pronane due to high price correlation among them (see Table 1.1).

The limitation of statistical techniques in prediction of non renewable natural resources commodities price such as Autoregressive Integrated Moving Average (ARIMA) (Dooley & Lenihan, 2005; Yusof, et al., 2010) has been highlighted in numerous literatures (Khazem, 2008; Kumar & Thenmozi, 2007; Liu, 2009; Yu, Wang, Wen, & Lai, 2008). Besides ARIMA, other statistical technique such Exponential Smoothing (Rao, Shu, Xiao, 2011) is also not favorable in prediction of the said time series data. This is due to the linear structure of the techniques which make them incapable to capture the nonlinear patterns in the energy fuels price time series data (Li, Masuda, & Nagai, 2013). Meanwhile, the Classical Component Model (Chang, 1992) lack of capability in considering the other factors that would influenced the prediction model (Chang, 1992) which made it unsuitable for the complex nature of energy fuels price (Jammazi & Aloui, 2012; Khashman & Nwulu, 2011; Khazem, 2008). The limitation of these techniques in dealing with nonlinear patterns in data has led researchers to suggest a CI based predictor tool, which is Artificial Neural Network (ANN) (Shadbolt & Taylor, 2002). By applying ANN, the prediction model is trained based on example based learning approach (Kulkarni & Haidar, 2009; Zhang, Patuwo, & Hu, 1998).

However, even though ANN has enjoyed considerable success (Abdullah & Zeng, 2010; Kulkarni & Haidar, 2009; Khazem, 2008),the shortcomings of this approach could not be concealed. The Empirical Risk Minimization (ERM) that is adopted by ANN intends to only minimize the training error, thus making ANN exposed to over fitting and consequently led to poor generalization (Cheng, Qian, & Guo, 2006;

Xiang & Jiang, 2009). In addition, the requirement of many control parameters to be tuned make ANN more complicated to be applied (Xiang & Jiang, 2009; Zhang, et al., 1998).

Recently, the application of SVM (Vapnik, 1995) has been proven to overcome the demerits of ANN. By applying Structural Risk Minimization (SRM), the principle tends to minimize the generalization error, i.e. true error on new data rather than the training error as applied in ERM (Gencoglu & Uyar, 2009; Elish & Elish, 2008). A variant of SVM, the LSSVM, (Suykens, et al., 2002) carries a great potential in solving practical problems in various fields including pattern recognition, prediction, classification and many more (Cheng, et al., 2010). However, as highlighted previously (see section 1.1), despite its advantages, LSSVM also comes with trade off where the generalization capability of LSSVM is dependent on the value of its parameters, namely regularization parameter, γ and kernel parameter, σ^2 which are commonly referred to as hyper-parameters (Jiang & Zhao, 2013). If both values of hyper-parameters are not well chosen, the produced results will be inaccurate.

With respect to that matter, obtaining ideal values for the hyper-parameters is not an easy task since it has a huge number of combinations. Thus, manual selection approach may not be an appropriate technique since it may lead to the selection of random value. Beside producing unreliable results (Dos Santos, Luvizotto, Mariani, & Dos Santos Coelho, 2012), it is time consuming and unsystematic (Chen, Wu, & Chen, 2008; Cheng, et al., 2010).

Generally, there are two major approaches in optimizing the LSSVM hyper-parameters that is experimental technique and meta heuristic technique (Afshin, 2007; Afshin, Sadeghian, & Raahemifar, 2007). For the first technique, in practice, Cross Validation (CV) is the most utilized. However, from literature, it is observed that, in terms of error rate, the result obtained tend to be unsatisfying (Yun, Chen, Wang, & Lai, 2009). Besides, it is also time consuming (Lendasse, et al., 2005). This leads to the hybridization of LSSVM with meta heuristic techniques which include Evolutionary Computation (EC) algorithm, such as GA (Mustafa, Sulaiman, Shareef & Khalid, 2012), PSO (Jiang & Zhao, 2013) and many others. With such hybridization, the hyper-parameters of interest are efficiently optimized. Hence, the generalization of LSSVM can be enhanced.

Nevertheless, the shortcomings of GA such as slow in convergence speed, suffer in local minimum and complexity in determining the control parameters could not be ignored since this would affect the generalization of LSSVM (Chen, Wang, Sun, & Liang, 2008). Similar situation also faced by PSO where besides the tendency to fall into local minimum (Park, Jeong, Shin, & Lee, 2010), it also raises difficulty in parameter selection due to several control parameters embedded in the algorithm (Chen, et al., 2008).

As a relatively new Swarm Intelligence (SI) algorithm, ABC comes with an advantage of employing fewer control parameters which is potentially more generic to adopt to a wider class of optimization (Babayigit & Ozdemir, 2012; Gao & Liu, 2012). The two parameters involved are the ones exist in any population-based

optimization techniques, namely population size and maximum number of iteration. Thus, with the advantages and modest requirement offered in the ABC, it has stimulated this study to propose the hybridization of ABC, which is enhanced with LSSVM, termed as *e*ABC-LSSVM for non-renewable natural resources commodity price prediction.

1.3 Problem Statement

For the sake of high accuracy, choosing a suitable prediction technique for time series data of interest requires careful attention. The incapability of linear statistical techniques in prediction of non linear time series data (Dooley & Lenihan, 2005; Yusof, et al., 2010)has led researchers to explore the application of CI technique such as the ANN. However, the adaptation of ERM principle make the ANN vulnerable with over fitting which consequently make it suffer generalization (Malliaris & Malliaris, 2008;Cheng, et al., 2006; Xiang & Jiang, 2009).In addition, the difficulty in tuning the control parameters worsens the situation (Xiang & Jiang, 2009; Zhang, et al., 1998).

The LSSVM is a good candidate to overcome the said limitations of ANN (Bao, et al., 2011; Jiang & Zhao, 2013; Xie, Li, Lu, & Xie, 2009). Nonetheless, to obtain the optimum value of LSSVM hyper-parameters, namely regularization parameter, γ and kernel parameter, σ^2 would be challenging since it has huge numbers of combinations. Hence, trial and error approach is not an appropriate technique (Dos Santos, et al., 2012). Meanwhile, the application of CV suffers in producing low

error rate and also inefficient in terms of computational time (Lendasse, et al, 2005; Yun, et al., 2009).

Recently, the LSSVM is being utilized along with EC algorithm. The hybridization of these approaches appears to be encouraging and capable to remedy the scarceness of single technique. This consequently leads to a better generalization, such as the hybridization of GA-LSSVM (Mustafa, et al., 2012) and the PSO-LSSVM (Jiang & Zhao, 2013). However, the demerits of GA-LSSVM viz. slow in convergence speed, trapped in local minimum and the choice of the embedded control parameters (Chen, et al., 2008) would demote the prediction performance. Quite similar situation is also faced in PSO-LSSVM where besides facing with the difficulty in parameters selection (Chen, et al., 2008), it is also exposed to fall into local minimum (Park, et al., 2010). Thus, the drawbacks of these techniques could not be neglected as it would affect the performance of LSSVM in prediction.

With respect to that matter, this study proposed the use of ABC to optimize the LSSVM hyper-parameters. However, despite the advantages offered in ABC, several inherent drawbacks of this technique need to be addressed. From literature (Babayigit & Ozdemir, 2012; Sharma, Pant, & Bhardwaj, 2011; Gao, Liu, & Huang, 2013), it is observed that the searching solution of ABC is good in exploration but poor in exploitation. This is due to the ABC dependency on a single equation for different phase of exploitation process. This consequently limits the searching behavior of the algorithm and leads the algorithm to fall into local minimum (Gao & Liu, 2012; Lee & Cai, 2011). The tendency of the ABC algorithm to fall into

boundary values (Karaboga & Akay, 2009) also requires attention as this would lead the LSSVM prediction technique vulnerable to over fitting and under fitting problem (Fu, Liu, & Sun, 2010;Lendasse, et al., 2005; Wu, et al., 2007).

The gaps that exist in highlighted prior works (see section 1.2) have paved the way for this study to propose the *e*ABC-LSSVM which addressed the said limitations. Firstly, as to address the tendency of the ABC algorithm from falling into local minimum, the searching behavior is enriched. This is done by inducing Levy mutation and introducing different strategies as to enrich and balance the performance between different phases in exploitation process. This is later known as *lv*ABC. On the other hand, as to prevent the algorithm from facing with over fitting and under fitting issues, conventional mutation is introduced (termed as *cm*ABC). The details on the *lv*ABC and *cm*ABC are described in Chapter 3. The potential that owned by the proposed approach would be beneficial in the field of interest.

1.4 Research Question

The research questions for this study are as follow:

1. How to automatically and efficiently optimize the hyper parameters of LSSVM?
2. How to address the limitation of searching behavior of the artificial bees in ABC algorithm?
3. How to prevent the LSSVM from facing with over fitting and under fitting problem?
4. How to improve the generalization in prediction and escape from premature convergence?

1.5 Objectives of Study

The goal of this study is to develop a price prediction algorithm for non renewable natural resources commodities price. In order to achieve the goal, the study is further divided into four objectives as follows:

i. To optimize LSSVM hyper parameters using ABC algorithm.

ii. To formulate a new solution generation functions to overcome the limitation of exploitation process of the artificial bees by using Levy mutation and different strategies.

iii. To propose a new decision making function for the artificial bees that prevents over fitting and under fitting problem.

iv. To construct *e*ABC-LSSVM algorithm that includes (ii) and (iii) which would overcome the generalization problem and escape from premature convergence.

1.6 Limitation of Study

In this study, several limitations are defined. Firstly, the performance of the proposed algorithm are evaluated based on error rates and accuracy without taken into account the computational time. Secondly, for the convergence, it only consider the convergence value, but not the speed. Lastly, the experiment are limited for daily time series data for non renewable natural resources commodities price.

1.7 Scope of Study

In this study, the employment of empirical data is based on existing research (Malliaris & Malliaris, 2008) and the variables are not influenced by other economic factors. The study involved using three sets of time series data which includes:

i. Energy fuels data set A (Data Set A)

ii. Energy fuels data set B (Data Set B)

iii. Metals data set (Data Set C)

Data Set A and B are to predict crude oil, heating oil, gasoline and propane price. Meanwhile Data Set C is to predict the gold price. Gold price is chosen due to similarity of its feature with energy fuels, viz. both are categorized as non-renewable natural resources (Kemp, 2004). This makes it possess fixed supplied but continuously increasing demand.

Daily time series data is utilized and other economic factor such as industrial production, inflation, Gross Domestic Product (GDP), and oil supply and demand inventory are excluded as they are only available on monthly or quarterly (Hammoudeh, et al., 2008; Kulkarni & Haidar, 2009).

This study emphasized on hybrid methods of EC with machine learning in prediction task, specifically in parameter tuning.

1.8 Significance of Study

This study is conducted to overcome the issue of optimizing the LSSVM hyper parameters by using ABC algorithm which is realized in energy fuels price and gold price prediction. Besides several advantages of original ABC and LSSVM embodied in the technique, the enhancements that have been introduced making it even better. Firstly, the action taken in enriching the searching behavior of the agents (i.e. bees) drives the prediction technique to avoid local minimum, hence global optimum can be achieved. On the other hand, an enhancement that has been introduced in decision making process assists the prediction technique from facing with over fitting and under fitting. With the combinations of all inherited advantages and features that are newly introduced, the *e*ABC-LSSVM is not only capable to produce better generalization, but also escape from premature convergence. With such advantages, the proposed *e*ABC-LSSVM will be beneficial in optimization and machine learning community, specifically in prediction task. In real application, the capability of *e*ABC-LSSVM in predicting the change of trend would be useful for investors and commodities traders in planning their investment and projecting their profit.

1.9 Thesis Outline

This thesis is structured as follows. Chapter 2 presents the literature reviews that are related to the optimization and prediction techniques, particularly the ones utilizing LSSVM optimized by EC algorithms.

In Chapter 3, the implemented methodology is described. This chapter incorporated of four main sections, namely data collection, designing the proposed algorithm, developing the proposed algorithm and finally evaluation. This is followed by the algorithm of ABC-LSSVM and *e*ABC-LSSVM which are presented in Chapter 4 and 5 respectively

Chapter 6 discuss and analyze the performance of *e*ABC-LSSVM and other benchmarking techniques in predicting the price of Data Set A, B and C. The benchmarking techniques include both hybrid and single techniques. Finally, Chapter 7 concludes this study and some recommendations for future works in the context of study are proposed.

CHAPTER TWO
LITERATURE REVIEW

2.1 Methods to Optimize the LSSVM Hyper Parameters

As highlighted in the previous chapter, the performance of LSSVM is highly dependent on the selection of its hyper parameters. With respect to that matter, this section reviews the existing work that employs cross validation procedure in optimization of LSSVM hyper parameters. It is follows by the discussion on the optimization of LSSVM using Evolutionary Computation (EC) algorithms, which includes Evolutionary Algorithm (EA) and Swarm Intelligence (SI) techniques.

2.1.1 Optimization of LSSVM using Cross Validation

Wang and Li (2010) highlight the importance of primary energy consumption prediction by presenting a prediction model utilizing LSSVM. Several factors that were chosen to be fed to the prediction models include industrial index and population. Empirical data sets were from 1990 to 2008, which was limited for the case in China. For evaluation purposes, two performance criteria were used namely Absolute Percentage Error (APE) and MAPE. For parameter tuning, CV technique was employed. By using CV, each training set, TR is divided into X partitions in equal size (TR_1, TR_2, …,TR_X). For each run, the LSSVM is trained X times and one of X is set as testing set. Later, the trained model is used in the testing set. Finally, by using the obtained parameters value, the average accuracy of X trials is calculate to estimate the final accuracy. The CV-LSSVM technique achieved the predetermined condition where the accepted error range is within 5%. In addition, as the researchers

employed yearly samples, it shows that LSSVM is capable to deal with small samples data sets as compared to ANN which frequently needs more data for training purposes (Haidar & Wolf, 2011; Zhang, et al., 1998).

In 2008, Liu and Wang presented a prediction technique that is based on LSSVM to predict foreign exchange rate market (Liu & Wang, 2008). In order to determine the optimized hyper-parameters of LSSVM, CV approach was applied. For experimental purposes, four pairs of exchange rates time series data were employed. The samples comprise of time series data from 2003-2007, which consist of 1196 observations. By using 86:14 data proportion for training and testing respectively, the efficiency of proposed technique is evaluated using RMSE, MAE, and MAPE. However, even though the CV-LSSVM is capable to provide promising results, the absence of comparison with other comparable prediction techniques made the experiment incomplete.

In related work, an EMD was integrated with LSSVM in dealing with exchange rate prediction (Lin, Chiu, and Lin, 2012). EMD was utilized to decompose the foreign exchange rates into several IMF components and one residual component. The LSSVM then performs the prediction of the IMFs and residual value independently. Comparing the result against those produced by EMD-ARIMA, single LSSVM and single ARIMA, the EMD-LSSVM outperforms other experimented prediction techniques based on several criteria, which includes MAE, MAPE, Correct Uptrend (CP) and Correct Downtrend (CD). For parameter tuning, the CV approach was employed.

Progressing further, in Mellit, Massi Pavan, and Benghanem (2013), short term prediction of meteorological time series data has been presented using CV-LSSVM. Prior to training task, three pre-processing techniques were performed namely data filtering which is to remove redundant components from the data sets, removing outliers from the data sets and finally data normalization. The evaluation of the presented technique is based on several statistical metrics which includes MAPE and RMSE. Comparison againts various ANN based predictor showed that the performance of CV-LSSVM is at par with Probabilistic Neural Network (PNN).

Meanwhile, Wu and Niu (2009) employed LSSVM to predict short-term power load in Inner Mongolia. The case study used one year load series from that region, and later divided them into 90:10 for training and testing respectively. For parameter tuning, CV technique was utilized. The results derived from this approach is regarded as efficient for the problem under study. However, no comparison with other technique was conducted.

Wang and Li (2009) integrated the EMD with LSSVM in predicting future wind speed and output power. The purpose of applying EMD is similar as described in Lin, et al. (2012) which is to decompose the data sets utilized into a series of different scales of IMF. By employing daily average wind speed from March to May 2006, the result obtained by the proposed model is compared against several techniques which include single LSSVM. The hyper-parameters of LSSVM are tuned based on CV approach. The comparison was made by taking into account the value produced in terms of MAPE and RMSE by each identified prediction model.

Final results indicated that the EMD-LSSVM outperformed the other experimented techniques by producing lower MAPE and RMSE.

In related work, Wang, Wang, Xie and Zhu (2010) demonstrated a CV-LSSVM for air temperature prediction. This study utilized monthly average temperature from 1951 to 2003. From that sample, only year 2003 data is allocated for testing set. The findings of the experiment conducted were evaluated using RMSE, Relative Error (RE) and prediction accuracy and involved another two prediction techniques, namely single LSSVM and RBFNN. Final results indicated that RMSE produced by EMD-LSSVM was 0.7202, which was the smallest among the experimented prediction techniques.

Meanwhile, in Gencoglu and Uyar (2009), the CV-LSSVM was presented in prediction of flashover voltage of insulators. In the study, several comparisons were performed between different types of kernels, which include RBF kernel and Polynomial kernel. The results indicated that RBF kernel yield lower RMSE. Meanwhile, the comparison against ANN is in favor to CV-LSSVM.

Previously, the application of LSSVM in a Bayesian evidence framework was examined by Van Gestel, et al (2001). In the presented study, two different cases were investigated, namely prediction for U.S. short term interest rate and daily index closing price return. Prior to training, normalization technique was performed to all data sets. Upon completing the experiments, the development of proposed model is proven to be successful as an improvement is made in MSE, MAE and Negative Log

Likehood (NLL) as compared to traditional methods, which includes autoregressive model.

In Ahmad, Hassan, and Majid (2012), a hybridization of GMDH with LSSVM was presented to predict electrical energy consumption. GMDH is the ANN based model which finds the optimal model based on inductive sorting out approach. For LSSVM parameter selection, CV technique was utilized. When evaluated based on MAPE, the presented technique offers lower MAPE as compared to the comparison algorithms, which includes single GMDH and LSSVM.

As discussed above, even the utilized CV has provided promising results for the context of study, however, without comparing CV with other techniques in the study conducted (Liu & Wang, 2008; Wang & Li, 2010; Wu & Niu, 2009), proving the efficiency of the presented method is therefore difficult. In addition, CV is revealed to be time consuming (Lendasse, et al., 2005) and tends to produce high error rates (Yu, et al., 2009). Due to that matter, an application of EC algorithm, which includes EA and SI techniques, become research interest among academic community in optimizing LSSVM hyper-parameters which is presented in the next section (see section 2.1.2). A summary of this section is tabulated in Table 2.1.

Table 2.1: Summaries of Time Series Prediction Technique using LSSVM Optimized by CV

Author	Research Area	Independent Variable	Method	Data Frequency	Evaluation Metric
Wang & Li (2010)	Primary energy consumption prediction	Gross Domestic Product (GDP), population and others	CV-LSSVM	Annually	APE and MAPE
Liu & Wang (2008)	Exchange rates prediction	JPY/USD, CAD/USD, USD/GBP and USD/EUR	CV-LSSVM	Daily	MAE, MAPE AND RMSE
Lin, et al.(2012)	Exchange rates prediction	USD/NTD, JPY/NTD, and RMB/NTD	CV-EMD-LSSVM	Daily	MAE, MAPE, CP, CD
Mellit et al.(2013)	Short-term meteorological time series data prediction	Solar irradiance, air temperature, relative humidity and pressure	CV-LSSVM	Daily	R^2, RMSE, Mean Bias Error (MBE), MAPE
Wu & Niu (2009)	Short-term load prediction	Load series in Inner Mongolia	CV-LSSVM	Daily	MSE
Wang & Li (2009)	Wind speed and output power prediction	Average wind speed	EMD-LSSVM	Daily	MAPE and RMSE
Wang, et al.(2010)	Air temperature prediction	Average temperature	EMD-LSSVM	Monthly	RMSE, RE
Gencoglu &Uyar	Flashover voltage of insulators prediction	Insulator height, insulator diameter, and several others	CV-LSSVM	-	RMSE, R^2, Mean Error Function (MFE)

(2009)					
Van Gestel, et al. (2001)	Financial time series prediction	Interest rate, closing index prices	Weekly	CV-LSSVM	MSE, NLL and MAE
Ahmad, et al. (2012)	Electrical energy consumption prediction	Electrical energy consumption time series data	-	CV-GMDH-LSSVM	MAPE

2.1.2 Optimization of LSSVM using Evolutionary Computation Algorithm

The Evolutionary Computation (EC) is comprised of Evolutionary Algorithm (EA) and SI technique (Karaboga, et al., 2012). The EA is inspired by the natural evolution phenomenon while the imitation of insects or animals is featured in SI technique (Karaboga & Akay, 2009). Figure 2.1 illustrates the EC class.

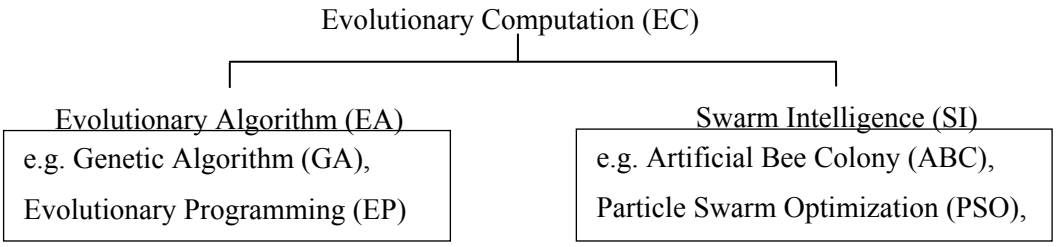


Figure 2.1: Class of Evolutionary Computation

Categorized under EA group, GA is the most prominent and widely used technique as compared with other techniques within the same domain (Karaboga & Akay, 2009, Yang, 2010). It was introduced by John Holland and later was popularized by David Goldberg (Haupt & Haupt, 2004). It models genetic selection and imitate the natural evolution phenomenon to solve numerical optimization problems. The GA consists of five main components viz. random number generator, fitness evaluation, reproduction, crossover and mutation operations. The last three components are categorized as genetic operator. In GA, the possible solutions are represented by chromosomes.

The hybrid technique of GA-LSSVM has been demonstrated to predict transmission loss in deregulated power system (Mustafa, Sulaiman, Shareef, & Khalid, 2011). In

the study, the application of GA-LSSVM is used to learn which generator is responsible for the losses by determining the optimal value for LSSVM hyper-parameters. The data was broken down into three sets; 57% of samples are used for training, 14% are set for validation and the rest 29% is reserved for testing. Simulations were performed and the results obtained showed satisfactory result. Similar approach also has been tested by the authors to predict reactive power which is compared against result produced by ANN (Mustafa, et al., 2012). The results obtained suggested that both methods produced similar outputs; nonetheless GA-LSSVM possesses an advantage in terms of computational time.

Yu, et al., (2009) presented the GA-LSSVM for stock market prediction. In this study, the first GA was utilized for feature selection while the second GA is for parameter tuning. The proposed model is compared against the results produced by several techniques which include conventional BPNN, ARIMA, SVM and variants LSSVM. From the experiment conducted, the GA-LSSVM is proven to be able to present some selection models that are easier to clarify by applying small number of features. However, for certain cases, the presented model was unable to outperform the other comparison algorithms.

In the maritime field, ship motion prediction by utilizing Adaptive GA (AGA) with LSSVM has been presented by Fu, et al., (2010). In the study, by referring to the fitness values of individuals, the probability of crossover and mutation are adaptively adjusted. To prove the efficiency of the AGA-LSSVM, comparison with the BPNN

and LSSVM were made. From the results obtained, it clearly shows that AGA-LSSVM excels by yielding lower MAE as compared to the experimented techniques. However, it would be appropriate if the comparison against GA-LSSVM is considered as to see the difference between the newly GA with the original one.

Meanwhile, Sun and Zhang (2008) examined the capability of GA-LSSVM in power spot prices prediction. Prior to training, the data was pre-processed by applying Min Max Normalization. The performance of the model was measured based on MAE and MAPE. After completing experiment process, the average of MAE and MAPE produced by GA-LSSVM yielded lower value as compared to value produced by ANN.

In Xie, et al., (2009), the comparison between the GA-LSSVM, standard SVM and LSSVM to predict dissolved gas concentration was performed. The findings of the study indicated that the GA-LSSVM provided better results than the other two approaches.

In Yang, Gu, Liang and Ling (2010), the GA-LSSVM was proposed for conductivity and tensile strength prediction for carbon fiber composites. In this study, the predictability of proposed model was guide based on two indices, namely Learning Mean Square Error (LMSE) and Generalization Mean Square Error (GMSE). To further assess the superiority of GA-LSSVM, comparisons against Principal Component Analysis-GABPNN (PCS-GABPNN) was made. Based on the results

obtained, the GA-LSSVM is capable to produce better generalization performance without having to use any data pre-treatment as required in the identified competitor. In Zou (2009), the application of GA-LSSVM was proposed to predict the condition of hydro turbine generating units (HGUs). Good prediction model for the context study is beneficial in providing backups to predictive maintenance system in mechanical equipment fault diagnosis. As compared to CV-LSSVM and BPNN, GA-LSSVM performed slightly better than CV-LSSVM while BPNN performed dismally due to over fitting problem.

In 2005, LSSVM optimized by GA has been used to deal with prediction of chaotic time series data (Wang, et al., 2005). Tested on benchmark time series data, both training and testing set were divided equally. The comparison against CV-LSSVM suggested that GA-LSSVM is capable to produced lower error rate, relative to the RMSE.

Further development of GA-LSSVM has been presented by Liao and Balzen (2013) in solving network flow prediction. In the study, the modification is put forward to GA, where Self-adaptive GA is introduced. The modification involved the standard operation of crossover and mutation. Using hourly network access flow, the data sets were first normalized using Min Max Normalization. The conducted experiment indicated that Self-adaptive GA-LSSVM is superior over the other identified approach. Nevertheless, the comparison with the standard GA-LSSVM was absent in the study.

Introduced by Kennedy and Eberhart in 1995 (Kennedy & Eberhart, 1995), PSO is considered as a well-known SI technique which was inspired by social behavior of birds flocking or fish schooling. The application of PSO can be seen broadly utilized in solving various optimization issues (Karaboga & Akay, 2009). In this technique, it is initialized with a group of random particles which represent possible solutions, and then searches for global optimal by updating generations. As compared to GA, there is no evolution operator such as crossover and mutation in PSO, which make it simpler in implementation. However, the choice in parameter selection such as acceleration constant and weight still need to be considered (Chen, et al., 2008).

The efficiency of PSO-LSSVM has been demonstrated by Jiang and Zhao (2013) in critical heat flux prediction (CHF). Applied in nuclear science field, accurate prediction of CHF is vital not only for safety but also in designing nuclear reactors. For that matter, several relevant inputs are considered for training and later, the samples of data sets are divided into training and testing. The performance of the proposed method is guided based on three indices, viz. the R^2, Mean Relative Error (MRE) and RMSE. Compared against ANFIS, the result obtained is in favor to the PSO-LSSVM.

Prediction for water quality utilizing hybrid of PSO with LSSVM was studied by Xiang and Jiang (2009). In this study, 3256 data points of water figures from Lixiu River water database were employed as empirical data. The results obtained by the

presented technique are compared against the results produced by using CV-LSSVM, BPNN, ARIMA and GML. The incapability of CV in this study supports the results obtained in Yang (2009) and confirms the fact stated in previous the studies (Lendasse, et al. 2005; Yu, et al., 2009). However, while the PSO-LSSVM outperforms CV-LSSVM, BPNN and ARIMA, it loses to GML as prediction accuracy presented by the GML technique is higher.

Unlike the previous study of PSO-LSSVM (Jiang & Zhao, 2013; Xiang & Jiang, 2009), in Shen, Zhang and Ma (2009), the dynamic inertia weight PSO (WPSO) is employed to optimize LSSVM hyper-parameters. The optimal value of parameters obtained is then used by LSSVM for prediction task. Upon completing simulation task, the RE, which is selected as performance evaluation metric of the study indicated that the PSO-LSSVM outperformed both experimented prediction techniques, viz. BPNN and LSSVM. Since the PSO used in this study has been modified, a comparison with the standard PSO needs to be performed.

While Xie, et al., (2009) demonstrated the GA-LSSVM for dissolved gases prediction, Liao, et al., (2011) examined the capability of PSO-LSSVM for the similar task. In this study, dissolved gas data from several electric power companies in China were employed as empirical data. As to improve the generalization of LSSVM, the raw data were first normalized before training task. For comparison purposes, four single prediction models were chosen, including three ANN based prediction models and SVR. Upon completing the simulation task, the PSO-LSSVM

was concluded to be superior than the identified techniques relative to MAPE and R^2. However, despite the good performance of the model, the author admitted that the produced results might be unreliable due to the insufficient size of actual data.

Experimented on chaotic time series data, Liu and Yao (2009) presented the use of PSO-LSSVM. For empirical purposes, two benchmark data sets were employed namely Mackey–Glass time series data and Lorenz time series data. As compared to fuzzy prediction based on SVM, the PSO-LSSVM showed superior performance for both cases.

As Fu, et al., (2010) presented AGA-LSSVM for ship motion prediction, Zhou and Shi (2010) demonstrated the hybridization of PSO with LSSVM for the similar context of study. Taking the similar approach as in Fu, et al., (2010), the optimization technique, which is PSO, is modified as to enhance the generalization performance of LSSVM in prediction. The modification of PSO is done by integrating it with Simulated Annealing (SA), termed ad HPSO. Tested on a real data, the performance of proposed method which was guided by RMSE produced an encouraging result. However, the evaluation does not involve any other prediction techniques.

Carbon price prediction using hybrid ARIMA and LSSVM has been demonstrated by Zhu and Wei (2013). For optimization of hyper-parameters of interest, PSO algorithm was employed. However, the PSO was only used to tune the LSSVM hyper-parameters, while the ARIMA parameters are derived empirically. The

research data employed was daily carbon price which covers the period from 2005 to 2010. The performance of proposed model was compared against several single and hybrid techniques, which includes ARIMA-ANN and also LSSVM and the evaluation is based on RMSE. Findings of the study suggested that the presented model demonstrate better predictive capability by producing lower error.

Besides GA and PSO, another EC algorithm which includes Artificial Fish Swarm Algorithm (AFSA) (Li, Shao & Qian, 2002), Ant Colony Optimization (ACO) (Dorigo & Stutzle, 2004), Differential Evolution (DE) (Storn & Price, 1997) and Gravitational Search Algorithm (GSA) were also presented in literature as one of the optimization tool for LSSVM.

AFSA which was first proposed by Li, et al., (2002) is inspired by the intelligent swarming behavior of fish. In AFSA, after initialization of population, the algorithm will go through the selection of the best behavior to be executed, which involved between swarm and follow behavior. The behavior which produced the minimum result will be chosen for the next stage. Even though AFSA possesses an advantage in avoiding local minimum, it suffers from imbalance between exploitation and exploration process, which would affect the optimization process (Peng, 2011). In literature, Chen, et al. (2008) examined the hybridization of LSSVM with AFSA. Using electrical load data set as an empirical data, the produced results indicated that AFSA-LSSVM is capable to yield smaller average RE as compared to PSO-LSSVM.

On the other hand, ACO (Dorigo & Stutzle, 2004) imitates the foraging behavior of ants, which acts as agents in searching the shortest path between the colony and the food source. This algorithm is reported to guarantee the convergence; however the time to converge is inconsistent (Selvi & Umarani, 2010). Besides, as this algorithm is developed based on routing problems, the theoretical is reported to be quite difficult to be applied in other applications (Selvi & Umarani, 2010). In optimizing LSSVM hyper-parameters, Fang and Bai (2009) demonstrated the ACO which has been applied for share price prediction. As to eliminate the fluctuant component in raw data, Wavelet Transform (WT) was employed which served as preprocessor to LSSVM. The proposed model is measured based on MSE and the comparison was made against single SVM and ACO-SVM without WT. The results obtained indicated that ACO-LSSVM offered promising results by producing lower MSE.

Meanwhile, DE (Storn & Price, 1997) is inspired by the mechanism of natural selection which considered as extension of GA. The difference between DE and GA is, in DE, all possible solutions have an equal chance in evaluation task, while in GA, the chance of updating the solution is depends on fitness value. The application of DE involved the predefined of several control parameters which affect the optimization process of DE and is revealed to have tendency of falling into local minimum (Dos Santos, et al., 2012). This situation leads to the modification of DE (termed as CDEK) which has been demonstrated for optimizing the LSSVM hyper-parameters (Dos Santos, et al., 2012). In the study, CDEK-LSSVM was tested on

estimation of thermal process. Empirical experimentation suggested that the proposed model demonstrate better predictive power than the standard DE-LSSVM.

Previously, in 2004, the hybridization of DE-LSSVM has been presented which experimented on three benchmark functions (Feoktistov & Janaqi, 2004). Empirical results showed that the DE-LSSVM offered good convergence performance, however, in terms of prediction accuracy, the results obtained is unsatisfying.

Progressing further, Zhang, Niu, Li and Li (2013) investigated the usefulness of online LSSVM in prediction of turbine heat rate. In the study, the GSA was utilized to optimize the hyper-parameters of LSSVM. GSA (Rashedi, Nezamabadi-pour, & Saryazdi, 2009) is a population-based algorithm which is motivated from Newtonian gravity. Apart from population size and number of iteration, several control parameters need to be predefined by the user, which includes gravitational constant and inertia mass. Upon completing the experiment, the experimental results suggested that GSA-LSSVM offered better generalization capability as compared to the other identified prediction techniques.

In related work, Li (2009) presented a hybrid of LSSVM with Bacterial Colony Chemotaxis (BCC) for short term load prediction. Categorized as bionic algorithm, BCC is colony-intelligent optimization which is based on bacterium's reaction to chemo attractants to find the optimum. Upon completing the experiment, lower MAPE was produced by BCC-LSSVM as compared to several identified techniques,

which includes BPNN and CV-LSSVM. In terms of training time, the demerits of CV that was highlighted in Lendasse, et al., (2005) was proven in the study where it took much longer time to complete the training task as compared to time recorded in BCC-LSSVM.

Meanwhile, in Li, Guo, Zhao, Su and Wang (2012), the LSSVM is hybridized with Fruit Fly Optimization (FFO) for electric load prediction. The FFO (Pan, 2012) is inspired from the food searching behavior. In the study, the FFO is employed as an optimizer to LSSVM and the FFO-LSSVM is compared against several identified techniques which include single LSSVM and regression technique. Final results suggested that the FFO-LSSVM is capable to produce lower error rate relative to MAPE, MSE and AAE.

As discussed above, both GA and PSO are widely utilized in optimizing the LSSVM hyper parameters. However, despite its extensive use of GA in literature, several deficiency of this technique has been reported. There is high possibility of GA to face with stagnation in the local minimum, premature convergence, low convergence speed and difficulty in parameters selection (Chen, et al., 2008).As for PSO, it is also revealed that the PSO tends to fall into local minimum (Park, et al., 2010). Besides GA and PSO, the application of other EC algorithms which includes AFSA, ACO, DE and GSA are also presented in literature. However, the applications of these techniques are less obtrusive. Therefore, in this research, the proposed algorithm addressed the drawbacks of the reviewed existing works.

Table 2.2 summaries the existing works on time series prediction technique based on LSSVM optimized by EC algorithm.

Table 2.2: Summaries of Time Series Prediction Technique using LSSVM Optimized by EC Algorithm

Authors	Research Area	Independent Variable	Method	Data Frequency	Evaluation Metric
Mustafa, et al. (2011)	Transmission loss prediction	Load curve and load demand	GA-LSSVM	Hourly	MSE
Mustafa, et al. (2012)	Reactive power prediction	Load curve and load demand	GA-LSSVM	Hourly	MSE
Yu, et al. (2009)	Stock market prediction	S&P 500 index, Dow Jones Industrial Average (DJIA) index and New York Stock Exchange (NYSE) index	GA-LSSVM	Monthly	Hit ratio
Fu, et al. (2010)	Ship motion prediction	Heading angle	AGA-LSSVM	Seconds	MAE
Sun & Zhang (2008)	Power spot prices prediction	Power spot prices	GA-LSSVM	Daily	MAE and MAPE
Xie, et al. (2009)	Dissolved gas concentration prediction	Dissolved gas analysis data	GA-LSSVM	-	Absolute error
Yang, et al. (2010)	Conductivity and tensile strength prediction modeling for carbon fiber composites	Temperature, average fiber, fiber content, conductivity, and several others	GA-LSSVM	-	LMSE and GMSE
Wang, et al. (2005)	Chaotic time series data prediction	Henon map time series data	GA-LSSVM	-	RMSE
Liao & Balzen (2013)	Network flow prediction	Network access flow	Self-Adaptive GA-LSSVM	Hourly	RMSE and MAPE
Jiang & Zhang(2013)	Critical heat flux prediction	Mass flow rate, pressure, the heated length to tube diameter	PSO-LSSVM	-	Relative Error

		ratio and equilibrium quality			
Xiang & Jiang (2009)	Water quality prediction	Water figures of Lixiu River	PSO-LSSVM	-	MAPE
Shen, et al. (2009)	Stock return prediction	Stock return of 5 days backward and closing prices of 3 days forward, and several others	WPOS-LSSVM	Daily	Error
Liao, et al.(2011)	Dissolved gases prediction	Dissolved gas data110kV and 220kV	PSO-LSSVM	-	MAPE and R^2
Liu & Yao (2009)	Chaotic time series prediction	Mackey-Glass time series data and Lorenz time series data	PSO-LSSVM	-	RMSE
Zhou & Shi (2010)	Ship motion prediction	A series of roll and pitch time series data of a real ship during her seakeeping trial	HPSO-LSSVM	-	RMSE
Zhu & Wei (2013)	Carbon price prediction	Carbon price data sets	PSO-ARIMA-LSSVM	Daily	RMSE, Dstat
Chen, et al. (2008)	Electricity load prediction	Electricity load data	AFSA-LSSVM	Hourly	RE
Fang & Bai (2009)	Share price prediction	Share price data	ACO-LSSVM	Daily	MSE
Dos Santos, et al. (2012)	Thermal process prediction	Temperature, relative humidity	CDEK-LSSVM	Second	R^2
Feoktistov & Janaqi (2004)	Prediction	Benchmark functions	DE-LSSVM	-	-

Zhang, et al. (2013)	Turbine heat rate prediction	Heat rate	GSA-LSSVM	Hourly	R^2, APE and MAPE
Li (2009)	Short-term load prediction	Load value of last three hours, the load value of the very hour in last three days, and several others	BCC-LSSVM	Hourly	MAPE and computational time
Li, et al. (2012)	Electric load prediction	Electricity consumption	FFO-LSSVM	Annually	MAPE, MSE and AAE

2.2 Methods to Enhance Searching Behavior in ABC Algorithm

Improving the searching behavior of the ABC algorithm is vital as it would encourage the algorithm to perform wider search. Such an approach will then assist the algorithm to escape from local minimum trapping caused by premature convergence (Jahjouh, 2012). In literature, many modifications on the searching behavior of the ABC algorithm has been proposed and this includes Rajasekhar, et al., (2011), Zhang, Guan, Tang and Tang (2011), Sharma, Pant, and Bansal (2012), Subotic (2011), Hsieh, Hsiao, and Yeh (2012), Alatas (2010), and others. This indicates the importance of this issue in the area of interest. It is worth noting to mention that by addressing the local minimum issue, will directly contribute to overcome the premature convergence problem (Jahjouh, 2012).

In Rajasekhar, et al., (2011), the standard ABC algorithm is improved by implementing Levy mutation in producing a new solution. The purpose is to guide the bees in searching for global optimum. However, the predetermined value of α which corresponds to LPD limits the shape of distribution. This consequently affects the searching behavior of the bees. In addition, even though the study introduced a new formula in producing new solutions, the concept of depending on a single equation is still retained, which has been one of the critical concerns in literature (Babayigit & Ozdemir, 2012; Sharma, et al., 2011; Subotic, 2011).

In related work, a modification in ABC has been presented by Zhang, et al., (2011). In the presented model, instead of using the greedy selection scheme, the new

equation is introduced which combine the former neighbor of old solution with the better solution selected in greedy selection. With such modification, it provides larger search range. Compared against the standard ABC, the presented ABC improved the global searching in some optimization problems. Nonetheless, for certain case, it fails to make further progress at final stage.

As to enhance the searching process in ABC, three modifications have been introduced by Sharma, et al., (2012) namely colony size reduction, perturbation scheme improvement and introducing a new control parameter, namely modification rate (*MR*). In this version, the old solution is improved if the generated random number is less than *MR*, otherwise it is retained. Tested on eight benchmark functions, the findings of the study indicated that the modified ABC performs better than the standard ABC. Similar approach also can be seen in Brajevic, Suba and Subotic (2010).

Multiple onlookers scheme was demonstrated in Subotic (2011). In the study, the solutions have to be sorted based on probability before entering the onlooker bee phase. For this matter, a new solution is introduced which involved another three parameters which are sorted by probability. Realized in thirteen benchmark functions, the empirical results suggested that the presented model outperforms the standard ABC in terms of best values, mean values, worst values and standard deviation. Nevertheless, a new pre determined parameter still need to be taken into account.

In Hsieh, et al., (2012), a hybrid PSO with ABC was presented which is realized in financial prediction. In the study, each bee was given a property of PSO, viz. velocity and flying direction. The objective of modification is for optimization of SVM parameter. As to investigate the efficiency of the presented model, it was compared against several other methods which include SVM optimized by standard ABC. Findings of the study suggested that the PSO-ABC-SVM capable to produce better results as compared to the other identified techniques. However, the combination of two optimization algorithms may cost longer computational time.

In related work, Alatas (2010) presented the chaotic mechanism in improving the searching behavior of the artificial bees in ABC algorithm. By using this approach, the chaotic variable is generated according to the selected map, and later it is used in producing the new solution. Based on the results obtained, the chaotic produced better outputs than the other identified techniques. However, it is noted that the modification introduced caused more extra function evaluations in chaotic search (Gao & Liu, 2011).

Meanwhile, Zhu and Kwong (2010) demonstrated a modified ABC (GABC) which integrates global best information (gbest) into solution search expression. Similarly like the other studies, the objective of the modification is to improve the exploitation process of the artificial bees in the search space. Other related work of GABC also has been proposed in Roy and Jadhav (2015). The modified version directs the

search trajectory towards global optimal. Nonetheless, there is no guarantee to avoid from local minimum.

Progressing further, Gao, et al. (2012) demonstrated a modified ABC which is inspired from DE algorithm. In the presented model, each bee perform the searching process only around the best solution recorded from the previous iteration. The feasibility of the presented model is experimented on 26 benchmark functions. The empirical results show that the improved ABC superior than the other comparison techniques. Quite similar approach also can be seen in Gao and Liu (2011).

An Interactive ABC (IABC) has been presented for numerical optimization problems (Tsai, Pan, Liao, & Chu, 2009). In IABC, the modification involved the movement of the onlooker bees which is done based on the universal gravitation force theory. The objective is to enrich the searching capability of the standard ABC. Comparison against standard ABC and PSO was in favor to IABC. However, the modification introduced caused the algorithm took a longer time to find the near best solution.

From the extensive literature review that has been done in this section, it can be seen that there are still some gaps that can be filled in order to enhance the performance of ABC. From various improvements that have been proposed, it shows that the study in this area is still encouraging and critical to explore. Thus, it has paved a way for this study to propose new algorithm to overcome the shortcoming of single ABC. Table 2.3 summaries the existing works on this section.

Table 2.3: Summaries on Methods to Overcome Searching Behavior in ABC Algorithm

Author	Research Area	Method	Evaluation Metric
Rajasekhar, et al. (2011)	Modification of searching behavior in ABC	*L*-ABC	Standard deviation, error
Sharma, et al. (2012)	Modification of searching behavior in ABC	RABC	Standard deviation, mean
Zhang, et al. (2011)	Modification of searching behavior in ABC	Modified ABC	Mean of the best values
Subotic (2011)	Modification of searching behavior in ABC	MO-ABC	Best, mean, worst values and standard deviation
Brajevic, et al. (2010)	Modification of searching behavior in ABC	Improved ABC	Average and standard deviation
Hsieh, et al. (201)	Modification of searching behavior in ABC	EABC-PGSVM	Best hit ratio
Gao, et al. (2012)	Modification of searching behavior in ABC	ABC/best	Mean and standard deviation
Gao & Liu (2011)	Modification of searching behavior in ABC	IABC	Mean and standard deviation
Alatas (2010)	Modification of searching behavior in ABC	IABC	Success rates
Zhu & Kwong (2010)	Modification of searching behavior in ABC	GABC	Mean and standard deviation
Roy & Jadhav (2015)	Modification of searching behavior in ABC	GABC	Simulation time, minimum, maximum and average cost
Tsai, et al. (2009)	Modification of searching behavior in ABC	IABC	The cost of time to find the near best solution

2.3 Methods to Prevent Over fitting and Under fitting in LSSVM and SVM

In time series prediction, over fitting and under fitting during learning is one of the critical issue (Jammazi & Aloui, 2012; Shrivasta, Ch, & Panigrahi, 2011) that need to addressed. This is because over fit or under fit learning models may lead the algorithm poor generalization (Shi, et al., 2013). For that matter, preventing the model from facing with such situation is necessary, for the benefits generalization performance.

Hegazy, Soliman and Abdul Salam (2013) employed the PSO algorithm in selecting the best value of the LSSVM hyper parameters and also preventing the over fitting and under fitting. The PSO-LSSVM is later used in prediction of daily stock market. Comparison against ANN model indicates that the PSO-LSSVM capable to produce better prediction accuracy. However, the absent of value of LSSVM hyper parameters in the study made the discussion on over fitting incomplete. Similar approach was also presented by Zhu and Wei (2011) which demonstrated a hybrid PSO-LSSVM. The LSSVM which was integrated with ARIMA is used to predict the carbon price. The empirical results obtained shows that, even an encouraging results are recorded, however, based on the minimum values recorded by the kernel parameter, the over fitting problem in the study are still unaddressed.

Meanwhile, in Pahasa and Ngamroo (2011), the PSO with *k*-fold CV is utilized in optimization of LSSVM hyper parameters which is believed capable to detect and prevent the over fitting situation. The efficiency of the CV-PSO-LSSVM is

simulated in power system stabilization. Measured based on MAE, the presented model is superior to the competitor method under study. However, as highlighted in section 2.1, the employment of CV would raise the computational time issue.

A stacked LSSVM model has been introduced to overcome the problem of over fitting and under fitting in LSSVM (Zhang, Wang, & He, 2012). In stacked LSSVM, each LSSVM model is trained on a bootstrap re-sampling replication of the training data. Since the LSSVM model is trained separately, the stacked LSSVM is expected capable to overcome the over fitting and under fitting problem during learning. The stacked LSSVM which is tested on a batch processes show an encouraging performance as compared to the stacked RBFNN. Even though the over fitting and under fitting were highlighted, but no discussion was presented.

In Yeh, Chi and Hsu (2010), a hybrid approach of data envelopment analysis (DEA) with LSSVM for financial prediction was presented. Concerning the over fitting issue in LSSVM, the problem is addressed using *k*-fold CV procedure. The effectiveness of the presented model was verified by comparing against ANN based model. Findings of the study suggested that the CV-DEA-LSSVM model posses the ability to produce high prediction accuracy. Again, it can be expected to raise an issue of long computational time.

Regularized LSSVM for financial prediction was demonstrated by Khemchandani, Jayadeva and Chandra (2009). The presented model assigns fuzzy membership

values to the training data to capture the input and output relationship in a better way and reduces over-fitting risk as well. Findings of the study showed that the regularized LSSVM is superior than the single LSSVM. However, to assign fuzzy membership values in large data set is quite complicated task.

Jain, Smith, Culligan and Taylor (2014) employed similar approach in avoiding the over fitting problem in SVM algorithm. The predictive power of the proposed CV-SVM is realized in prediction of energy consumption of multi-family residential buildings. The obtained results indicated that the CV-SVM produced positive outcomes. Nevertheless, no comparison with other technique was performed.

Similarly like in Jain, et al, (2014), the CV procedure was also reported in Papadimitriou, Gogas, and Stathakis (2014) in preventing the over fitting of SVM during learning. The CV-SVM is later realized in prediction of energy markets. The empirical findings shows that the CV-SVM recorded an average performance, which is less than 80% prediction accuracy.

Previously, Qi, Hu, Peng, and Ren (2011) had investigate the efficiency of CV-SVM for electrical evoked potentials prediction. Having the similar purpose like the previous studies, the CV is employed to serve as an avoider for the model from facing with over fitting issue. As to improve the learning process, the SVM is enhanced with multiple weights (SVM-MW). The CV-SVM-MW was empirically tested with actual data of electrical evoked potentials. Empirical results suggested

that the CV-SVM-MW outperform the other identified techniques, which include BPNN and single SVM.

Similar approach was also presented by Lee (2009) who utilized the CV to prevent the over fitting and under fitting problem in SVM. Tested on stock trend prediction, the presented approach was compared against the results produced by BPNN. Final output shows that the CV-SVM outperform BPNN significantly.Unlike the previous studies which employed the CV, the GA-CV has been presented in Nieto, Fernandez, Juez, Lasheras, and Muniz (2013). Realized in prediction of cyanotoxins presence in the Trasona reservoir, quite a number of variables are fed to the prediction model which includes water temperature, ambient temperature, and many others. Even claimed to have good results, however, no comparison with other techniques was performed. Thus, the justification is unconvincing.

As discussed above, even CV is widely utilized, however, CV is inefficient since it is time consuming. Meanwhile, the PSO and GA face with the risk of local minimum trapping and premature convergence (Chen, et al., 2008; Park, et al., 2010). In addition, even the works discussed highlighted the over fitting and under fitting issue, nonetheless, there is no discussion were presented. Therefore, this study addressed the problems by improving the decision making process in ABC algorithm. Table 2.4 summaries the existing works on method to prevent over fitting and under fitting problem in LSSVM and SVM.

Table 2.4: Summaries on Methods to Prevent Over fitting and Under fitting in LSSVM and SVM

Author	Research Area	Independent Variable	Method	Data Frequency	Evaluation Metric
Hegazy, et al. (2013)	Prevent over fitting/under fitting in LSSVM	Relative Strength Index (RSI), Money Flow Index (MFI), Exponential Moving Average (EMA) and others	PSO-LSSVM	Daily	MSE
Pahasa & Ngamroo (2011)	Prevent over fitting/under fitting in LSSVM	Past and current values of power deviation and control signal features	CV-PSO-LSSVM	-	MAE
Zhang, et al. (2012)	Prevent over fitting/under fitting in LSSVM	Normal operating data,	Stacked LSSVM	-	Root Mean Square Error Prediction (RMSEP), MAE
Yeh, et al. (2010)	Prevent over fitting/under fitting in LSSVM	Net income, DEA and others	CV-DEA-LSSVM	-	Prediction Accuracy
Zhu & Wei (2011)	Prevent over fitting/under fitting in LSSVM	Carbon price	PSO-LSSVM-ARIMA	Daily	RMSE, Dstat
Khemchandani, Jayadeva and	Prevent over fitting/under fitting in	Financial data of IBM, Microsoft, Google and many others	Regularized LSSVM	-	NMSE

Chandra (2009)	LSSVM				
Jain, et al. (2014)	Prevent over fitting/under fitting in SVM	Electrical consumption, current temperature, current solar flux	CV-SVM	Daily, hourly, 10 minutes interval	Coefficient of variation (CV)
Papadimitriou, et al. (2014)	Prevent over fitting/under fitting in SVM	Index prices	CV-SVM	Daily	Prediction Accuracy
Qi, et al. (2011)	Prevent over fitting/under fitting in SVM	Electrical evoked potentials	CV-SVM-MW		SPPA, DSA, TIC
Lee (2009)	Prevent over fitting/under fitting in SVM	NASDAQ index	CV-SVM	Daily	Standard deviation, mean
Nieto, et al. (2013)	Prevent over fitting/under fitting in SVM	Water temperature, ambient temperature, conductivity	GA-CV--SVM	-	-

2.4 Evaluation in Time Series Predictions

In this section, the metrics and methods employed in time series predictions are reviewed.

2.4.1 Metrics

In evaluating time series prediction models, the utilization of appropriate evaluation metric is important. This is to justify the obtained results from the conducted experiments. In literature, various metrics have been employed and one of the popular metric is Mean Absolute Percentage Error (MAPE) (Hyndman & Koehler, 2006). As represented by its name, MAPE is a metric that is based on percentage error. MAPE is a scale independent metric, thus making it suitable for evaluating different time series data. The metric is measured by obtaining the sum of all absolute percentage errors and computing their average value. In literature, this metric have been widely utilized in Alizadeh, et al. (2012) and Yusof, et al. (2010) which adopted the metric for crude oil price prediction, while Varahrami (2011), and Liu and Wang (2008) utilized this metric in the prediction of gold price and foreign exchange rate market respectively. The utilization of MAPE can also be seen in Lin, et al. (2012) and Liao and Balzen (2013) in financial prediction and network flow prediction using LSSVM based prediction technique.

On the other hand, a number of studies prefer to use Root Mean Square Error (RMSE) (Hyndman & Koehler, 2006) as a prediction performance measure. Unlike MAPE, the RMSE is scale dependent (Hyndman & Koehler, 2006) which indicates

that it is unsuitable for comparing the performance of different time series data. Example of research work that utilizes this metric includes Yusof, et al. (2010), Zhang, et al. (2010), Abdullah and Zeng (2010). Yu, et al. (2008), Haidar and Wolf (2011) and Wang, et al. (2005) who applied it in crude oil rice prediction. This metric also can be seen in network flow prediction that uses Self Adaptive GA-LSSVM (Liao & Balzen, 2013)

Another well known scale dependent metric is the Mean Square Error (MSE) where this type of metric is useful for making comparison of performance between different methods applied to the same data set, but this metric should not be used when involving different data sets, as wrongly used in Malliaris and Malliaris (2008). Thus, make the obtained results can be argued. By using MSE, the result is obtained by averaging the square error. The MSE also has been used in Jammazi and Aloui (2012), Wu and Niu (2009), Van Gestel, et al., (2001) and Mustafa, et al., (2012) for evaluating the prediction performance in prediction tasks of different fields.

Meanwhile, the prediction accuracy also been favorable for some researchers. The prediction accuracy obtained is reflected to the error rate recorded using MAPE (Hyndman & Koehler, 2006). By using this metric, the prediction accuracy is interpreted in straight forward manner. This metric has been used as a performance measurement in Tehrani and Khodayar (2011) who applied it for crude oil price prediction. Besides, Dooley and Lenihan (2005) used this metric for evaluating the performance of ARIMA in metal price prediction.

In Jiang and Zhao (2013), the coefficient of determination (R^2) was employed to serve as performance evaluation metric in critical heat flux prediction. The R^2 measures the goodness of fit of the regression model where the closer the value of it to 1, the better it is (Jiang and Zhao, 2013). This metric also has been used in Haidar and Wolf (2011) and Yu, et al. (2008) who both applied it in crude oil price prediction using ANN based prediction model respectively.

On the other hand, the directional prediction statistic which is also known as Dstat is used to evaluate the directional prediction, which is expressed in percentage. This metric is useful to evaluate the directional performance of the prediction model (Yu, Wang, & Lai, 2007) as demonstrated in Wang, et al., (2005), Yu, et al., (2008) and Zhou and Lai (2011).

Besides the reviewed metrics, there are also other metrics that have been used as a performance evaluation metrics in time series prediction such as Sum Square Error (SSE) (Hussein, et al., 2011), Mean Bias Error (MBE) (Mellit, et al., 2013), Learning Mean Square Error (LMSE) and Generalization Mean Square Error (GMSE) (Yang, et al., 2010), Theil's U (Yusof, et al, 2010), CP and CD (Lin, et al. 2012), Hit Rate (Haidar & Wolf, 2011) and several others. However, these metrics are not as widely used as compared to the metrics discussed previously.

In this study, as it used different time series data, the scale independent metrics were utilized, namely MAPE, symmetric MAPE (sMAPE) and Root Mean Square

Percentage Error (RMSPE). The sMAPE was used to cater the problem when the difference between the actual and predicted values are too large. For this situation, the symmetric MAPE (sMAPE) is preferable. By using sMAPE, it will give more accurate results as its fluctuates between -200% to 200% while the MAPE has no limits (Hyndman & Koehler, 2006). Besides the stated metrics, the prediction accuracy and Theil's U were also employed as it is suitable in comparing the performance of different data sets(Hyndman & Koehler, 2006) and Theil's U (Armstrong, 2001). A summary of the metrics utilized in time series prediction is presented in Table 2.5.

Table 2.5: Summaries of Metrics utilized in Evaluation of Time Series Prediction

Author	Research Area	Method	Data Frequency	Evaluation Metric
Alizadeh, et al. (2012)	Crude oil price prediction	ANN-FR	Annually	MAPE
Yusof, et al. (2010)	Crude oil production prediction	ARIMA	Monthly	RMSE, MAPE, MAE, Theil's U
Varahrami (2011)	Gold price prediction	MLFF-BP and GMDH	Daily	RMSE and MAPE
Liu & Wang (2008)	Exchange rates prediction	CV-LSSVM	Daily	MAE, MAPE and RMSE
Lin, et al.(2012)	Exchange rates prediction	CV-EMD-LSSVM	Daily	MAE, MAPE, CP, CD
Liao & Balzen (2013)	Network flow prediction	Self-Adaptive GA-LSSVM	Hourly	RMSE and MAPE
Zhang, et al. (2010)	Crude oil price prediction	Fuzzy Time Series	-	RMSE
Abdullah & Zeng, (2010)	Crude oil price prediction	ANN-Q	Monthly	RMSE and Dstat
Yu, et al. (2008)	Crude oil price prediction	An AI-agent based Trapezoidal Fuzzy Ensemble	Daily	RMSE, Dstat and Coefficient of determination (R^2
Haidar & Wolf (2011)	Crude oil price prediction	ANN	Daily	Hit rate, $R^2$and RMSE
Wang, et al. (2005)	Crude oil price prediction	TEI@I methodology	Monthly	RMSE and Dstat
Malliaris & Malliaris (2008)	Crude oil, heating oil, gasoline and propane price prediction	BPNN	Daily	MSE and AAE
Jammazi & Aloui (2012)	Crude oil price prediction	Wavelet decomposition and NN	Monthly	MSE and MAE
Wu & Niu (2009)	Short-term load prediction	CV-LSSVM	Daily	MSE
Van Gestel, et al. (2001)	Financial time series prediction	CV-LSSVM	Weekly	MSE, NLL and MAE
Mustafa, et al. (2012)	Reactive power prediction	GA-LSSVM	Hourly	MSE

Tehrani & Khodayar (2011)	Crude oil price prediction	GA-FFNN	Monthly	Prediction accuracy
Dooley & Lenihan (2005)	Lead and zinc price prediction	ARIMA	-	Prediction error
Jiang & Zhao (2013)	Critical heat flux prediction	PSO-LSSVM		RMSE,R^2
Hussein, et al. (2011)	Gold price prediction	RBFNN	Daily	SSE
Melit, et al. (2013)	Short term meteorological time series data prediction	CV-LSSVM	Daily	R^2, RMSE, Mean Bias Error (MBE), MAPE
Yang, et al. (2010)	Conductivity and tensile strength prediction	GA-LSSVM	-	LMSE and GMSE

2.4.2 Methods

In time series predictions, the methods utilized can be categorized into two major classes, namely statistical approach and Computational Intelligence (CI) approach. In literature, time series prediction initiates with the employment of statistical approach. Later, due to the demerits of statistical approach, the CI approach became popular in the academic community.

2.4.2.1 Statistical Approach

Even though the application of ARIMA in financial prediction is regarded as unsuitable and inefficient (Liu, 2009; Yu, et al., 2008), the utilization of this technique in prediction, including for metal price is still of interest for several academia. Dooley and Lenihan (2005) presented a price prediction model based on ARIMA for two metals prices, namely lead and zinc. The samples for both metals prices are from November 1988 to 1999. As highlighted in many studies, the unsuitability of the proposed method in the field of interest is proven in this study when the proposed model was unable to offer satisfied prediction error in several conducted experiments.

In Yusof, et al., (2010), prediction for Malaysia's crude oil production was presented using ARIMA. In developing the model, there are three main stages, namely model identification which involves statistical procedure, model estimation which is related to parameter estimation and finally model validation which is used to evaluate how well the model fits the data. Research data was covered from 2005 to 2010 in

monthly basis and the performance is guided based on several statistical metrics including Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). No comparison with other techniques was made in experiment task. Thus, the predictability of the proposed model was unconvincing. Besides, the application of ARIMA in crude oil price prediction is much debated in literature due to inefficiency of this technique in dealing with non-linearity of crude oil data (Abdullah& Zeng, 2010; Xiao, et al. 2012). Table 2.6 summaries the existing work on time series prediction technique using statistical approach.

Table 2:6: Summaries of Time Series Prediction Technique using Statistical Approach

Author	Research Area	Independent Variable	Method	Data Frequency	Evaluation Metric
Dooley & Lenihan (2005)	Lead and zinc price prediction	Cash of lead and zinc, 3 and 15 months forward of lead and zinc	ARIMA	-	Prediction error
Yusof, et al. (2010)	Crude oil production prediction	Malaysia crude oil production	ARIMA	Monthly	RMSE, MAPE, MAE, Theil's U

2.4.2.2 Computational Intelligence Approach

In literature, it is observed that the application of ANN has been favorable choice as a prediction tool for the energy fuels price prediction, specifically crude oil price. Due to the popularity of ANN, Malliaris and Malliaris (2008) attempted to predict four energy fuels price, namely crude oil, heating oil, gasoline and propane. Prior to the model development, the price correlation coefficient between the energy fuels was calculated. This is to ensure the significance of each input fed to the model. For comparison purposes, the proposed model was compared against Regression Model and Simple Model and the evaluation is relative to Mean Square Error (MSE) and Average Absolute Error (AAE). Empirical results indicated that in predicting crude oil, heating oil and gasoline price, ANN demonstrates better predictive power. However, in prediction of propane price, the proposed model was defeated by Simple Model. In this study, two points can be made. First, the absence of data normalization to the research data employed would affect the predictability of ANN (Abdullah & Zeng, 2010; Zhang, et al., 1998). Secondly, the employment of MSE and AAE in evaluating different time series data is inappropriate since it would cause biasness in the produced results as both metric are scale dependent (Hyndman & Koehler, 2006).

Jammazi and Aloui (2012) presented a crude oil price prediction based on combined wavelet decomposition and ANN. In this study, the monthly West Texas Intermediate (WTI) crude oil prices between 1988 and 2010 are used as empirical data. Similar with Malliaris and Malliaris (2008), no data normalization was

conducted prior to training phase. As to avoid the over fitting issues, particular attention was given in dealing with the design of ANN model which includes the determination of architecture. Final results suggested that, by comparing with conventional BPNN, the proposed method proved to be superior in the context of study.

Previously in 2010, Fuzzy Time Series which combined people's subjective attitude and objective history values has been proposed for crude oil price prediction (Zhang, et al., 2010). Similarly with Jammazi and Aloui (2012), WTI crude oil price is employed for empirical purposes. The data sets covered from January 1991 to September 2001 and the evaluation is made based on RMSE. While the results obtained is reported to be good, the absence of other competing method makes the efficiency of the presented method doubtful.

The combination of machine learning approach with ANN-Quantitative (ANN-Q) in prediction of crude oil prices has been demonstrated by Abdullah and Zeng (2010). To train the input variables, Back Propagation Neural Network (BPNN) algorithm was used. In this study, besides quantitative data, qualitative data which were obtained from experts' opinion were utilized. Nevertheless, this would raise an issue since the point of views of the experts can be varied and biased (Kulkarni & Haidar, 2009). Upon completion of the experiment, the final results indicated that more improvement needs to be considered since the ANN-Q was unable to outperform the prediction technique presented in Wang, Yu, and Lai (2005).

In Wang, et al., (2005), crude oil price prediction based on TEI@I approach has been presented which combined Text Mining, Econometrics and intelligent techniques. For experimental purposes, besides the crude oil price, several other factors were also included to be fed to the prediction model, which are irregular events. However, the irregular events are considered as outliers which would affect the predictability of prediction model (Hyndman & Athanasopoulos, 2012; Tang & Hammoudeh, 2002). Findings of the study were guided by RMSE and Dstat, and comparison against ANN, ARIMA and Simple Integration technique was made. The results obtained showed that even the TEI@I approach excels in most cases, however, in certain case, it loses to Simple Integration.

In 2009, Kulkarni and Haidar examined the capability of ANN in predicting crude oil price (Kulkarni & Haidar, 2009). In the study, the Multilayer Feed Forward Neural Network (MLFNN) with Back Propagation (BP) algorithm is utilized. Similar situation with Jammazi and Aloui (2012), serious heed has to be paid in determining the ideal architecture of ANN. For experiment purposes, daily WTI crude oil prices were employed, ranging from September 1996 to August 2007. Without comparing with other comparable methods, the results obtained which were evaluated based on RMSE and hit ratio achieved satisfying results.

Yu, et al., (2008) presented an Artificial Intelligent (AI) agent-based trapezoidal fuzzy ensemble to predict crude oil price. The model involves four main stages to be accomplished. Firstly, single AI-based predictor is created, where different AI

techniques are employed as prediction agent in producing single prediction result. Next, to combine the different predictions, fuzzyfication is utilized. The next stage is fuzzy prediction ensemble and followed by the aggregated prediction defuzzification. The result obtained was in favor to the presented model, while among single model chose as competitor, ARIMA produced the worst result. This situation reinforces the fact that ARIMA is unsuitable to deal with nonlinear characteristic in crude oil prices (Khazem, 2008; Kumar & Thenmozi, 2007).

In Tehrani and Khodayar (2011), a hybridization of GA and Feed Forward Neural Network (FFNN) with BP algorithm has been demonstrated for crude oil price prediction. In the study, GA was employed to improve the learning algorithm and reduce the complexity in determining the control parameters of ANN. Later, the prediction process is continued by the FFNN. The experimental process involved two time series data of crude oil prices, viz. WTI and Iran crude oil prices and comparison was conducted against conventional ANN. Upon completing the experiment, it is indicated that the prediction accuracy produced by GA-FFNN are closer to actual data.

A prediction for crude oil price using ANN and Fuzzy Regression (ANN-FR) has been demonstrated in Alizadeh, et al., (2012). The time series data employed were in annually basis. However, this would limit the data sample significantly (Kulkarni & Haidar, 2009). For comparison purposes, variants of ANN models were selected as

competitors. Findings of the experiment indicated that ANN-FR provided lower error rates, which reflects the prediction accuracy obtained.

In Haidar and Wolff (2011), the prediction of crude oil price is presented utilizing ANN. The empirical data was WTI crude oil prices, which is from 1986 to 2010. Even it is claimed that huge amount of data is required for ANN for generalization purposes, however, the employment of old data would defect the prediction accuracy (Kulkarni & Haidar, 2009). Facing similar issues as other work that employ ANN, serious attention is put forward in dealing with the complicated architecture of ANN. Prior to training, a test for non-linearity was conducted as to match the suitability of data and ANN prediction model. Findings of the study showed that, in terms of Hit rate, the ANN performs better than RW, however, based on R^2, the situation is contrary.

Yu, Wang and Lai (2008) attempted to predict crude oil price using Empirical Mode Decomposition (EMD)-based Neural Network ensemble learning paradigm. In the study, the prediction task is done by using Adaptive Linear Neural Network(ALNN). Firstly, the crude oil price was first decomposed into a number of Intrinsic Mode Function (IMF). Later, the ALNN is used to predict each of the IMF. For each prediction results, a weight is assigned and then, the obtained results are combined together. The evaluation is later made based on RMSE and Dstat. Even empirical results of the study showed an encouraging result, however, this approach is ineffective since if one of the ALNN yields a poor prediction results, the results

might be affected since all results are summed together. This may result in imprecise prediction (Zhou, Lai, & Yen, 2012).

In Liu (2009), the GA is integrated with BPNN for the gold price prediction. In the study, GA is utilized as an optimizer to the value of linking weight of BPNN. For experimental purposes, 240 intraday data set of gold price was employed. From the data set, 200 days were served as training set and the rest 40 days were used as test set. Based on the illustration provided, the GA-BPNN is concluded to outperform single BPNN. However, no numerical results were given for testing set. Thus, the value of prediction error achieved is unknown.

Zhou and Lai (2011) demonstrated an improved EMD online learning-based which is hybrid with BPNN in prediction of gold price. The improved EMD is used to divide the data into several parts where each is trained by BPNN while the online learning algorithm is to prevent the model from falling into local minimum. No normalization technique was performed and the evaluation of prediction performance is guided based on MAE and Dstat. Upon completing the simulation task, the IEMD Online Learning-BPNN model showed improvement from Yu, et al., (2008). Similar study on this also can be seen in Zhou, et al., (2012).

The application of Multi-Layer Feed Forward (MLFF) with BP algorithm and Group Method Data Handling (GMDH) with GA has been presented by Varahrami (2011) in prediction of gold price. In the study, the moving average cross over inputs of

gold price was used as input factor and the performance is evaluated base on RMSE and MAPE. During simulation task, the author experimented with different sets of architectures for both prediction models, specifically in determining the number of hidden layers and neurons. The final results obtained indicated that the GMDH performed better than MLFF relative to the metrics employed.

In related work, a multiple Radial Basis Function Neural Network (RBFNN) prediction model for gold price has been presented in Hussein, et al., 2011. After data collection, data pre-processing was performed to facilitate the network learning and to avoid computational difficulties. Based on the results provided, even the multiple RBFNN model is capable to outperform Auto Regressive Model, however it loses to single RBFNN. Besides the lack of input factor, the inappropriate portion for training and testing is one of the reasons that contribute to the unsatisfied result obtained. This is due to too much portions were set for training which finally led to over fitting.

From the above discussion, it is observed that besides the over fitting issue, the application of ANN requires careful attention in determining the network topology and also the embedded control parameters which would directly affect the capability of ANN in prediction task. These limitations are addressed by using LSSVM which is presented in the next section. Table 2.7 summarizes the existing work on time series prediction using ANN.

Table 2.7: Summaries of Time Series Prediction Technique using CI Approach

Author	Research Area	Independent Variable	Method	Data Frequency	Evaluation Metric
Malliaris & Malliaris (2008)	Crude oil, heating oil, gasoline and propane price prediction	Daily price of crude oil, heating oil, gasoline and propane prices and derivative features	BPNN	Daily	MSE and AAE
Jammazi & Aloui (2012)	Crude oil price prediction	WTI crude oil price	Wavelet decomposition and NN	Monthly	MSE and MAE
Zhang, et al. (2010)	Crude oil price prediction	WTI crude oil price	Fuzzy Time Series	-	RMSE
Abdullah & Zeng, (2010)	Crude oil price prediction	WTI crude oil price, supply, demand and several others	ANN-Q	Monthly	RMSE and Dstat
Wang, et al. (2005)	Crude oil price prediction	WTI crude oil prices and irregular events	TEI@I methodology	Monthly	RMSE and Dstat
Kulkarni & Haidar (2009)	Crude oil price prediction	WTI crude oil prices and futures contracts	ANN	Daily	RMSE and Hit rate
Yu, et al. (2008)	Crude oil price prediction	WTI and Brent crude oil prices	An AI-agent based Trapezoidal Fuzzy Ensemble	Daily	RMSE, Dstat and Coefficient of determination (R^2
Tehrani & Khodayar (2011)	Crude oil price prediction	WTI and Iran crude oil prices	GA-FFNN	Monthly	Prediction accuracy

Alizadeh, et al. (2012)	Crude oil price prediction	Oil supply, crude oil distillation capacity and several others	ANN-FR	Annually	MAPE
Haidar & Wolf (2011)	Crude oil price prediction	WTI crude oil prices	ANN	Daily	Hit rate, $R^2$and RMSE
Yu, et al. (2008)	Crude oil price prediction	WTI and Brent crude oil prices	EMD-FNN-ALNN	Daily	Dstat and RMSE
Liu (2009)	Gold price prediction	Gold price, closing gold price of previous five days	GA-BPNN	Daily	-
Zhou & Lai (2011)	Gold price prediction	Gold price	IEMD Online Learning-BPNN	Daily	MAE and Dstat
Varahrami (2011)	Gold price prediction	Gold price, moving average crossover of gold price	MLFF-BP and GMDH	Daily	RMSE and MAPE
Hussein, et al.(2011)	Gold price prediction	Gold price	RBFNN	Daily`	Sum Square Error (SSE)

2.5 Least Squares Support Vector Machines

SVM has appeared as an active study in machine learning community and extensively used in various fields including in prediction, pattern recognition and many more (Xie, Shao, & Zhao, 2011). Presently, SVM has been proven to offer better solutions than ANN and reported to be good in generalization performance (Yu et al., 2009). Introduced by Vapnik and colleagues (Christianini & Shawe-Taylor, 2000), this technique has solved the over fitting and local minimum which are found in ANN (Xiang & Jiang, 2009). The main theory of this approach is to map the nonlinear input data from input space to high dimensional feature space where in high dimensional feature space, the linear separation can be done (see Figure 2.2).

Due to the merits offered by SVM, the application of this technique has been preferred among research community, including in solving prediction task. However, since SVM requires the solution of nonlinear equations (Quadratic Programming), it possess complicated model, high computational complexity, large memory requirement and long running time (Liu, et al., 2005; Yu, et al., 2009). These limitations have led to the reformulation of LSSVM (Suykens, et al., 2002).

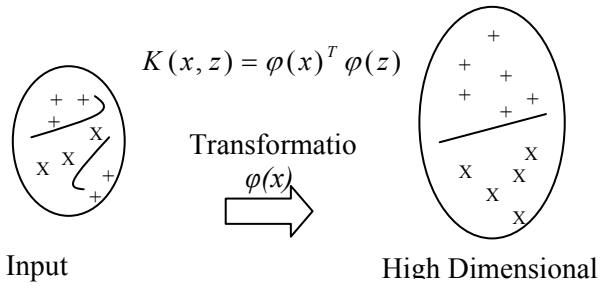


Figure 2.2: Mapping of the Input Space to a High Dimensional Feature Space. Adapted from Least Squares Support Vector Machines (p.38), by J.A.K. Suykens, J.D. Brabanter, B.D. Moor and J. Vandewalle, 2002, Leuven, Belgium: World Scientific Publishing Co. Pte. Ltd. Copyright 2002 by the World Scientific Publishing Co. Pte. Ltd.. Adapted with permission.

The LSSVM (Suykens, et al., 2002) is a variant of standard SVM which offers better solution strategy. As a modified version of standard SVM, LSSVM uses square errors instead of nonnegative errors in the cost function and applies equality constraint rather inequality constraint of SVM in the problem formulation. As a result, one solves a linear equations instead of Quadratic Programming (QP) solver which in practice is harder to use (Suykens, et al., 2002). The adaptation of QP also raises computational complexity in training which is at least quadratic with respect to the number of training data (Huang, Zhou, Ding, & Zhang, 2012). With the reformulation of LSSVM, it simplifies complex calculation which led to easier and faster training task. Hence, a simpler optimization problem can be obtained. In addition, LSSVM offers less control parameters, which are γ and σ^2, as compared to three control parameters required in SVM (C, σ^2 and ε) (Ou & Wang, 2009; Suykens, et al., 2002). In addition, in terms of prediction task, LSSVM is proven to

be better than SVM (Ou & Wang, 2009; Tarhouni, Laabidi, Zidi, & Ksouri-Lahmari, 2011). As the hyper-parameters optimization of LSSVM involved the proper selection of hyper-parameters (i.e. γ and σ^2), the optimized values of both hyper-parameters are crucial in order to achieve good generalization. A large value of γ will cause LSSVM over fit the training data while a small value of γ will expose the model to under fitting problem. On the other hand, for σ^2, a large value will lead to under fitting problem while over fitting is likely to occur when a small value is derived (Fu, et al., 2010; Lendasse, et al. 2005; Wu, et al., 2007).

Formally, given a training set of N points $\{x_i, y_i\}^N$ with the input values x_i and the output values y_i, for nonlinear regression, the data are generated by the nonlinear function $y(x) = f(x_i) + e_i$, the objective is to estimate a model of the following form (Suykens, et al., 2002):

$$y(x) = w^T \varphi(x_i) + b + e_i \tag{2.1}$$

where w is the weight vector, $\varphi(.):R^n$ is the nonlinear function which maps the input space into a higher dimensional feature space, b denotes the biasness and e_i is the error between the actual and predicted output at the ith training data. The input, x_i and output, $y(x)$ are described in Chapter 3. The coefficient vector w and biased term b can be obtained through the optimization problem which is formulated as follows (Suykens, et al., 2002):

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \tag{2.2}$$

Subject to the equality constraints

$$y_i = w^T \varphi(x_i) + b + e_i \text{ , } i = 1, 2 \text{ , } \ldots, N$$

The first part of (2.2) is used to regulate the weight sizes and penalize large weights. On the other hand, the second part of (2.2) indicates the error in training data.

Applying the Lagrangian multiplier to (2.2) yields:

$$L(w, b, e; \alpha) = J(w, e) - \sum_{i=1}^{N} \alpha_i \{w^T \varphi(x_i) + b + e_i - y_i\} \tag{2.3}$$

where α_i are Lagrange multipliers called support values that can be positive or negative in LSSVM formulation due to the equality constraints, γ is the regularization parameter which balances the complexity of the LSSVM model, i.e. $y(x)$, and the training error. Differentiating (2.3) with *w, b,* e_i *and* α_i, the conditions for optimality of this problem can be obtained by setting all derivatives equal to zero, as expressed in the following:

$$\begin{aligned} &\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\ &\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ &\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \\ &\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0 \end{aligned} \qquad i = 1, 2, \ldots, N \tag{2.4}$$

By elimination of *w*and e_i, the optimization problem can be transformed into the following linear equations:

$$\begin{bmatrix} 0 & y^T \\ y & \Omega + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix} \tag{2.5}$$

With $y = [y_1; \ldots; y_N]$, $\alpha=[\alpha_1, \ldots, \alpha_N]$, I is the identity matrix and 1_v =[1;…;1]. The kernel trick is applied as follows:

$$\begin{aligned} \Omega il &= \varphi(x_i)^T \varphi(x_l) \\ &= K(x_i, x_l) \end{aligned} \quad i, l = 1, \ldots, N \tag{2.6}$$

The resulting of LSSVM model for regression in (2.1) becomes:

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{2.7}$$

Where α and $b$are the solutions of (2.5). In (2.7), there are several available kernel functions $K(x, x_i)$, namely Gaussian kernel or Radial Basis Function (RBF) kernel, Polynomial kernel or Linear kernel function. In this study, the RBF kernel is used since its suitability in dealing with nonlinear cases (Bessedik & Hadi, 2013; Wu, et al., 2007) and give good performance in many prediction cases (Liao, Zheng, Grzybowski, & Yang, 2011). It is expressed as:

$$K(x, x_i) = e^{\frac{\|x - x_i\|^2}{2\sigma^2}} \tag{2.8}$$

whereσ^2 is a tuning parameter which is associated with RBF kernel. By using kernel function, it allows the data which are not linearly separable in input space to become linearly separable in high dimensional feature space. Another tuning parameter, which is regularization parameter, γ can be seen in (2.2).

2.6 Artificial Bee Colony Algorithm

ABC algorithm is motivated from the intelligent behavior of honey bees swarm (Karaboga, 2005). In the colony of artificial bee, there consists of three groups of bees, namely Employed Bees (EB), Onlooker Bees (OB) and Scout Bees (SB). Each

of them serves a significant role in achieving a common objective, which is maximizing the amount of nectar. Both OB and SB are also known as unemployed foragers (Karaboga, et al., 2012). The population size (number of colony members) are equally distributed, where the first half of the colony is comprised of the EB while the latter composed of the OB. Meanwhile, the SB is basically the EB which switch its role at certain condition. This situation is described later. The number of food sources, *SN* is equal to the number of EB. This means that one EB is associated with a single food source. In the ABC algorithm, the exploitation process is handled by EB and OB, while the exploration process is carried out by the SB.

Theoretically, the duty of EB is to search for new food sources, where in real-world problems, new food source positions represent the possible solutions. Later, the amount of nectar, which indicates the quality of solutions or fitness value, is calculated. Then, the information gathered by EB is shared with the OBs which are waiting in the hive or dance area. Here, decision making process is controlled by the OBs, where they have to exploit a nectar source depending on the information passed by the EBs. For the purpose of decision making, OBs watch waggle dance performed by the EBs, before doing the selection of food source according to the probability, which is proportional to the quality of that food source. Later, the OBs will employ herself at the most profitable source. The OBs also determine the food source to be abandoned and assign the associated EB as SB. In SB phase, the SB finds the new food sources at which point it once again becomes EB. It searches the space near the hive randomly.

Initially, all food source positions are discovered by SB. The food source positions represent the possible solutions of the problem under study. Suppose the solution space of the problem is *D*-dimensional, where *D* is the number of parameters to be optimized, viz. γ, σ^2 (see Section 2.5) and α (see section 2.7). Initial food sources positions are produced randomly within the range of the boundaries of the parameters of interest (Karaboga, 2005). It is expressed as follows:

$$x_{ij} = x_j^{\min} + rand(0,1)(x_j^{\max} - x_j^{\min}) \tag{2.9}$$

where $x_j^{\max}$ and $x_j^{\min}$ are the upper and lower bound of parameters of interest respectively. After initialization of the population, the fitness of food source is calculated and is defined as follows (Karaboga, 2005):

$$fit_i = \frac{1}{(1 + obj.Fun_i)} \tag{2.10}$$

where *obj.Fun* is the objective function. In this study, the objective function is MAPE (see Chapter 3, Section 3.5.2).

In EB phase, for each food source position, one EB is assigned to it. A new food source is obtained according to (2.11) (Karaboga, 2005):

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}) \tag{2.11}$$

where $i = 1, 2, \ldots, SN$ and $j = 1, 2, \ldots, D$.*k*is a randomly selected index in the colony which has to be different from *i*. φ is a random number within the range [-1, 1]. The position between x_{ij} and x_{kj} represent the distance from one food source to another.

As the difference between both positions decreases, the perturbation on the position x_{ij} decreases too.

In (2.11), if the values of parameters produced exceed their boundary, they are automatically shifted onto the boundary values (Karaboga & Akay, 2009). In ABC algorithm, only one parameter of each possible solution is changed at each iteration, which is determined randomly. After producing the new solution v_{ij}, its fitness is calculated and compared to the previous solution x_{ij}. Here, greedy selection scheme is applied, where, if the fitness of new solution is better than previous one, the bee memorizes the new solution; otherwise the previous solution is retained. If the second condition occurs, the number of trial which represent numbers of attempts in improving the solution increases by 1.

After all EB complete their searching process, the new food source positions obtained are shared with the OBs. This represents multiple interactions among the bees in the colony. An OB evaluates all the nectar information provided by EB and selects a food source with a probability related to its nectar quality. This indicates the fitness values of the solution. The selection is based on roulette wheel selection. The probability is defined as follows (Karaboga, 2005):

$$p_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i} \tag{2.12}$$

Where fit_i is the fitness of the solution *v.* SN is the number of food sources positions. By using (2.12), a good food source will attract more OB than at poor food source.

The food source selection by OB represents the positive feedback in ABC algorithm. Even though both EB and OB phases use similar way for food source exploitation, however, in EB phase, every solution will be updated while in OB phase, only the selected one will go through that process. Similarly as in EB phase, the OB produce modification on the selected food source position using (2.11) and the greedy selection process is performed after that. In SB phase, if the fitness of a found food source by EB has not been improved for a given number of trial (denoted by Limit), it is abandoned. This action represents the negative feedback in ABC algorithm and the EB of that food source becomes a SB and makes a random search using (2.9).The random movement by SB represents the fluctuations characteristic in SI. Basic steps of ABC algorithm are as follows:

1. *Initialize the food source positions (population)*

2. *Each EB is assigned on their food sources.*

3. *Each OB selects a source based on the quality of her solution, produces a new food source in selected food source site and exploits the better source.*

4. *Decide the source to be cast aside and appoints its EB as SB for discovering new food sources.*

5. *Memorize the best food source position (solution) found so far.*

6. *If requirement is met, output the best solution, otherwise repeat steps 2-5 until the maximum number of iteration is achieved.*

Data representation of ABC algorithm is as tabulated in Table 2.8.

Table 2.8: Data Representation of eABC-LSSVM Algorithm

Data Representation	**Variable**
D	Number of parameters to be optimized, which is 3, namely γ, σ^2 and α
New food source position	Possible solution to obtain optimized value of γ, σ^2 and α
Quality of nectar	Fitness function, which is the inverse of objective function
Objective Function	MAPE

2.7 Levy Probability Distribution

Levy Probability Distribution (LPD) was first introduced by Paul Levy in 1930 (Levy, 1937). It is a class of probability distributions which offer wider distribution (Lee & Yao, 2004; Wang & Li, 2008). The advantage of LPD relies on its tuning parameter, namely α where different values of α will yield different shapes of probability distribution (see Figure 2.3). This characteristic provides an effective strategy for searching since it will provide wider search space and reduce the possibility of revisiting a similar location (Lee & Yao, 2004). Hence, the better solution is likely to be achieved.

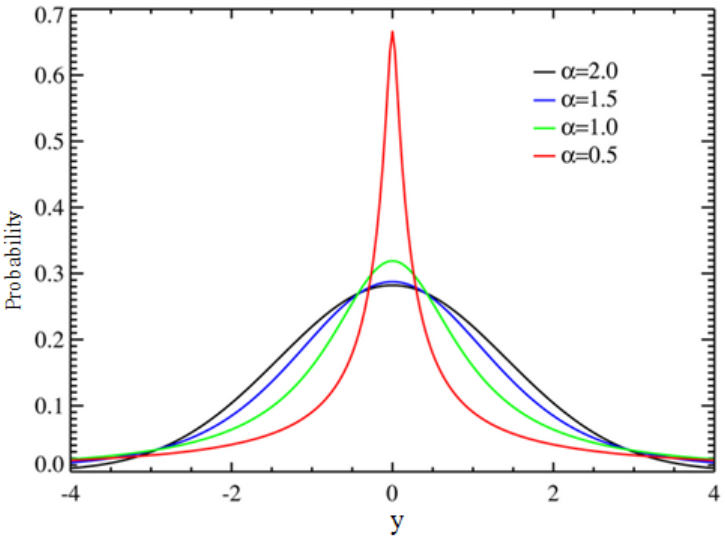


Figure 2.3: LPD with Different Value of α. Adapted from "Evolutionary Programming Using Mutations Based on Levy Probability Distribution,' by C.Y. Lee and X. Yao, 2004, IEEE Transactions on Evolutionary Computation, 8(1) p. 2. Copyright by IEEE Transactions on Evolutionary Computation. Adapted with permission.

The LPD has the following form (Levy, 1937):

$$L_{\alpha,\theta}(y) = \frac{1}{\pi}\int_0^{\infty} e^{-\theta q^{\alpha}} \cos(xy)dx \qquad y \in R \tag{2.13}$$

From (2.13), with y as the output and x is input, the distribution is symmetric with respect to y=0 and has 2 parameters, α and θ. α controls the shape of the distribution, subject to $0 < \alpha < 2$ while θ is the scaling factor satisfying $\theta > 0$.For α=1, it results to Cauchy distribution while when the $\alpha \to 2$, it becomes Gaussian distribution, which is excluded from LPD (Lee & Yao, 2004).Besides the fixed α value in both

distributions, the application of both Gaussian and Cauchy distribution is limited for searching in small and large area of search space respectively. Thus, the application of Gaussian and Cauchy distribution would limit the searching behavior (Lee & Yao, 2004).

Meanwhile, it is noted that finding different value of α for different problems is vital for sake the of the best results (Lee & Yao, 2004). With respect to that matter, the Levy distribution offers effective searching behavior which benefits from its adjustable parameter, α. With such dynamic characteristic, different α values can be obtained for different problems, depending on which searching stage it is in.

Since the analytic form of the integral(2.13) is unknown for general α, an algorithm for generating Levy random number is used (McCulloch, 1996).

For θ, it can be set to $\theta = 1$ without loss of generality. To describe this, rescale y to $y' = by$ with some constant b. Then, from (2.13), the following relation can be obtained (Lee & Yao, 2004):

$$L_{\alpha,\theta}(by) = \frac{1}{b} L_{\alpha,\theta'}(y) \tag{2.14}$$

where $\theta' = \theta b^{-\alpha}$. In particular, by setting $\theta' = 1$, (2.14) becomes (Lee & Yao, 2004):

$$L_{\alpha,\theta}(y') = \theta^{\frac{-1}{\alpha}} L_{\alpha,1}(y) \tag{2.15}$$

Implying that θ is nothing but an overall scaling factor. Thus, with the distribution of $\theta = 1$, the distribution of any other θ can be obtained (Lee & Yao, 2004).

With such characteristic, the addition of LPD in EC technique has become favorable among researcher (Lee & Yao, 2004; Rajasekhar, et al,. 2011). The integration of LPD encourages the optimization algorithm in searching for global minimum rather than local minimum. The implementation of LPD in EC algorithm is discussed in the following section.

2.7.1 Levy Probability Distribution in Evolutionary Computation

As to improve the efficiency of DE, the scaling factor of DE is adaptively controlled based on LPD, termed as LDE (He & Yang, 2012).In each iteration, the scaling parameter is produced based on LPD. Demonstrated on various range of benchmark functions, the performance of LDE against variants of DE is conducted based on 30 independent runs. Final results indicated that the LDE proved to be competitive than the other comparison algorithms.

In Hoorfar, Lakhani, and Jamnejad (2007), the application of LPD in Evolutionary Programming (EP) has been presented to solve the problem of antennas (termed as EP-LMO). Similarly like Rajasekhar, et al. (2011), the value of α is predetermined. As to examine the efficiency of EP-LMO, it was compared against EP variants. Upon completing the simulation task, the obtained results was in favor to EP-LMO.

Previously, the implementation of adaptive and non-adaptive LPD in EP has been presented by Lee and Yao (2004). Demonstrated on multivariate functional optimization, the comparison between the LEP and variants EP was evaluated based on minimum fitness. Empirical evidence suggested that LEP outperforms the other

identified techniques significantly. Similar study on this also has been presented in Lee and Yao (2001).

Meanwhile, the effectiveness of LPD in PSO has investigated by Richer and Blackwell (2006). The modification is done by a replacement of particle updating rule using LPD (termed as Levy PSO). Unlike in Rajasekhar, et al., (2011) and Hoorfar, et al (2007), the value of α *is* undetermined as to encourage the searching process. Realized on several benchmark functions, the Levy PSO performs better than the other variants of PSO.

An implementation of LPD in DE which combined the Extremal Optimization (termed as HEODE) has been demonstrated by Fu and Yu (2009). The inducing of LPD in DE is to address the tendency of DE to fall into local minimum, hence improve the searching capability of DE. The HEODE was used to train the ANN in dealing with soft-sensor of jet fuel endpoint of main fractionator of hydro cracking unit. Compared against another variant of DE, the HEODE produced better results as compared to competitor under study.

Based on the above discussion, it can be seen that with such a unique characteristic of the Levy mutation, it has been proven to be favorable in enhancing the global search ability of EC algorithm. Summary of this section is presented in Table 2.9.

Table 2.9: Summaries of Levy Probability Distribution in Evolutionary Computation Algorithm

Author	Research Area	Independent Variables	Method	Data Frequency	Evaluation Metric
He & Yang (2012)	Modification of standard DE	-	LDE	-	Average obtained minimum error
Hoorfar, et al. (2007	Modification of standard EP	-	LMO	-	Fitness evaluation
Lee & Yao (2004)	Modification of standard EP	-	LEP	-	Minimum fitness
Richer & Blackwell (2006)	Modification of standard PSO	-	Levy PSO	-	Average fitness
Fu & Yu (2009)	Modification of standard DE		HEODE		Best fitness, optimal rate, mean of iterations, mean of fitness

2.8 Summary

This chapter reviewed related studies which covered issues on prediction utilizing various techniques, covering from statistical to CI techniques. Due to linear structure of statistical technique, it is regarded as inefficient in dealing with prediction of the time series data of interest. This has led to the application of ANN. Nonetheless, several demerits of ANN make it suffer generalization. By using LSSVM, the disadvantage of ANN is addressed. However, efficient optimization algorithm is needed in order to increase the capability of LSSVM in prediction task. For that matter, particular concerns are given on optimizing the LSSVM hyper parameters using EC algorithm which has been proven to overcome the limitation of manual and CV approach. Meanwhile, the application of ABC as an individual in prediction is reviewed and its merit and demerit are highlighted. In optimizing the embedded control parameters of machine learning, the ABC has been favorable due its simplicity in application. However, several limitations of the ABC algorithm is identified such as limitation in searching behavior and tendency to cause the LSSVM algorithm face with over fitting and under fitting problem.

Based on literature review that has been done, the proposed algorithm is designed, which is presents in the next chapter. The discussion covered the discussion on implemented methodology which consist of four main phases, viz. data collection, designing the proposed model, developing the proposed model and evaluation. In algorithm design, the said limitations of ABC algorithm are discussed and addressed accordingly.

CHAPTER THREE
METHODOLOGY

3.1 Introduction

This chapter presents the methodology implemented in this study. The stages incorporated of four major tasks, viz. data collection and preparation, algorithm design, algorithm development, and evaluation. The simplified form of the methodology implemented is illustrated in Figure 3.1

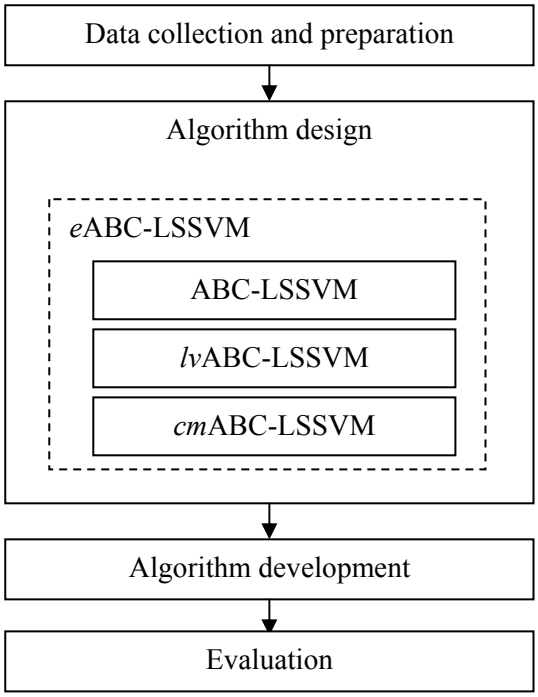


Figure 3.1: Methodology

In this thesis, three set of time series data sets have been utilized, namely Data set A, B and C. The description of these data sets are provided in the next section.

3.2 Data Collection and Preparation

The first time series data set employed in this study is the one utilized in Malliaris and Malliaris (2008), termed as Data Set A. The description of the data set is tabulated in Table 3.1.

Table 3.1: Data Set A

Set	Period	Input	Symbol
A	1st December 1997 - 27th November 2002	Crude oil price	CL
		Heating oil price	HO
		Gasoline price	HU
		Propane price	PN

Data set A consist of 1248 days (trading days) and is available at *Barchart* website ("Barchart," 2011). Sample of the data set is shown in Table 3.2.

Table 3.2: Sample of Data Set A

Date	CL	HO	HU	PN
1/12/1997	18.6300	0.5144	0.5338	0.3260
2/12/1997	18.7000	0.5190	0.5316	0.3188
3/12/1997	18.6000	0.5220	0.5328	0.3213
4/12/1997	18.5900	0.5218	0.5272	0.3198
5/12/1997	18.7000	0.5276	0.5262	0.3200

The second time series data employed in this study, term as Data Set B is as described in Table 3.3

Table 3.3: Data Set B

Set	Period	Input	Symbol
B	2nd December 2002 –31st October 2007	Crude oil prices	CL
		Heating oil prices	HO
		Gasoline prices	HU
		Propane prices	PN
		Interest rate	IR

Similarly as in Data set A, the CL, HO, HU and PN are obtained from *Barchart* website while the interest rate data is downloaded from Board of Governors of the Federal Reserve System website ("Board of Governors of the Federal Reserve System," 2012). Data set B consist of 1225 days (trading days) and examples of the data is shown in Table 3.4.

Table 3.4: Sample Data Set B

Date	CL	HO	HU	PN	IR
2/12/2002	27.2300	0.7747	0.7212	0.4875	1.2200
3/12/2002	27.2800	0.7755	0.7370	0.4913	1.2100
4/12/2002	26.7300	0.7437	0.7150	0.4900	1.2000
5/12/2002	27.2800	0.7525	0.7340	0.4938	1.2000
6/12/2002	26.9300	0.7441	0.7181	0.4938	1.1900

As to test the efficiency of the proposed technique, the metal data set, viz. gold prices, termed as Data Set C, was utilized. The inputs employed are tabulated in Table 3.5.

Table 3.5: Data Set C

Set	Period	Input	Symbol
C	2nd January 2009 – 3rd October 2012	Gold price	GC
		Silver price	SI
		Palladium price	PA
		Crude oil price	CL

The data which was also collected from *Barchart* website is of prices for 943 days (trading days). Sample of Data set C is as provided in Table 3.6.

Table 3.6: Sample of Data Set C

Date	GC	SI	PA	CL
2/1/2009	875.9500	11.5804	186.0000	46.3400
5/1/2009	859.5000	11.2410	186.0000	48.8100
6/1/2009	864.1600	11.4700	199.0000	48.5800
7/1/2009	842.9400	11.0340	205.0000	42.6300
8/1/2009	857.2500	11.1332	198.0000	41.7000

Besides the above inputs, another three inputs that are considered to be added to the prediction technique are crude oil price, foreign exchange rates (U.S. dollar against the seven major currencies, viz. Euro Area, Canada, Japan, United Kingdom, Switzerland, Australia and Sweden) and interest rates (Hammoudeh, et al., 2008). The justification of considering these additional inputs is because they are common macroeconomic factors that influence the price of commodities (Hammoudeh, et al., 2008; Zhang & Wei, 2010). However, prior to adding those inputs to prediction technique, an initial stage is performed in order to determine the range of correlation between the above common factors with the main input data. They will be

considered as input to the prediction technique if the predefined condition is fulfilled (see section 3.2.1). In order to prepare the data for experiment, all the datasets need to be tested for correlation and non linearity. These tests are discussed in the following sections.

3.2.1 Test for Correlation

Correlation is a measure that can be used to describe the relationship that exist among the data (Wang, et al., 2010). The purpose of this test is conducted is to identify relevant attributes to be used as independent inputs for the proposed model. Only correlated attributes will be used in the model. In this study, the price correlation coefficient among time series data of interest was tested using Pearson Product Moment Correlation Coefficient (PPMCC) (Lomax, 2007). For this matter, SPSS 15.0 software was used (Colman & Pulford, 2008). The definition for PPMCC is defined as r= 0.8 - 1.0 which indicates high correlation (Ahmed Farid & Salahudin, 2010). The higher value of correlation coefficient reflects the level of significance among the input features.

Even though crude oil price, exchange rate and interest rate are fundamental indicators which influence commodity prices, specifying the correlation between them is also important. This is to ensure the significant of the factor to the commodity prices of interest in the specific period since the price correlation among commodities prices varies from one period to another (Wang, et al., 2010). Thus, for prediction purposes, by neglecting irrelevant factors, it will be beneficial for the sake

of efficiency in prediction (Kotsiantis, Kanellopoulos, & Pintelas, 2006). Table 3.7 to 3.9 show the correlation between selected inputs for energy fuel prices Data Set A, B and C respectively. The results for correlation of these inputs with other input are provided in Appendix A.

Table 3.7: Price Correlation among Input Data Set A

	CL	**HO**	**HU**	**PN**
CL	1	-	-	-
HO	0.9649	1	-	-
HU	0.9597	0.9262	1	-
PN	0.8422	0.8473	0.8812	1

Table 3.8: Price Correlation among Input Data Set B

	CL	**HO**	**HU**	**PN**	**IR**
CL	1	-	-	-	-
HO	0.9541	1	-	-	-
HU	0.9833	0.9552	1	-	-
PN	0.9637	0.9337	0.9590	1	-
IR	0.8495	0.8486	0.8442	0.7920	1

Table 3.9: Price Correlation among Input Data Set C

	GC	**SI**	**PA**	**CL**
GC	1	-	-	-
SI	0.9071	1	-	-
PA	0.8449	0.9076	1	-
CL	0.8107	0.8372	0.8642	1

3.2.2 Test for Non-linearity

The purpose of conducting non-linearity test is to support the justification of employing a non-linear prediction technique (i.e. LSSVM). In this study, the Brock, Dechert, and Scheickman (BDS, hereafter) test (Brock, Scheinkman, Dechert, & LeBaron, 1996) is employed using EViews 7 (Belaire & Contreras, 2002). The BDS tests is a standard test for non-linearity effect in the data. Based on BDS, when the z-Statistic, $z > 2$ at significance level of 0.05, it is evidence that the time series data exhibit non-linear behaviour (Vlad, 2010).

The BDS test results for the time series data of interest are summarized in Appendix B. From the tables, it is learned that the z-Statistic are significantly greater than the critical values. This indicates that the time series data of interest demonstrate a significant nonlinear behavior which supports the usage of a nonlinear technique for prediction task. The results obtained also reinforce claims in literatures that state CL price is govern by nonlinear features (Abdullah & Zeng, 2010; Alizadeh, et al., 2012; Bao, et al., 2011; Haidar &Wolf, 2011). This is due to these data are influenced by various factors which contribute to its complexity (Abdullah & Zeng, 2010; Jammazi & Aloui, 2012). Similar results also can be seen in other three variables, namely HO, HU and PN. This is not surprising since they are not only the refined products of CL but also possess high correlation with CL price (Malliaris & Malliaris, 2008). This characteristic is also featured in GC price (Liu, 2009) (see Appendix B).

3.2.3 Data Normalization

The success of machine learning algorithms rely on the quality of the data they operate on (Kotsiantis, et al., 2006).Thus, data transformation such as data normalization is essential in order to standardize the values of all input and output involved into specific range. This ensures that the larger input values do not dominate the smaller input values which will make the training of prediction model become difficult. Hence, helps to improve the prediction accuracy and efficiency of data mining algorithm (Al-Shalabi, Shaaban, & Kasasbeh, 2006). In this study, Min Max Normalization is employed (Al-Shalabi, et al., 2006). The definition of Min Max Normalization is shown in the following equation:

$$v' = \left((v - \min_a)/\left(\max_a - \min_a\right)\right) * \left(new\max_a - new\min_a\right) + new\min_a \quad (3.1)$$

where, v' = New value for variable v

v = Current value

$\min_a$ = Minimum value in data set

$\max_a$ = Maximum value in data set

$new\max_a$ = New maximum value in data set

$new\min_a$ = New minimum value in data set

Min Max Normalization maps a value of v of A to v' in the range [$new\max_a$, $new\min_a$] (i.e. 0 - 1) by solving the equation above. The normalized input for sample in Table 3.2, 3.4 and 3.6 are as tabulated in Table 3.10, 3.12 and 3.13 respectively.

Table 3.10: Samples of Normalized Input for Data Set A

Date	CL	HO	HU	PN
1/12/1997	0.2989	0.2804	0.2970	0.1803
2/12/1997	0.3015	0.2860	0.2943	0.1698
3/12/1997	0.2977	0.2896	0.2957	0.1734
4/12/1997	0.2973	0.2894	0.2889	0.1712
5/12/1997	0.3015	0.2965	0.2877	0.1715

Table 3.11: Samples of Normalized Input for Data Set B

Date	CL	HO	HU	PN	IR
2/12/2002	0.0289	0.0347	0.0211	0.0139	0.0988
3/12/2002	0.0296	0.0351	0.0299	0.0175	0.0965
4/12/2002	0.0216	0.0221	0.0177	0.0162	0.0941
5/12/2002	0.0296	0.0257	0.0282	0.0198	0.0941
6/12/2002	0.0245	0.0222	0.0194	0.0198	0.0918

Table 3.12: Samples of Normalized Input for Data Set C

Date	GC	SI	PA	CL
2/1/2009	0.0593	0.0263	0.0060	0.1546
5/1/2009	0.0441	0.0173	0.0060	0.1855
6/1/2009	0.0484	0.0234	0.0257	0.1826
7/1/2009	0.0289	0.0119	0.0347	0.1082
8/1/2009	0.0421	0.0145	0.0242	0.0966

3.3 Algorithm Design

In this phase, the algorithm of *e*ABC-LSSVM is designed accordingly which is aligned with the objectives of study. This phase started with the design of ABC-LSSVM algorithm, followed by the design for *lv*ABC-LSSVM, *cm*ABC-LSSVM and finally the *e*ABC-LSSVM algorithm which combined the *lv*ABC-LSSVM and *cm*ABC-LSSVM algorithms.

3.3.1 ABC-LSSVM

For the purpose of parameter tuning, ABC algorithm is utilized to automatically tune the hyper-parameters of LSSVM. To describe how the ABC-LSSVM works, LSSVM is embedded in the ABC algorithm as a fitness function evaluation. Thus, the optimized value of hyper-parameters can be obtained after a maximum number of iteration has been achieved. In this study, the objective function is served by MAPE, where the lower the MAPE, the better the prediction accuracy (see section 3.5.2).

3.3.1.1 Flow of ABC-LSSVM

Mainly, there are four phases in ABC-LSSVM algorithm, namely initialization phase, EB phase, OB phase and SB phase. Prior to the initialization, the normalized training and validation data is fed to the prediction model.

3.3.1.1.1 Initialization Phase

The initialization of the food source positions is done by making use of (2.9) (see Chapter 2, section 2.6). The food source positions indicate the possible solutions which are the parameters to be optimized, namely the hyper-parameters, γ and σ^2. In this study, the number of possible solutions position is set to 10. Figure 3.2 illustrates the example of possible solutions in X.

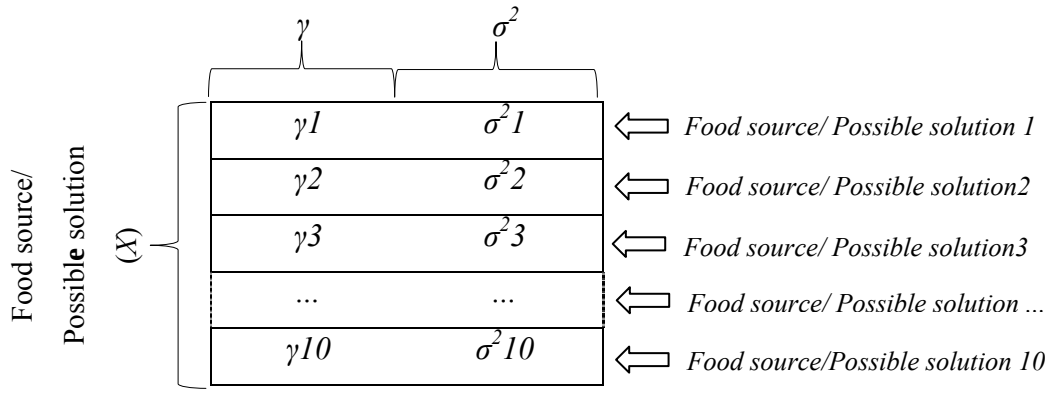


Figure 3.2: Representation of Food Source Position as Possible Solutions in X

The fitness function of the hyper-parameters is evaluated based on training and validation set using LSSVM function which is integrated in ABC algorithm. In this study, the fitness function evaluation is as defined in (2.10) (see Chapter 2, section 2.6). Table 3.13 shows the example of initial fitness function.

The goal is to find the ideal combination of hyper-parameters (i.e. possible solutions) that will minimize the objective function, which is guided by MAPE (see section 3.5.2). Before the EB phase begins, the cycle number is set to 1.

Table 3:13: Initial Fitness Functions

Possible Solutions in *X*	**MAPE (%)**	**Fitness Functions**
Possible solution 1	*m1*	$f1 = 1/(1+m1)$
Possible solution 2	*m2*	$f2 = 1/(1+m2)$
Possible solution 3	*m3*	$f3 = 1/(1+m3)$
...	...	...
Possible solution 10	*m10*	$f10 = 1/(1+m10)$

3.3.1.1.2 EB Phase

Each EB produces a new solution by using (2.11) (see Chapter 2, section 2.6). If the produced new solution surpasses the boundary value, it is automatically shifted onto the boundary value. Later, the newly produced solution is evaluated based on training and validation sets using the LSSVM. Then, a comparison between the fitness function of the new and old solution is performed. If the fitness function of the new solution is superior to the old one, the bees memorize the new solution. With that, the number of trials which represent numbers of attempts in improving the solution is reset to 0. However, if the fitness function of the old solution is better than the new one, the old solution is retained and the number of trial by the respective EB increases by one. The maximum number of trial is indicated by *Limit* parameter (see section 3.5.1.3). For example, the comparison between the fitness function of old solution, *f2* and the fitness function of new solution, *new_f2* is in favor to the old one, *f2*. For this matter, the old solution is memorized and the associated EB has to increase its trial counter to 1. The example of solution obtained is as shown in Table 3.14.

Table 3.14: EB Phase

Possible Solutions in *X*	**Old Fitness Functions**	**New Fitness Functions**	**Fitness function after Greedy Selection**
Possible solution 1	*f1*	*new_f1*	*new_f1*
Possible solution 2	*f2*	*new_f2*	*f2*
Possible solution 3	*f3*	*new_f3*	*new_f3*
...	...	...	...
Possible solution 10	*f10*	*new_f10*	*new_f10*

After all EB complete the search process, the solutions produced in *X* are shared with the OB. The probability value for each solution in *X* is computed, which is related to their fitness (see Table 3.15). The formula is as defined in (2.12) (see Chapter 2, section 2.6). For example, the highlighted cell in Table 3.15 represents the selected possible solutions to be modified by a certain OB, let say OB*n,* while the rest is ignored for modification process.

Table 3.15: Probability

Possible solutions in *X*	**Fitness Functions**	**Probability**
Possible solution 1	*new_f1*	*p1*
Possible solution 2	*f2*	*p2*
Possible solution 3	*new_f3*	*p3*
...	...	...
Possible solution 10	*new_f10*	*p10*

3.3.1.1.3 OB Phase

An OB chooses the solutions provided by EB in *X* based on the probability value which is related to the fitness values of the solution. A solution is selected for modification in the OB phase if the value of probability is higher than the generated random number (within the range [0, 1]). Otherwise, it is neglected. In this example, the selected possible solutions in *X* that are chosen to for modification by OB*n* are *Possible solution 1*, *2* and *3* (see Table 3.15).

After the selection is made, the OB will modify the selected solution using (2.11) (see Chapter 2, section 2.6). As in the EB phase, the newly produced solution is

checked whether it exceeds the boundary value or not. If yes, it is automatically shifted to the boundary values. Later, the newly produced solution is evaluated based on LSSVM. Similarly as in EB phase, the comparison between new and old fitness function of the solution is performed, where, if the new is better than the old one, the new solution is retained. Otherwise, the old one is kept and the trial counter is incremented by 1.

By referring to Table 3.16, the highlighted cells represent the chosen solutions that have been modified by OB*n*, while the rest of the solutions are ignored for modification process.

Table 3.16: OB Phase

Possible Solutions	**Old Fitness Functions by EBs**	**New Fitness Functions by OB *n***	**Fitness function after Greedy Selection by OB *n***
Possible solution 1	*new_f1*	*Latest_f1*	*Latest_f1*
Possible solution 2	*f2*	*Latest_f2*	*Latest_f2*
Possible solution 3	*new_f3*	*Latest_f3*	*Latest_f3*
...	...	...	...
Possible solution 10	*new_f10*	*Not selected for modification, thus new_f10 is retained in X*	

3.3.1.1.4 SB Phase

In SB phase, the number of trial is checked. If the number of trial exceeds the *Limit* value, the responsible EB is introduced as SB. A new solution is produced by making use of (2.9) (see Chapter 2, section 2.6) to replace the abandoned solution. Otherwise, the best solution so far is memorized.

These steps (3.3.1.1.2, 3.3.1.1.3, and 3.3.1.1.4) are continuously carried out until the maximum iteration number (*MCN*) is achieved. After the iterations end, the obtained optimal γ and σ^2 are used in testing set using LSSVM for prediction purposes. The flow and algorithm of ABC-LSSVM are given in Chapter 4.

3.3.2 Enrich the Exploitation Process in Search Space (*lv*ABC)

In ABC algorithm, it can be seen that the exploitation process in both EB and OB phases are highly dependent on a single equation, which represents the applied strategy (see Chapter 2, section 2.6). Applying a single equation for different phases limits the searching behavior significantly. Such limitation may produce a poor optimization results and lead to local minimum. Thus, to enrich the searching behavior of the bees, in the EB phase, instead of applying (2.11) (see Chapter 2, section 2.6), the following equation is introduced:

$$v_{ij} = x_{ij} + \left(x_{ij} - x_{kj}\right) \times L \tag{3.2}$$

On the other hand, in OB phase, the following equation is used as to replace (2.11):

$$v_{ij} = x_{ij} + L \tag{3.3}$$

where v_{ij} represents new solution, x_{ij} indicates old solution, x_{kj} represents neighborhood solution which is randomly selected. This means that x_{ij} and x_{kj} indicate different solution positions. The difference between both position solutions is the distance from one food source to another (i.e. impact of neighborhood). Meanwhile, L is represented by a random number generated by LPD. The main objective is to enrich the searching process by applying different equations (i.e strategies) in both phases. With the inducing of Levy mutation (see Chapter 2,

section 2.7) in each strategy, it will encourage large step length in the search space and perform wider exploration at any stage of convergence. This provides efficient driving force in producing more distinct values in produced solution and contributes in escaping local minimum (Richer & Blackwell, 2006).

As shown in (3.2) and (3.3), the similarity in EB and OB phases is, Levy mutation is induced in producing new solution. Meanwhile, the difference is the impact of neighborhood is retained in EB while in OB phase, it is omitted. The justification is, in ABC, as the searching process approaches the optimal solution, the step length is reduced adaptively (Karaboga & Akay, 2009). However, if the found solution is not the optimal solution, this would lead the algorithm to suffer premature convergence and trapped in local minimum. For that matter, to omit the impact of neighborhood in both EB and OB phases may cause loss of the important information. On the other hand, to keep it in both phases would cause the algorithm trapped in local minimum. Due to that matter, as the results produced in EB still has a chance for modification in OB phase, in *lv*ABC, the impact of neighborhood is keep in EB phase.

Meanwhile, in OB phase, as the OB is responsible to improve the selected solution taken from EB, the impact of neighborhood is omitted. This is due to the risk of false signal that might be given by the impact of neighborhood cannot be taken since the OB is the last phase in exploitation process. Thus, with the inducing of Levy mutation and well balance between EB and OB phases, the searching behavior of

ABC is enriched which is beneficial in avoiding local minimum. In this study, all parameters involved are automatically tuned by *e*ABC-LSSVM.

In addition, in searching for global optimal, it is an advantage when the global optimal is adequately far away from current solution found. This is not only assists the technique to converge in acceptable manner, but helps in avoiding the technique from being trapped in local minimum as well (Wang & Li, 2008). The algorithm of *lv*ABC-LSSVM is given in Chapter 5.

3.3.3 Preventing Over fitting and Under fitting (*cm*ABC)

In ABC algorithm, whenever the value of parameters of interest exceeds the pre-defined boundary values, it will be shifted to the boundary value (Karaboga & Akay, 2009). Nonetheless, in LSSVM, the maximum and minimum value of hyper-parameters do not necessary mean the optimal (Lendasse, et al., 2005; Wu, et al., 2007). Besides, it may result in the prediction technique being vulnerable to over fitting and under fitting (Fu, et al., 2010;Lendasse, et al., 2005; Wu, et al., 2007). This will consequently affect the generalization of LSSVM in prediction task.

With respect to that matter, an enhancement is made in decision making. This is done by adapting a conventional mutation. By applying a mutation strategy, the model is induced to explore other areas in order to look for global optimal rather than local minimum (Haupt & Haupt, 2004). This operation is executed by multiplying the generated random number with the range of predefined boundary. In

this study, the boundaries are set to the range of [1, 1000]. The equation is expressed as follows (Haupt & Haupt, 2004):

$$new\_param = (ub - lb) * rand\_num \quad (3.4)$$

where *new_param* = new parameter

rand_num= random number between [0, 1]

ub = upper bound

lb = lower bound

The *cm*ABC-LSSVM algorithm can be seen in Algorithm 5.2 which is presented in Chapter 5.

3.3.4 *e*ABC-LSSVM

In any meta heuristic algorithm, there are two critical tasks which ensure the global optimal is achievable, viz. exploitation and exploration (Yang, 2010). In ABC algorithm, the exploitation task requires the bees to focus on the searching in local region. This is done by fully exploiting and applying the information obtained from previous solution to find better solutions. On the other hand, the exploration is where the solution found by the bees is diverted for the purpose to further explore the search space in global region. A well-balanced between exploitation and exploration process is an advantage in achieving global optimal (Yang, 2010).

However, the basic structure of ABC is good at exploration but poor in exploitation. This is due to the dependency of ABC in a single equation for different phases in exploitation process (Babayigit & Ozdemir, 2012; Sharma, et al., 2011, Gao, et al., 2013). This situation would affect the optimization process of ABC and leads the algorithm to fall into local minimum (Gao & Liu, 2012; Lee & Cai, 2011).

Meanwhile, it is also observed that, in ABC algorithm, whenever the value of parameters of interest exceeds the pre-defined boundary values, it will be shifted to the boundary value (Karaboga & Akay, 2009). Nonetheless, it is worth to remind that in LSSVM, the maximum and minimum value of hyper-parameters do not necessary mean the optimal (Wu, et al., 2007). Besides, it may result in the prediction technique vulnerable with over fitting and under fitting, which consequently affects the generalization capability of LSSVM in prediction task.

With respect to that matter, the *e*ABC-LSSVM algorithm is designed to address the highlighted problems. As described in section 3.3.2, the problem of local minimum is addressed using Levy mutation and different strategies in EB and OB phase, (termed as *lv*ABC). Meanwhile, the problem of over fitting and under fitting is addressed using conventional mutation (termed as *cm*ABC) (see section 3.3.3).

3.3.4.1 Flow of *e*ABC-LSSVM

The *e*ABC-LSSVM algorithm consists of four phases, namely initialization, EB, OB and SB phase. Before the initialization, the normalized training and validation data is fed to the prediction model.

3.3.4.1.1 Initialization Phase

The food source positions are initialized using (2.9) (see Chapter 2, section 2.6). This represents the possible solutions of the problem under study which are the parameters to be optimized, namely γ, σ^2, and α. The number of possible solutions is equal to number of EB. In this study, it is set to 10. Figure 3.3 shows the example of possible solutions which are the parameters of interest in *X*.

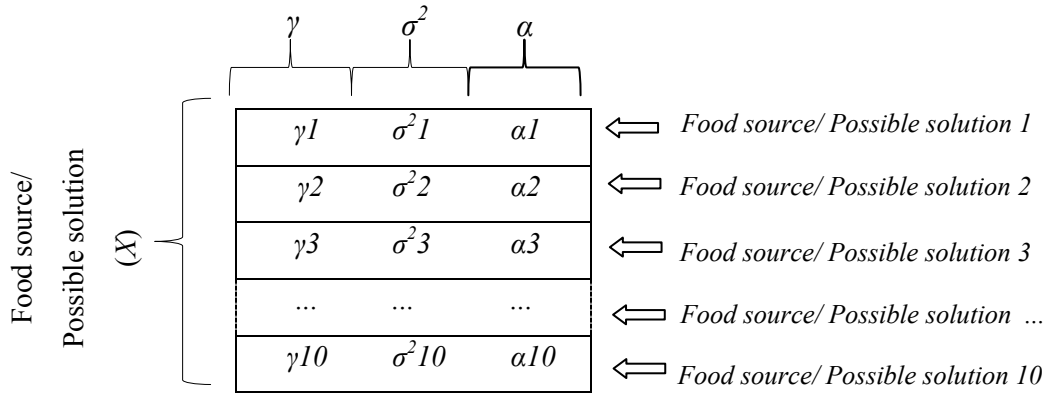


Figure 3.3: Representation of Food Source Position as Possible Solutions in X

For fitness function evaluation, the combination of γ and σ^2(i.e. hyper-parameters) is evaluated based on validation set, after the training process, using the LSSVM function which is embedded in the ABC algorithm. In this study, the fitness function evaluation is calculated using (2.10) (see Chapter 2, section 2.6). The example of initial fitness function is as shown in Table 3.17.

Table 3.17: Initial Fitness Functions

Possible Solutions in *X*	**MAPE (%)**	**Fitness Functions**
Possible solution 1	*m1*	$f1 = 1/(1+m1)$
Possible solution 2	*m2*	$f2 = 1/(1+m2)$
Possible solution 3	*m3*	$f3 = 1/(1+m3)$
...	...	...
Possible solution 10	*m10*	$f10 = 1/(1+m10)$

Meanwhile, the α is used to assist the *e*ABC-LSSVM algorithm in obtaining optimal value of hyper-parameters, but is not used directly in fitness function evaluation in LSSVM. The goal is to find the best combination of LSSVM hyper-parameters that will minimize the objective function, which is served by MAPE (see section 3.5.2). Before entering the EB phase, the cycle number is set to 1.

3.3.4.1.2 EB Phase

In this phase, the new solution is produced using (3.2). If the obtained new solution exceeds the boundary value, (3.4) is applied. After the new solution is produced, its fitness function is evaluated based on training and validation sets using the LSSVM. Later, the newly produced solution is compared against the old one based on greedy selection. If the fitness value of new solution is better than the old one, the new solution is memorized and the number of trial is reset to 0. However, if the fitness value of old solution is better than the new solution, the old solution is retained and the number of trial increases by 1. The maximum number of trial is indicated by *Limit* parameter (see section 3.5.1.3). Let's say the fitness function of the old solution, *f2* is better than the fitness function of the new solution, *new_f2*, thus, the

old solution is kept and the responsible EB has to increase its trial counter to 1. Table 3.18 shows the example of results produced in EB phase.

Table 3.18: EB Phase

Possible Solutions in *X*	**Old Fitness Functions**	**New Fitness Functions**	**Fitness function after Greedy Selection**
Possible solution 1	*f1*	*new_f1*	*new_f1*
Possible solution 2	*f2*	*new_f2*	*f2*
Possible solution 3	*f3*	*new_f3*	*new_f3*
...	*...*	*...*	*...*
Possible solution 10	*f10*	*new_f10*	*new_f10*

After all EBs have completed the searching process, the solutions obtained in *X* are shared with OB. Then, the probability values for each solution in *X* by means of their fitness are calculated for selection process in OB phase. For example, the highlighted cell in Table 3.19 indicates the selected possible solutions to be modified by a certain OB, let's say OB*n.* Then the rest is neglected for modification process.

Table 3.19: Probability

Possible solutions in *X*	**Fitness Functions**	**Probability**
Possible solution 1	*new_f1*	*p1*
Possible solution 2	*f2*	*p2*
Possible solution 3	*new_f3*	*p3*
...	*...*	*...*
Possible solution 10	*new_f10*	*p10*

3.3.4.1.3 OB Phase

Each OB evaluates all the solutions produced by EB in *X* and chooses a solution with a probability value which is related to the fitness of the solution in *X* using (2.12). A solution is selected for modification in OB phase when the probability value is greater than the generated random number (within the range [0, 1]), then the solution is selected for modification process in OB phase. Otherwise, it is neglected. As the fitness value of the solution increases, the probability of the solution to be chosen increases too. In our example, the selected possible solutions in *X* that are chosen to be modified by OB*n* are *Possible solution 1*, *2* and *3* (see Table 3.19).

After choosing the selected food source from *X*, the OB will modify the selected solution using (3.3). Similarly as in EB phase, the new produced solution is checked whether it surpasses the boundary value or not. Equation (3.4) is applied if the solution surpasses the boundary values. Later, the newly produced solution is evaluated based on LSSVM. As in EB phase, the comparison between the fitness function of new and old solution is performed. If the fitness function of new solution is better than the fitness function of the old solution, the old solution is replaced and trial number is reset to 0. Otherwise, the old solution is retained and trial number is increased by 1. Note that in this example, only *Possible solution 1*, *2* and *3* are selected by the OB*n* for modification process (see Table 3.20).

Table 3.20: OB Phase

Possible Solutions	**Old Fitness Functions by EBs**	**New Fitness Functions by OB *n***	**Fitness function after Greedy Selection by OB *n***
Possible solution 1	*new_f1*	*Latest_f1*	*Latest_f1*
Possible solution 2	*f2*	*Latest_f2*	*Latest_f2*
Possible solution 3	*new_f3*	*Latest_f3*	*Latest_f3*
...	...	...	...
Possible solution 10	*new_f10*	*Not selected for modification, thus new_f10 is retained in X*	

3.3.4.1.4 SB Phase

In SB phase, the trial number is checked. If the *trial* number exceeds the *Limit* value, the responsible EB is assigned as SB and a new solution will be produced using equation (2.9) (see Chapter 2, section 2.6). Otherwise, the best solution so far is memorized.

The processes in 3.3.4.1.2, 3.3.4.1.3 and 3.3.4.1.4 are continuously proceed until the maximum iteration number (*MCN*) is achieved. In this study, the *MCN* is set to 100. After the iterations end, the obtained optimal hyper-parameters are used in testing set using LSSVM for prediction purposes. The flow of *e*ABC-LSSVM and the algorithm are given in Chapter 5.

3.4 Algorithm Development

The *e*ABC-LSSVM is developed using LS-SVMlab Toolbox (Pelkmans et al., 2002) and is executed using MATLAB platform on Intel® Core™ i3-2330M processor, CPU @ 2.20GHz with 6.00 GB of RAM, in 64-bit Windows 7 operating system.

3.5 Evaluation

This section describes experimental setup of the conducted experiment, utilized performance evaluation metrics and benchmarking techniques.

3.5.1 Experimental Setup

This phase consists of arrangement of input and output, data proportion for training, validation and testing, and as well as properties setting for the identified prediction techniques.

In this study, it has involved 36 experiments which has been conducted using ABC-LSSVM, *lv*ABC-LSSVM, *cm*ABC-LSSVM and *e*ABC-LSSVM on three groups of data set, namely Data Set A, B and C. In Data Set A and B, the prediction task was conducted to predict four time series data: the price of crude oil, heating oil, gasoline and propane. Meanwhile in Data Set C, the prediction task was performed to predict the gold price.

The experiment using ABC-LSSVM is conducted to investigate the capability of single ABC in optimizing the LSSVM hyper parameters. Upon completing the experiment, the limitation of ABC in searching behavior is identified. For this matter, a series of experiment using *lv*ABC-LSSVM are conducted. As to address the over fitting and under fitting problem, the experiments utilizing *cm*ABC-LSSVM are performed, which are realized on the same time series data. As to see whether the combination of *lv*ABC-LSSVM and *cm*ABC-LSSVM capable to produce better

generalization and escape from premature convergence, the experiments using *e*ACB-LSSVM are conducted.

3.5.1.1 Input and Output Variables

The input data is as the ones utilized in Malliaris and Malliaris (2008). For each prediction model, the input variables are daily closing spot price, percent change in daily closing spot price from the previous day, standard deviation over the previous 5 trading days and standard deviation over the previous 21 trading days (Malliaris & Malliaris, 2008). For Data Set A and B, the output variables are daily spot price of crude oil, heating oil, gasoline and propane from day 21 onwards (CL21, HO21, HU21 and PN21 respectively) while for Data Set C, the output are gold price from day 21 onwards (GC21). Tables 3.21 to 3.23 indicate the variables assigned to the inputs involved for Data Set A, B and C respectively.

Table 3.21: Assigning Input and Output Variable for Data Set A

Input	Variable	Output
Daily closing spot price of crude oil, heating oil, gasoline and propane	CL, HO, HU, PN	CL21, HO21, HU21, PN21
Percent change (%Chg) in daily closing spot prices from the previous day of CL, HO, HU and PN	CL%Chg, HO%Chg, HU%Chg, PN%Chg	
Standard deviation (sd) over the previous 5 days trading days of CL, HO, HU and PN	CLsd5, HOsd5, HUsd5, PNsd5	
Standard deviation (sd) over the previous 21 days trading days of CL, HO, HU and PN	CLsd21, HOsd21, HUsd21,PNsd21	

Table 3.22: Assigning Input and Output Variable for Data Set B

Input	Variable	Output
Daily closing price of crude oil, heating oil, gasoline, propane and daily rate of interest	CL, HO, HU, PN, IR	CL21, HO21, HU21, PN21
Percent change (%Chg) in daily closing spot prices from the previous day of CL, HO, HU, PN and percent change in daily rate from the previous day of IR	CL%Chg, HO%Chg, HU%Chg, PN%Chg, IR%Chg	
Standard deviation (sd) over the previous 5 days trading days of CL, HO, HU, PN and IR	CLsd5, HOsd5, HUsd5, PNsd5, IRsd5	
Standard deviation (sd) over the previous 21 days trading days of CL, HO, HU, PN and IR	CLsd21, HOsd21, HUsd21,PNsd21, IRsd21	

Table 3.23: Assigning Input and Output Variable for Data Set C

Input	Variable	Output
Daily closing price of gold, silver, palladium and crude oil	GC, SI, PA, CL	GC21
Percent change (%Chg) in daily closing spot prices from the previous day of GC, SI, PA and CL	GC%Chg, SI%Chg, PA%Chg, CL%Chg	
Standard deviation (sd) over the previous 5 days trading days of GC, SI, PA and CL	GCsd5, CIsd5, PAsd5, CLsd5	
Standard deviation (sd) over the previous 21 days trading days of GC, SI, PA and CL	GCsd21, SIsd21, PAsd21, CLsd21	

The daily spot price will help the model to fix to current price location while the purpose of including the derivative input is to help the model to learn the underlying

relationship that is constant over time (Malliaris & Malliaris, 2008). The standard deviation for 5 and 21 days represents the volatility of the price/rate in a week (sd5) and a month (sd21) during business days.

3.5.1.2 Training, Validation and Testing

The data set is split into three independent subsets, namely training set (for model fitting), validation set (for model assessment and to prevent over fitting) and testing set (for real assessment of how well the model generalize) (Marsland, 2009). The proportion for each subset is tabulated in Table 3.24. The data arrangement is based on Arithmetic Progression formula (Tan, 2010). By using this formula, the data is arranged as shown in Table 3.25. The example makes use the CL data from Data Set A.

Table 3.24: Training, Validation and Testing

	Training	**Validation**	**Testing**
Data Proportion (%)	70	15	15
Data Arrangement	*Train1* = (2:6:*N*) *Train2* = (3:6:*N*) *Train3* = (4:6:*N*) *Train4* = (5:6:*N*) *Train5* = (6:6:*N*) *Train* = *horzcat*(*Train1,Train2,Train3, Train4,Train5*) Where *N* is sum of training and validation data	*Validation* = 1+ (*n*-1) x *d,* where *d* = 6, *n* = 1,2,3, …, *N* 1:6:*N*	Testing Set A: 1062-1248 Testing Set B: 1042-1225 Testing Set C: 802-943

Table 3.25: Example of data arrangement for Data Set A

Date	Day	Normalized Data of CL Data Set A	Training	Validation
30/12/1997	1	0.2592 (Validation Day 1)	0.2607	0.2592
31/12/1997	2	0.2607 (Training Day1)	0.2221	0.2366
02/01/1998	3	0.2527(Training Day178)	0.2240	0.2176
05/01/1998	4	0.2321(Training Day355)	0.2481	0.2385
06/01/1998	5	0.2359(Training Day532)	0.2298	0.2122
07/01/1998	6	0.2260(Training Day 279)	0.1992	0.2027
08/01/1998	7	0.2366 (Validation Day2)	0.1718	0.1851
09/01/1998	8	0.2221(Training Day2)	0.1710	0.1756
12/01/1998	9	0.2214(Training Day179)	0.1294	0.1286
13/01/1998	10	0.2153(Training Day356)	0.1424	0.1359
14/01/1998	11	0.2153(Training Day533)	0.2271	0.2282
15/01/1998	12	0.2149(Training Day 280)	0.1771	0.1924
16/01/1998	13	0.2176 (Validation Day 3)	0.1821	0.1740
20/01/1998	14	0.2240 (Training Day 3)	0.1672	0.1695
...	...	...	...	...
02/04/2002	1061	0.6462(Training Day 708)	0.5240	0.5565

From the table, it can be seen that, the selection of training and validation set is done every six days. For example, for the training set, the data from day 2 (see third column) become the training data for day 1 (see fourth column), and after 6 days, the selection is done again by selecting the data from day 8 which become as training data day 2. The selection proceeds continuously until all the data are distributed to the respective set.

3.5.1.3 Properties Setting

Prior to experimental processes, the properties of the proposed technique and the identified techniques are set, which is tabulated in Table 3.26.

For BPNN, the learning rate and momentum are decided through experimentation within the range 0 - 1 (Zhang, et al., 1998) while for SVM, the *C* parameter are set within 2^{-5}- 2^{5} (Ou & Wang, 2009).

Table 3.26: Properties of Prediction Techniques Utilized

Properties	***e*ABC/ABC/ *lv*ABC/*cm*ABC**	**GA**	**DE**	**PSO**	***e*PSO**	**BPNN**
SN	10	-	-	-	-	-
Number of bees/population size	20	20	20	20	20	-
Limit	*SN*D*	-	-	-	-	-
Maximum iteration	100	100	100	100	100	1000
Crossover probability	-	0.9	0.9	-	-	-
Mutation probability	-	0.1	-	-	-	-
Scaling factor	-	-	0.7	-	-	-
Acceleration constant	-	-	-	1.5, 1.7	1.5, 1.7	
Initial and final weight	-	-	-	0.9, 0.4	-	-

3.5.2 Performance Evaluation Metrics

In this study, the comparisons between prediction techniques involve different time series data. Thus, choosing an appropriate performance evaluation metric is crucial in order to justify the results obtained from the conducted experiment. With respect to that matter, five statistical metrics were chosen due to its suitability in evaluating the time series data of interest, namely Mean Absolute Percentage Error (MAPE), symmetric MAPE (sMAPE), Root Mean Square Percentage Error (RMSPE), Prediction Accuracy (PA) (Hyndman & Koehler, 2006)and Theil's U (Armstrong, 2001). All stated metrics interpret the learning and generalization capabilities of

prediction model. The formula for the above mentioned metrics are shown in 3.5 to 3.9.

$$MAPE = \frac{1}{N}\left[\sum_{n=1}^{N}\left|\frac{y_n - y(x_n)}{y_n}\right|\right] \tag{3.5}$$

The interpretation of result produced by MAPE is as tabulated in Table 3.27(Yorucu, 2003):

Table 3.27: Range of Percentage by MAPE

Percentage	**Interpretation**
10% <	Highly accurate prediction
10-20%	Good prediction
20-50%	Reasonable prediction
>50%	Inaccurate prediction

$$PA = 100\% - (MAPE \times 100) \tag{3.6}$$

$$sMAPE = mean(200 \mid y_n - y(x_n) \mid /(y_n + y(x_n)) \tag{3.7}$$

$$RMSPE = \sqrt{\left(\frac{\sum_{n=1}^{N}\left(\frac{y_n - y(x_n)}{y_n}\right)^2}{N}\right)} \tag{3.8}$$

$$Theil'sU = \frac{\sqrt{\frac{1}{N}\Sigma_{n=1}^{N}(y_n - y(x_n))^2}}{\sqrt{\frac{1}{N}\Sigma_{n=1}^{N}(y_n)^2} + \sqrt{\frac{1}{N}\Sigma_{n=1}^{N}(y(x_n))^2}} \quad (3.9)$$

where $n = 1, 2, ..., N$

y_n = actual values

$y(x_n)$ = predicted values / approximate values by predictor models

N =Number of test data

After the prediction task, the normalized data are rescaled back based on the reverse of the Min Max Normalization (see 3.2.3). This means all evaluations are calculated based on the original scale of the data.

3.5.3 Benchmarking Techniques

In this study, the results from the proposed *e*ABC-LSSVM are compared with the results produced by variants of ABC-LSSVM which includes ABC-LSSVM, *lv*ABC-LSSVM and *cm*ABC-LSSVM. Besides the variants of ABC-LSSVM, other benchmarking techniques are GA-LSSVM (Mustafa, et al., 2012), PSO-LSSVM (Jiang & Zhao, 2013), *e*PSO-LSSVM, CV-LSSVM (Mellit, et al. 2013), DE-LSSVM(Feoktistov & Janaqi, 2004), BPNN(Malliaris & Malliaris, 2008) and SVM (Vapnik, 1995).Details on ABC-LSSVM are described in Chapter 4 while both *lv*ABC-LSSVM and *cm*ABC-LSSVM are discussed in Chapter 5. On the other hand, description on the utilized benchmark techniques can be seen in Chapter 2.

3.6 Summary

This chapter presents the methodology implemented in this study using the proposed algorithm, *e*ABC-LSSVM for the purpose of predicting the time series data of interest. This phase initiates with data collection which incorporated of several test such as test for correlation and also non linearity test. Later, data normalization was performed as to reduce the difficulty in calculation and increase the precision as well. In design algorithm, it consists of the design for ABC-LSSVM, *lv*ABC-LSSVM and *cm*ABC-LSSVM. Both *lv*ABC-LSSVM and *cm*ABC-LSSVM are embedded in *e*ABC-LSSVM. The *lv*ABC-LSSVM was designed to overcome the problem of limitation in searching behavior of the ABC algorithm. It is achieved by using Levy mutation and different strategies. Meanwhile, the *cm*ABC-LSSVM is introduced to prevent the algorithm from facing with over fitting and under fitting problem. For evaluation purposes, the proposed algorithm are compared with ten identified prediction techniques and guided by five statistical metrics as described earlier. Next chapter presents the flow of ABC-LSSVM and followed by the algorithm.

CHAPTER FOUR
ARTIFICIAL BEE COLONY-LEAST SQUARES SUPPORT VECTOR MACHINES ALGORITHM

4.1 Introduction

As highlighted previously (see Chapter 1, section 1.3) the performance of LSSVM is highly relies on the value of its hyper-parameters, viz. γ and σ^2.To ensure the optimality of value of hyper-parameters, the LSSVM is hybridized with ABC algorithm.

In ABC-LSSVM algorithm, the ABC algorithm is utilized as an optimization tool for the LSSVM hyper-parameters. Generally, the LSSVM is integrated with the ABC algorithm, where here, the LSSVM acts as a fitness function evaluation. The optimized value of LSSVM hyper-parameters can be achieved after a maximum number of iteration has been reached. In this study, the objective function is guided by MAPE, where the lower the MAPE, the better the prediction accuracy (see Chapter 3 section 3.5.2).

The flow of ABC-LSSVM is given in Figure 4.1 and 4.2 while the algorithm is as presented in Algorithm 4.1.

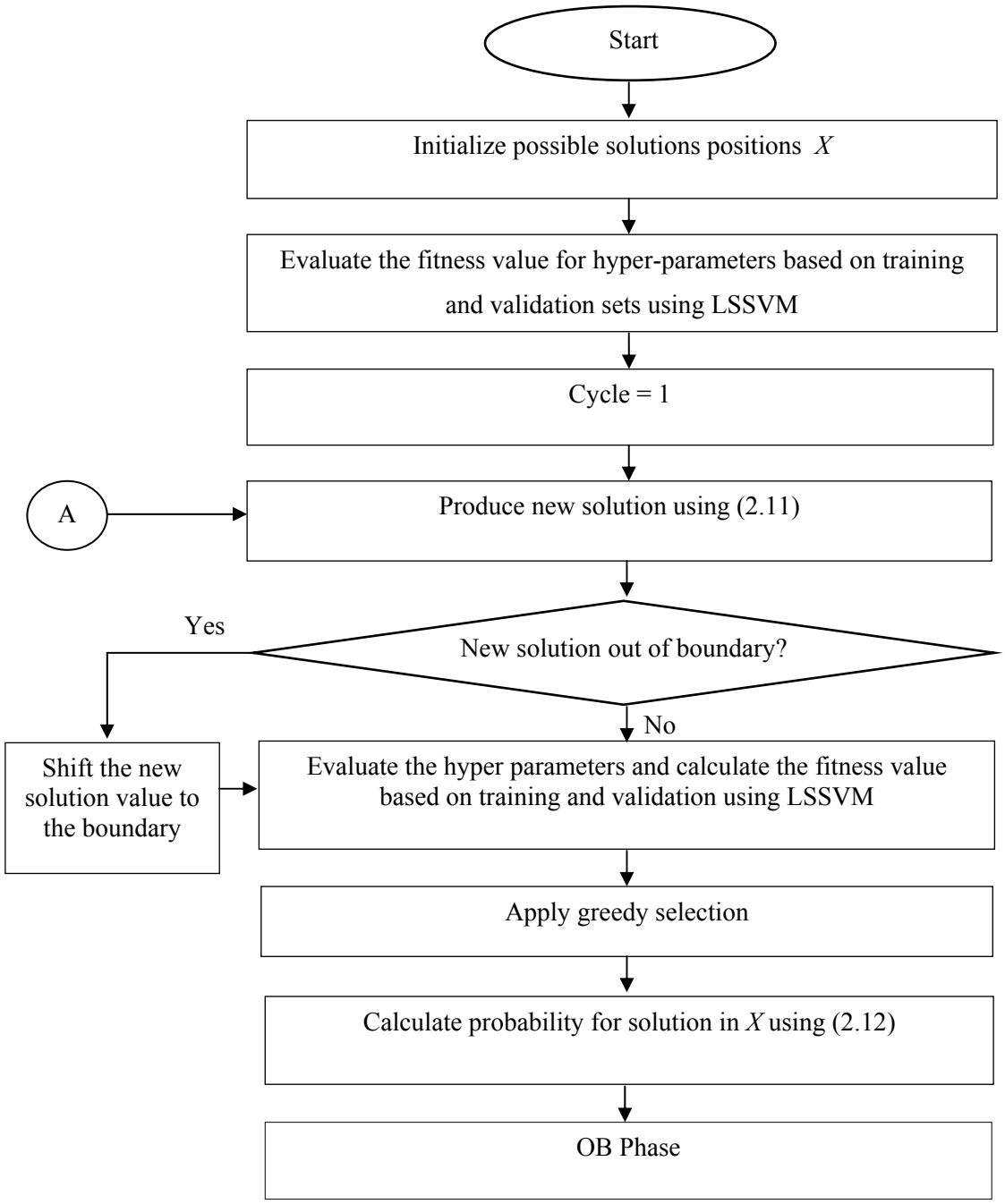


Figure 4.1: Flow of ABC-LSSVM in Initialization and EB Phase

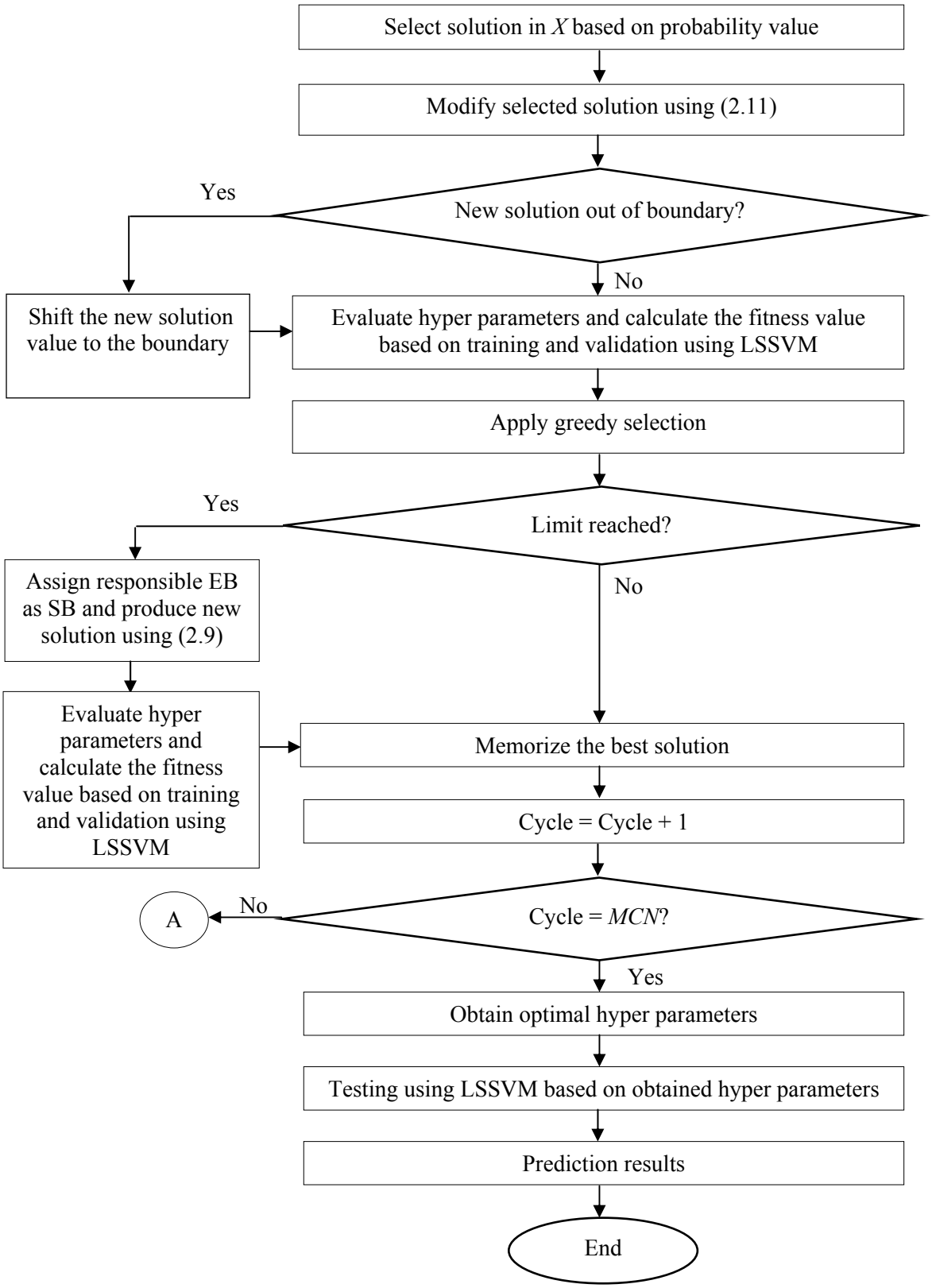


Figure 4.2: Flow of ABC-LSSVM in OB and SB Phase

Algorithm 4.1 **ABC-LSSVM algorithm**

1: Initialize possible solutions positions X
2: Evaluate the fitness value of hyper-parameters based on training and validation sets using LSSVM
3: Cycle = 1
4: **while** Cycle <= *MCN*
5:**for** each EB**(EB PHASE)**
6: Produce new solution using (2.11)
7:**if**the new solution is out of boundary, shift the new solution value to the boundary
8: Evaluate hyper-parameters and calculate the fitness value based on training and validation sets using LSSVM
9: **if** fitness value of new solution is better than fitness value of old solution
10: Keep new solution
11: Trial = 0.
12: **else**
13: Keep old solution
14: Trial = Trial +1
15: **end if**
16:**end for**
17: Calculate the probability values for solution in matrix X using (2.12)
18: **for** each OB**(OB PHASE)**
19: Select a solution in X based on probability value
20: Modify selected solution using (2.11)
21:**if**the new solution is out of boundary, shift the new solution value to the boundary
22: Evaluate hyper-parameters and calculate the fitness value based ontraining and validation sets using LSSVM
23:**if** fitness value of new solution is better than fitness value of old solution
24: Keep new solution
25: Trial = 0
26: **else**
27: Keep old solution
28: Trial = Trial + 1
29: **end if**
30:**end for**
31: **if (**max**)**Trial>*Limit*
32 **SB PHASE**
33: Assign responsible EB as SB and produce new solution to replace the abandoned food source using (2.9)
34: Evaluate hyper-parameters and calculate the fitness value based on training and validation using LSSVM
35:**else**
36: Memorized best solution
37:**end if**
38:**end for**
39: Cycle = Cycle +1
40:**end while**
41: Print optimal hyper-parameters
42: Testing using LSSVM

4.2 Summary

The hybridization of ABC with LSSVM, which is termed as ABC-LSSVM is presented in this chapter. By integrating both techniques, the optimized value of hyper-parameters can be obtained, which will influence the generalization performance of LSSVM in prediction task. However, due to limitations of ABC algorithm namely limitation in searching behavior and tendency to fall into boundary values, the *e*ABC-LSSVM is proposed. This is to ensure the generalization performance of the *e*ABC-LSSVM in prediction task. The flow and algorithm of *e*ABC-LSSVM are given in the next chapter.

CHAPTER FIVE
ENHANCED ARTIFICIAL BEE COLONY-LEAST SQUARES SUPPORT VECTOR MACHINES ALGORITHM

5.1 Introduction

In this chapter, the algorithms for *lv*ABC-LSSVM, *cm*ABC-LSSVM and *e*ABC-LSSVM are presented. Firstly, the *lv*ABC-LSSVM are designed to enrich the searching behavior of the artificial bees in the search space. This solution is to address the limitation of searching behavior in the search space that exist in the standard ABC algorithm. This limitation would caused the algorithm to face with local minimum trapping.

Meanwhile, the *cm*ABC-LSSVM contributes in preventing the algorithm from facing with over fitting and under fitting problem. The *cm*ABC-LSSVM is induced in decision making task, in both EB and OB phases. Later, both *lv*ABC-LSSVM and *cm*ABC-LSSVM are combined which contribute to better generalization as escape from premature convergence. The combinations of both algorithms are termed as *e*ABC-LSSVM.

Algorithm 5.1 ***Iv*ABC-LSSVM algorithm**

1: Initialize possible solutions, X
2: Evaluate the fitness value of hyper-parameters based on training and validation sets using LSSVM
3: Cycle = 1
4: **while** Cycle <= *MCN*
5: **for** each EB **(EB PHASE)**
6:Produce new solutions using (3.2)
7: **if** new solution out of boundary, shift the new solution value to the boundary
8: Evaluate hyper-parameters and calculate the fitness value based on training and validation sets using LSSVM
9: **if** fitness value of the new solution is better than fitness value of old solution
10: Keep new solution
11: Trial = 0
12: **else**
13: Keep old solution
14: Trial = Trial +1
15: **end if**
16: **end for**
17: Calculate the probability values for solutions in X using (2.12)
18: **for** each OB **(OB PHASE)**
19: Select a solution in X based on probability value
20: **Modify selected solution using (3.3)**
21: **if** new solution is out of boundary, shift the new solution value to the boundary
22: Evaluate hyper-parameters and calculate the fitness value based on training and validation sets using LSSVM
23: **if** fitness value of new solution is better than fitness value of old solution
24: Keep new solution
25: Trial = 0
26: **else**
27: Keep old solution
28: Trial = Trial +1
29: **end if**
30: **end for**
31: **if** (max)Trial>*Limit*
32: **SB PHASE**
23: Assign responsible EB as SB and produce new solution to replace the abandoned food source using (2.9)
34: Evaluate hyper-parameters and calculate the fitness value based on training and validation using LSSVM
35: **else**
36: Memorized best solution
37: **end if**
38: **end for**
39: Cycle = Cycle +1
40: **end while**
41: Print optimal γ, $\sigma^2$and α
42: Testing using LSSVM based on hyper-parameters

Algorithm 5.2 ***cm*ABC-LSSVM algorithm**

1: Initialize possible solutions, X
2: Evaluate the fitness value of hyper-parameters based on training and validation sets using LSSVM
3: Cycle = 1
4: **while** Cycle <= *MCN*
5: **for** each EB **(EB PHASE)**
6: Produce new solutions using (2.11)
7:ifnew solution out of boundary, apply (3.4)
8: Evaluate hyper-parameters and calculate the fitness value based on training and validation sets using LSSVM
9: **if** fitness value of the new solution is better than fitness value of old solution
10: Keep new solution
11: Trial = 0
12: **else**
13: Keep old solution
14: Trial = Trial +1
15: **end if**
16: **end for**
17: Calculate the probability values for solutions in X using (2.12)
18: **for** each OB **(OB PHASE)**
19: Select a solution in X based on probability value
20: Modify selected solution using (2.11)
21:ifnew solution is out of boundary, apply (3.4)
22: Evaluate hyper-parameters and calculate the fitness value based on training and validation sets using LSSVM
23: **if** fitness value of new solution is better than fitness value of old solution
24: Keep new solution
25: Trial = 0
26: **else**
27: Keep old solution
28: Trial = Trial +1
29: **end if**
30: **end for**
31: **if** (max)Trial>*Limit*
32: **SB PHASE**
23: Assign responsible EB as SB and produce new solution to replace the abandoned food source using (2.9)
34: Evaluate hyper-parameters and calculate the fitness value based on training and validation using LSSVM
35: **else**
36: Memorized best solution
37: **end if**
38: **end for**
39: Cycle = Cycle +1
40: **end while**
41: Print optimal γ, $\sigma^2$and α
42: Testing using LSSVM based on hyper-parameters

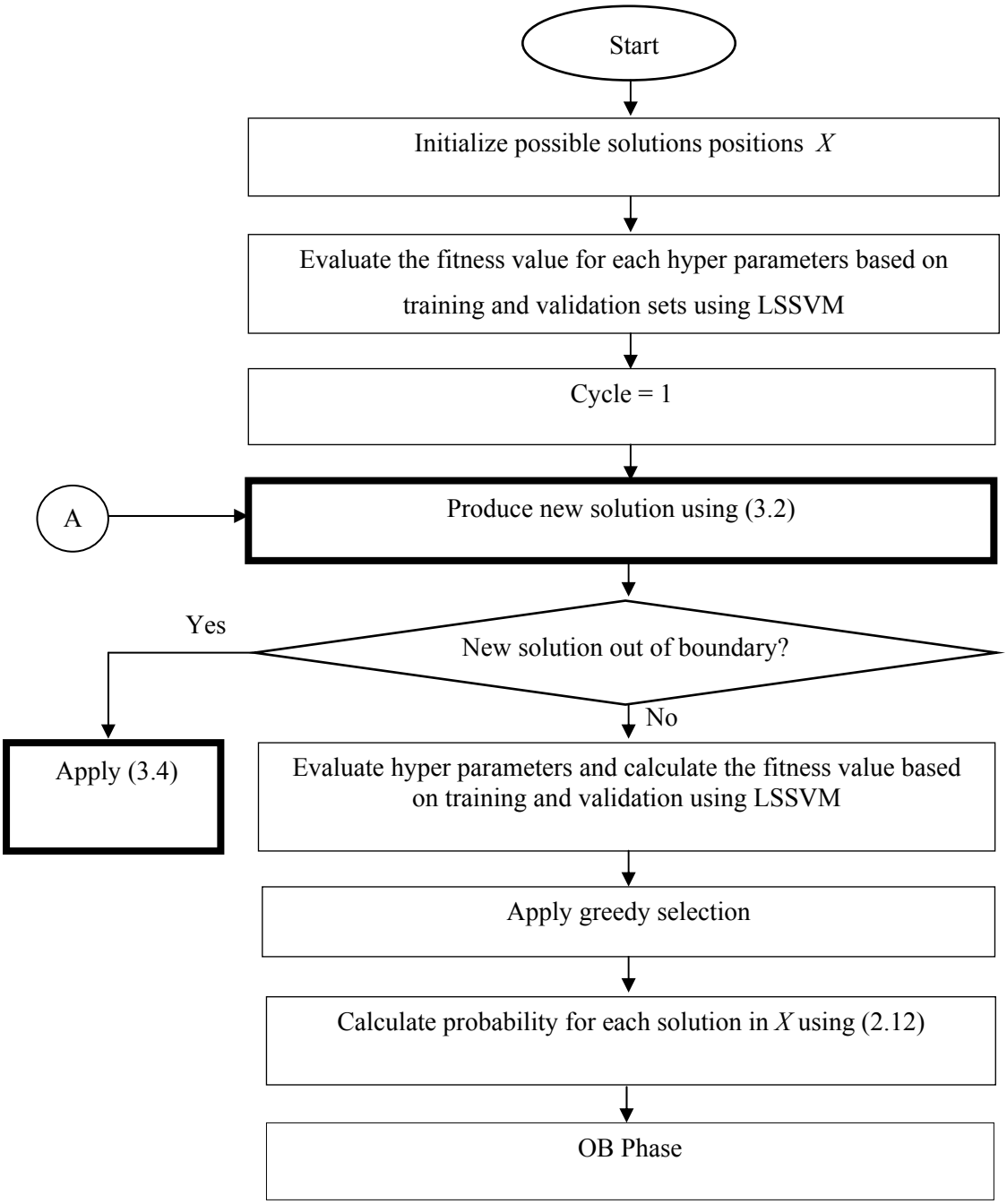


Figure 5.1: Flow of eABC-LSSVM in Initialization and EB Phase

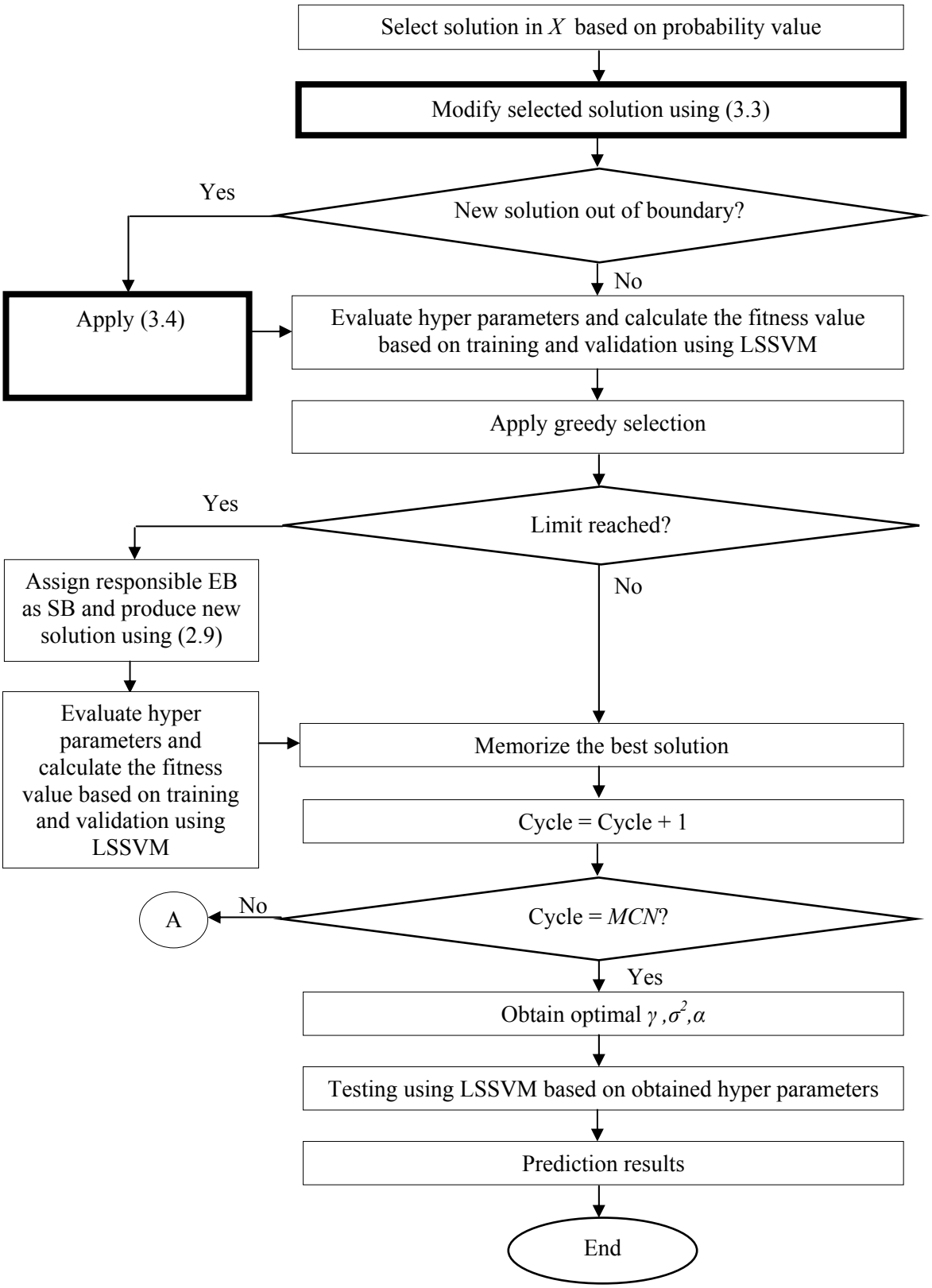


Figure 5.2: Flow of eABC-LSSVM in OB and SB Phase

Algorithm 5.3 ***e*ABC-LSSVM algorithm**

1: Initialize possible solutions, X
2: Evaluate the fitness value of hyper-parameters based on training and validation sets using LSSVM
3: Cycle = 1
4: **while** Cycle <= *MCN*
5: **for** each EB **(EB PHASE)**
6: Produce new solutions using (3.2)
7: if new solution out of boundary, apply (3.4)
8: Evaluate hyper-parameters and calculate the fitness value based on training and validation sets using LSSVM
9: **if** fitness value of the new solution is better than fitness value of old solution
10: Keep new solution
11: Trial = 0
12: **else**
13: Keep old solution
14: Trial = Trial +1
15: **end if**
16: **end for**
17: Calculate the probability values for solutions in X using (2.12)
18: **for** each OB **(OB PHASE)**
19: Select a solution in X based on probability value
20: Modify selected solution using (3.3)
21: if new solution is out of boundary, apply (3.4)
22: Evaluate hyper-parameters and calculate the fitness value based on training and validation sets using LSSVM
23: **if** fitness value of new solution is better than fitness value of old solution
24: Keep new solution
25: Trial = 0
26: **else**
27: Keep old solution
28: Trial = Trial +1
29: **end if**
30: **end for**
31: **if** (max)Trial>*Limit*
32: **SB PHASE**
23: Assign responsible EB as SB and produce new solution to replace the abandoned food source using (2.9)
34: Evaluate hyper-parameters and calculate the fitness value based on training and validation using LSSVM
35: **else**
36: Memorized best solution
37: **end if**
38: **end for**
39: Cycle = Cycle +1
40: **end while**
41: Print optimal γ, σ^2 and α
42: Testing using LSSVM based on hyper-parameters

5.2 Summary

The enhancements introduced in standard ABC-LSSVM, which is termed as *e*ABC-LSSVM has been presented in this chapter. The first enhancement involved the inducing of Levy mutation and different strategies which address the problem of limitation in searching behavior of the algorithm (*lv*ABC-LSSVM) while the second enhancement is introduced in decision making phase by introducing the conventional mutation (*cm*ABC-LSSVM). The purpose of introducing the *cm*ABC-LSSVM is to prevent the over fitting and under fitting problem. The combinations of *lv*ABC-LSSVM and *cm*ABC-LSSVM which is known as *e*ABC-LSSVM encourage the algorithm to achieve the global optimal.

In the next chapter, the performance of *e*ABC-LSSVM in non renewable natural resources commodities price predictions against identified prediction algorithms are discussed and analyzed. The evaluations are based on five indices, namely MAPE, PA, sMAPE, RMSPE and Theil's U.

CHAPTER SIX
RESULTS AND DISCUSSION

6.1 Introduction

This chapter discusses the results obtained from the series of experiments conducted. The experiments involved three sets of time series data, namely Data Set A, B and C. The descriptions of these data sets are provided in Chapter 3, section 3.2. The objective of the undertaken experiments is to evaluate the effectiveness of proposed algorithm compared to the state of the art techniques. The bold results indicated the best results produced by the respective technique.

6.2 *e*ABC-LSSVM on Energy Fuels

This section presents the results obtained by *e*ABC-LSSVM in energy fuels price prediction using Data Set A and B.

6.2.1 Data Set A

In this experiment, the price prediction for each energy fuels (CL, HO, HU and PN) was performed separately.

6.2.1.1 *e*ABC-LSSVM vs ABC-LSSVM and Its Variants

This section presents the comparison results obtained between the *e*ABC-LSSVM and variants ABC-LSSVM, which includes *lv*ABC-LSSVM and *cm*ABC-LSSVM.

6.2.1.1.1 Crude Oil Prices

From the results tabulated in Table 6.1, lowest MAPE was obtained by *e*ABC-LSSVM, which is 5.3975%. This is followed by *cm*ABC-LSSVM which yields 5.9836% of MAPE. Meanwhile, higher MAPE were recorded by both *lv*ABC-LSSVM and ABC-LSSVM when the value for γ and σ^2 are set to [334.1669, 1] and [333.7988, 1] respectively. A good agreement from sMAPE, RMSPE and Theil's U were also recorded which are in favor to *e*ABC-LSSVM. Meanwhile, Table 6.2 shows the results obtained from paired sample T-test. From the table, it shows that the statistical level of the difference of the means between the *e*ABC-LSSVM and the variants ABC-LSSVM is significant at 0.05% significance level. This implies that the enhancements that have been introduced are significant.

The statistical results in graphical presentation are depicted in Figure 6.1 while the comparison of convergence is shown in Figure 6.2. From Figure 6.2, the graph clearly shows that the *e*ABC-LSSVM, which is marked with redline, is able to escape from local minimum which finally contributed in lower MAPE in validation. Consequently, the value of γ and σ^2 with lowest MAPE are chosen as optimal value for parameter of interest. On the other hand, the fitness curve of ABC-LSSVM which is marked with green line decreases at the very beginning of the iterations and then the curve appear flat. This finally resulted the model to miss the global optimal and finally suffer in premature convergence.

Table 6.1: CL Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	15.2458	334.1669	808.7678	333.7988
σ^2	1.5832	1	4.1738	1
α	0.6898	1.0170	-	-
MAPE Training (%)	6.0579	**3.4165**	5.4333	3.4173
MAPE Validation (%)	**5.3802**	5.6527	5.5687	5.6519
MAPE Testing (%)	**5.3975**	5.9836	5.7081	5.9829
PA (%)	**94.6025**	94.0164	94.2919	94.0171
sMAPE (%)	**5.3802**	6.1137	5.6888	6.1129
RMSPE (%)	**0.0681**	0.0765	0.0736	0.0765
Theil's U	**0.0336**	0.0390	0.0361	0.0390
Minimum PA (%)	93.3037	93.0397	93.3196	**93.9763**
Average PA (%)	**94.0394**	93.5134	93.9344	93.9866

Table 6.2: Significant Test for CL Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.019
*e*ABC-LSSVM – *cm*ABC-LSSVM	.000
*e*ABC-LSSVM – ABC-LSSVM	.019

Figure 6.3 and 6.4 graphically show the comparison of the results obtained which highlights the pattern of the exploitation process in *e*ABC-LSSVM and ABC-LSSVM respectively. From Figure 6.3, it can be seen that a large step length is used in *e*ABC-LSSVM upon completing the searching of global optimal, which increases the chance of getting better results. This important characteristic which is lacking in ABC-LSSVM in terms of the smaller step length in search space, induce it to stuck in a local minimum since the beginning of the searching process. Hence, the prediction precision is affected.

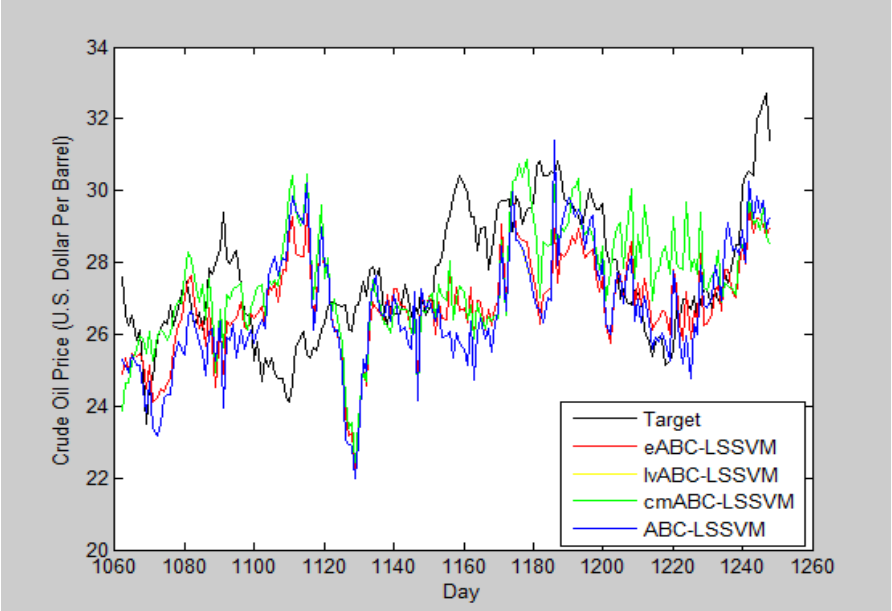


Figure 6.1: Comparison of CL Price Predictions Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

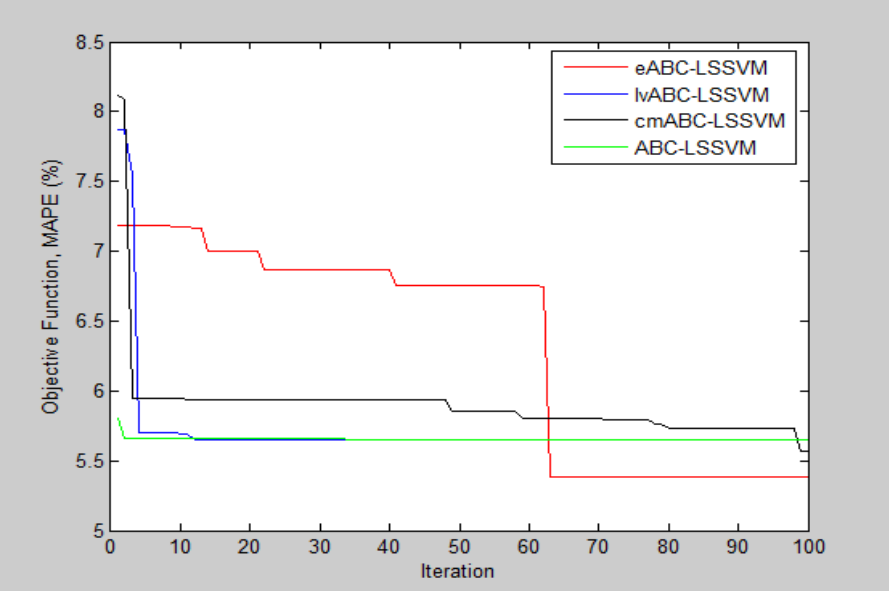
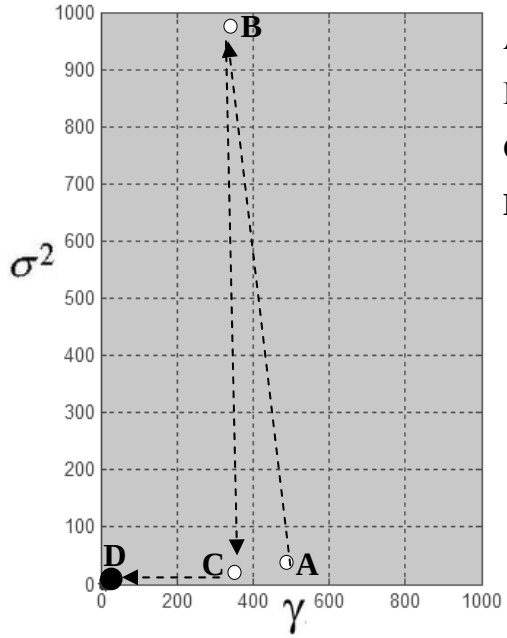


Figure 6.2: Comparison of Convergence for CL Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants



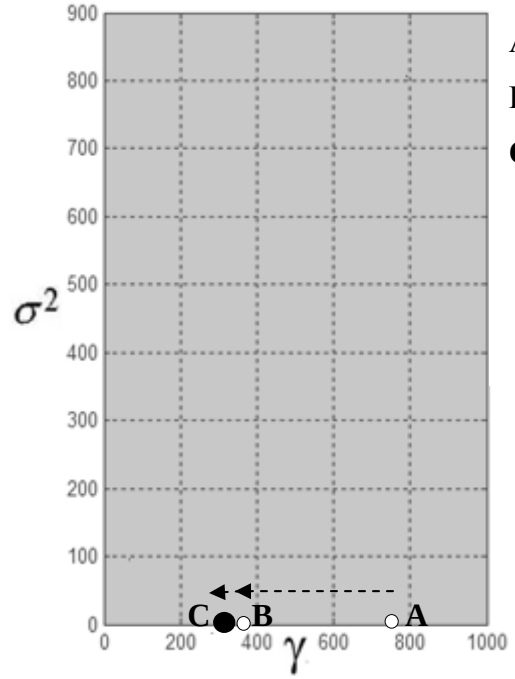
A [446.392, 19.0429]

B [353.899, 981.875]

C [346.595, 8.71729]

D [15.2458, 1.58319] *Optimal γ and σ^2

Figure 6.3: Exploitation of Search Space by eABC-LSSVM in CL Price Prediction Data Set A



A [788.148, 1]

B[383.757, 1]

C[333.799, 1] *Optimal γ and σ^2

Figure 6.4: Exploitation of Search Space by ABC-LSSVM in CL Price Prediction Data Set A

6.2.1.1.2 Heating Oil Prices

For HO price prediction (see Table 6.3), the prediction accuracy recorded by ABC-LSSVM was 92.8231%. With an increment of 0.8459% of prediction accuracy, the superiority is in favor to *e*ABC-LSSVM. It is also indicated that consistent results were also produced by the other metrics. This is followed by *lv*ABC-LSSVM which achieved 93.3636% of prediction accuracy. The significant results of each enhancement are tabulated in Table 6.4. For visualization purposes, the illustration is provided in Figure 6.5 which shows the comparison between target and predicted values by each prediction technique. By observing the red line which represents *e*ABC-LSSVM, the figure shows narrower span between *e*ABC-LSSVM with actual value, as compared to prediction value produced by other techniques.

Besides lowest MAPE in testing, *e*ABC-LSSVM is also capable to achieve the lowest MAPE in validation set, which represents the convergence value. The MAPE recorded for this set was 5.7943%. The illustration is shown in Figure 6.6. The superiority of *e*ABC-LSSVM in this case is also illustrated in Figure 6.7. From the figure, it can be seen that bigger steps are taken from position A to B, and later continued to C. After three small steps from position C to E, the progressive search is performed to a wider search area, which finally led to optimal solution at position F [80.2488, 944.0822]. However, contrary situation is presented in ABC-LSSVM make it stuck in local minimum at B [67.8285, 134.4249] (see Figure 6.8). Thus, by fully exploiting the search space, the *e*ABC-LSSVM provide a better search strategy which consequently assist in achieving global optimal as opposed to ABC-LSSVM.

Table 6.3: HO Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	80.2488	42.1570	205.0675	67.8285
σ^2	944.0822	278.6544	323.9908	134.4249
α	1.8135	0.2613	-	-
MAPE Training (%)	8.4098	8.3046	8.2074	**8.1882**
MAPE Validation (%)	**5.7943**	6.0162	6.2464	5.9865
MAPE Testing (%)	**6.3310**	6.6364	7.2388	7.1769
PA (%)	**93.6690**	93.3636	92.7612	92.8231
sMAPE (%)	**6.2715**	6.5124	7.0139	6.9577
RMSPE (%)	**0.0801**	0.0844	0.0926	0.0918
Theil's U	**0.0405**	0.0418	0.0448	0.0444
Minimum PA (%)	**92.5443**	88.8245	88.4186	88.4679
Average PA (%)	**92.8077**	92.7408	90.0733	88.7669

Table 6.4: Significant Test for HO Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.000
*e*ABC-LSSVM – ABC-LSSVM	.000

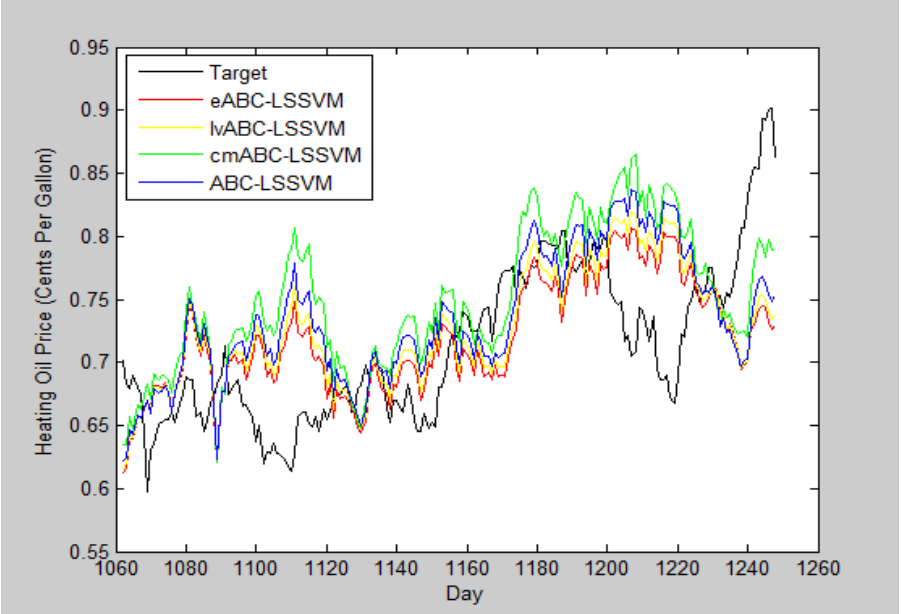


Figure 6.5: Comparison of HO Price Predictions Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

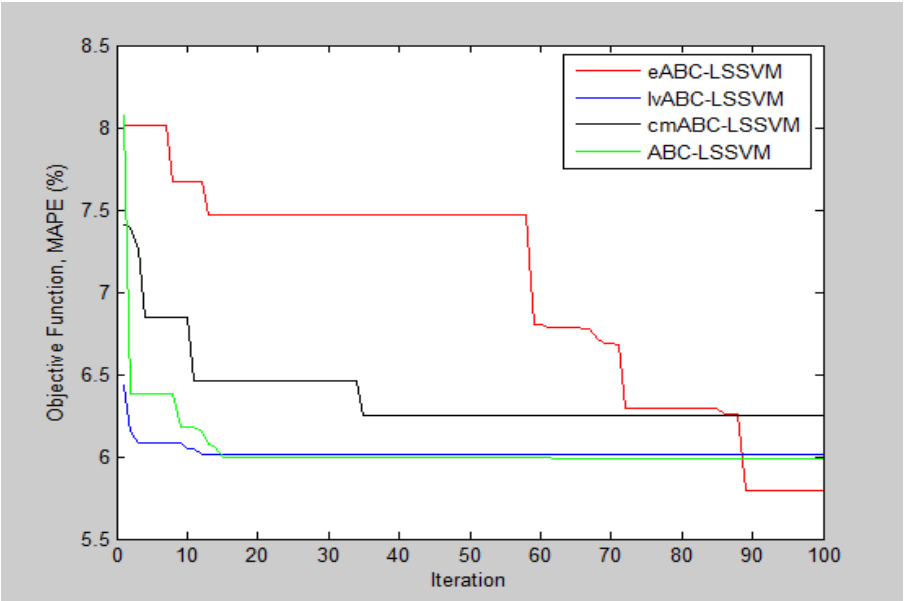
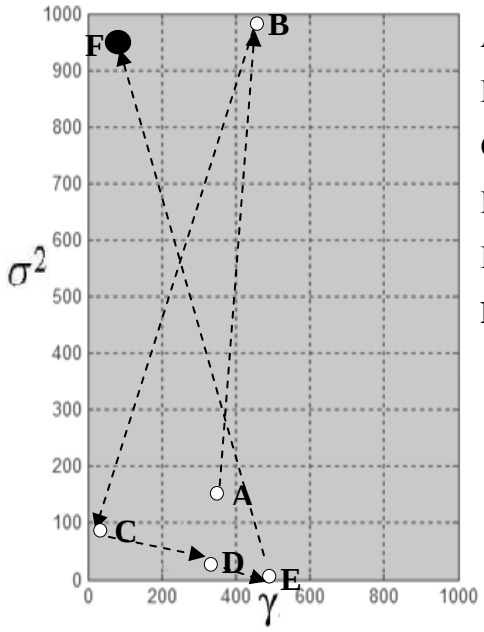


Figure 6.6: Comparison of Convergence for HO Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants



A [376.359, 147.655]

B [449.911,977.307]

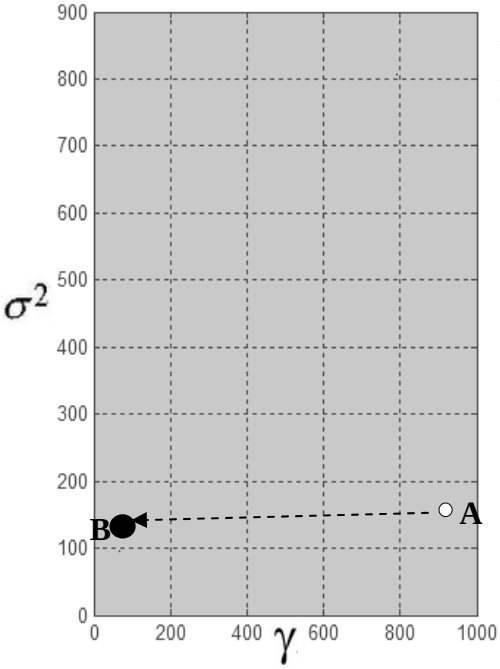
C [25.5976, 95.2171]

D [367.549, 4.5132]

E [523.525, 1.2617]

F [80.2488,944.0822] *Optimal γ and σ^2

Figure 6.7: Exploitation of search space by eABC-LSSVM in HO Price Prediction Data Set A



A [932.75, 151.117]

B [67.8285, 134.4249] *Optimal γ and σ^2

Figure 6.8: Exploitation of search space by ABC-LSSVM in HO Price Prediction Data Set A

6.2.1.1.3 Gasoline Prices

By referring to Table 6.5, the prediction accuracy recorded by ABC-LSSVM was 93.6221%. Meanwhile, similar increment was recorded by *lv*ABC-LSSVM and *cm*ABC-LSSVM, which was around 0.75%. The combination of both techniques leads to higher prediction accuracy, which was 0.9155%. This makes *e*ABC-LSSVM leads the pack by producing 94.5376% of prediction accuracy. In addition, lowest sMAPE, RMSPE and Theil's U were recorded by *e*ABC-LSSVM are also in accordance with the MAPE produced. The significance level for every enhancement is as tabulated in Table 6.6. From the table, a positive results are obtained, which prove the significance of each improvement that has been introduced.

For illustration purposes, the prediction results and convergence by experimented prediction techniques are plotted in Figure 6.9 and 6.10 respectively. Besides the searching behavior that has been enriched using *lv*ABC in *e*ABC-LSSVM (see Figure 6.11), the decision making process which involved *cm*ABC also assist the model from falling to the boundary value, as occurred in ABC-LSSVM. Besides has been tabulated in table 6.5, this situation also visualized in Figure 6.12.

Table 6.5: HU Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	121.3982	381.3253	979.2209	377.9901
σ^2	77.6900	67.2659	112.8787	1
α	1.3667	1.8578	-	-
MAPE Training (%)	10.1283	9.8794	9.9166	**4.0091**
MAPE Validation (%)	**6.5947**	6.8074	6.9661	6.8074
MAPE Testing (%)	**5.4624**	5.6183	5.6199	6.3779
PA (%)	**94.5376**	94.3817	94.3801	93.6221
sMAPE (%)	**5.5067**	5.6356	5.6409	6.5787
RMSPE (%)	**0.0693**	0.0703	0.0704	0.0821
Theil's U	**0.0356**	0.0358	0.0359	0.0423
Minimum PA (%)	**94.2743**	93.2265	93.2486	93.1995
Average PA (%)	**94.4717**	93.7423	93.5960	93.5778

Table 6.6: Significant Test for HU Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.000
*e*ABC-LSSVM – ABC-LSSVM	.007

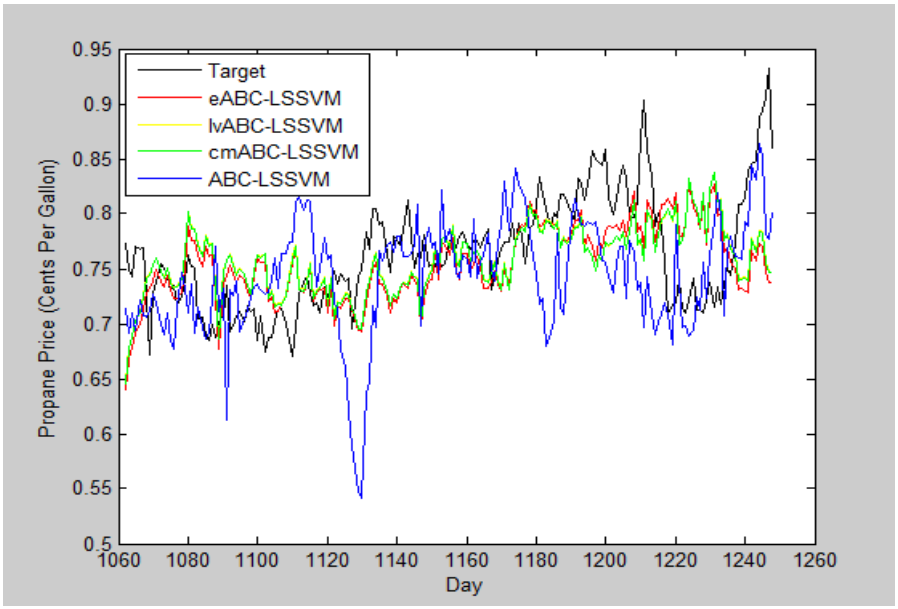


Figure 6.9: Comparison of HU Price Predictions Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

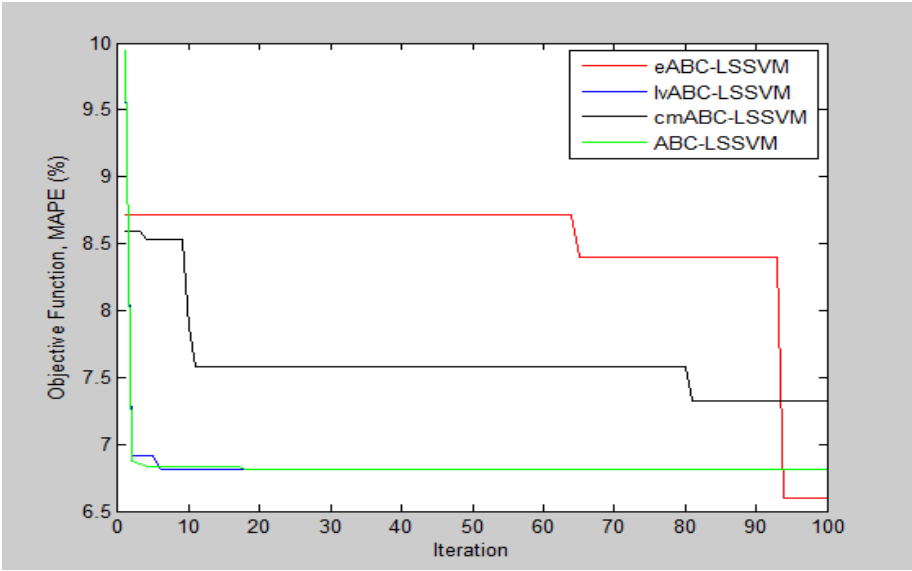
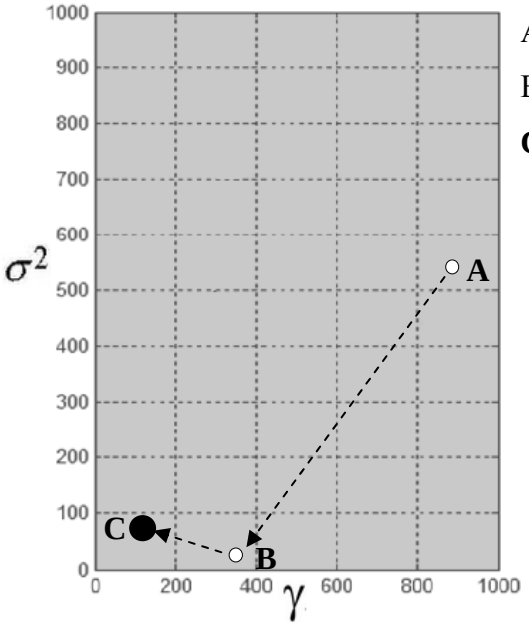


Figure 6.10: Comparison of Convergence for HU Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

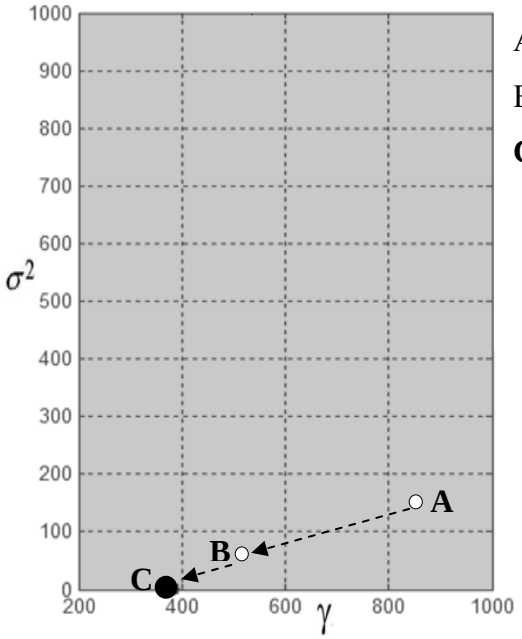


A [881.512, 516.149]

B [316.964, 4.8767]

C [121.3982,77.6900] * Optimal γ and σ^2

Figure 6.11: Exploitation of Search Space by eABC-LSSVM in HU Price Prediction Data Set A



A [814.909, 158.455]

B [511.898, 61.5458]

C [377.9901, 1] * Optimal γ and σ^2

Figure 6.12: Exploitation of Search Space by ABC-LSSVM in HU Price Prediction Data Set A

6.2.1.1.4 Propane Prices

The results of PN price prediction are summarized in Table 6.7. Taking all stated performance evaluation metrics, it can be seen that lowest prediction error is produced by *e*ABC-LSSVM, hence highest prediction accuracy is obtained as compared to the other techniques. This is followed by *lv*ABC-LSSVM with 13.1432% of MAPE. Meanwhile, for *cm*ABC-LSSVM and ABC-LSSVM, both techniques produced 16.1298% and 16.6509% of MAPE respectively. With significant difference between *e*ABC-LSSVM and the other techniques (see Table 6.8), it shows that the proposed technique is significantly better in PN price prediction as compared to the others.

Meanwhile, the graphical results which show the actual and predicted values by prediction models are illustrated in Figure 6.13.In terms of convergence, the *e*ABC-LSSVM is able to converge at lowest value, which is 6.3979% of MAPE while the rest of the techniques converge at similar value, which is around 7.7% of MAPE (see Figure 6.14). Figure 6.15 and 6.16 show the exploitation process for both *e*ABC-LSSVM and ABC-LSSVM respectively. Analysis on Figure 6.16 clearly shows that ABC-LSSVM is incapable to fully exploit the search space when the value of LSSVM hyper-parameters mostly falls at the boundary since the beginning of the iteration. Nonetheless, this is not the case in *e*ABC-LSSVM where efficiently usage of search space is portrayed by *e*ABC-LSSVM, which is finally beneficial in getting an ideal combination of hyper-parameters value, which is γ = 48.6424 and σ^2 = 536.4877 (see Figure 6.15).

Table 6.7: PN Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	48.6424	389.5877	494.9713	623.9340
σ^2	536.4877	406.6509	31.1097	1
α	1.6454	1.7126	-	-
MAPE Training (%)	12.2363	12.1837	10.6649	**3.3233**
MAPE Validation (%)	**6.3979**	7.7309	7.7708	7.7234
MAPE Testing (%)	**12.2780**	13.1432	16.1298	16.6509
PA (%)	**87.7220**	86.8568	83.8702	83.3491
sMAPE (%)	**11.4125**	12.1389	14.6117	14.8760
RMSPE (%)	**0.1443**	0.1543	0.1873	0.2193
Theil's U	**0.0631**	0.0677	0.0812	0.0954
Minimum PA (%)	**84.5807**	81.9341	79.1594	81.8547
Average PA (%)	**86.7378**	85.4704	80.6943	81.9769

Table 6.8: Significant Test for PN Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.000
*e*ABC-LSSVM – ABC-LSSVM	.007

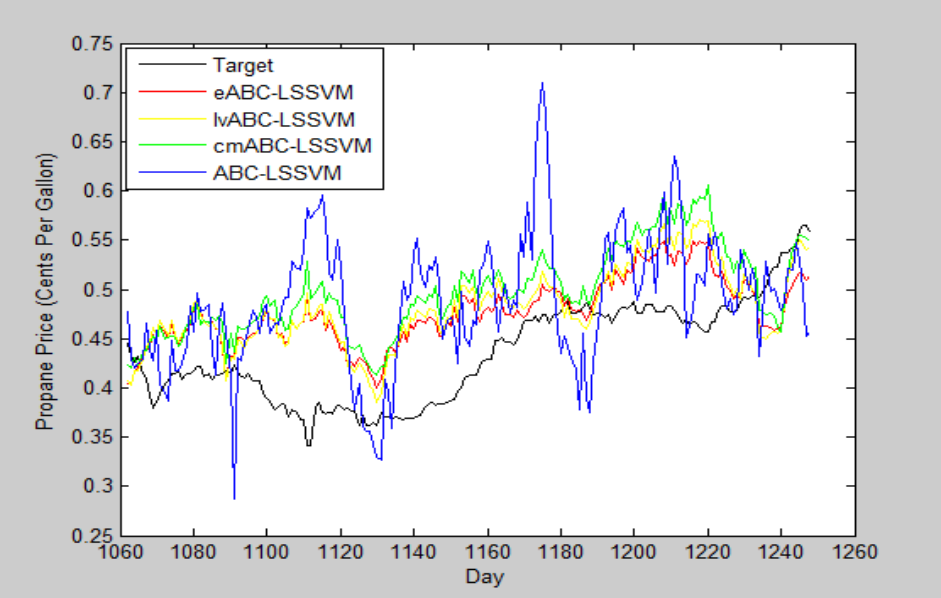


Figure 6.13: Comparison of PN Price Predictions Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

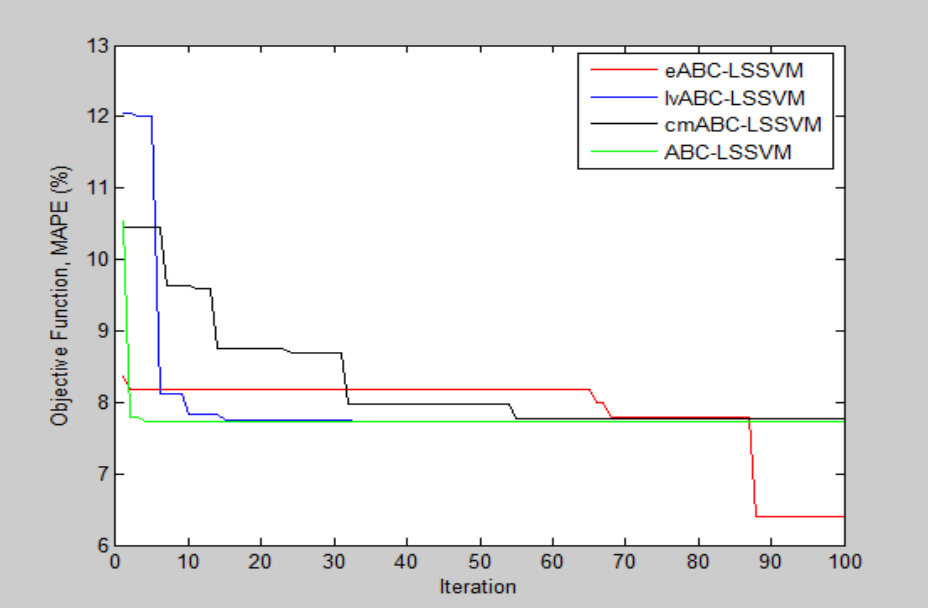
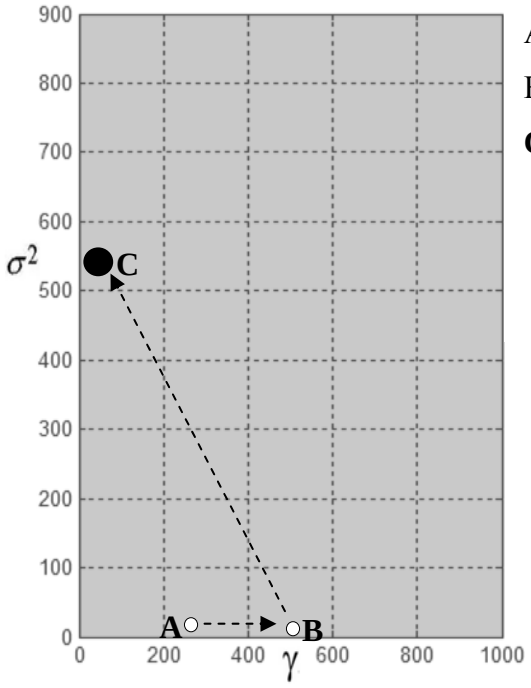


Figure 6.14: Comparison of Convergence for PN Price Prediction Data Set A: eABC-LSSVM vs. ABC-LSSVM and Its Variants

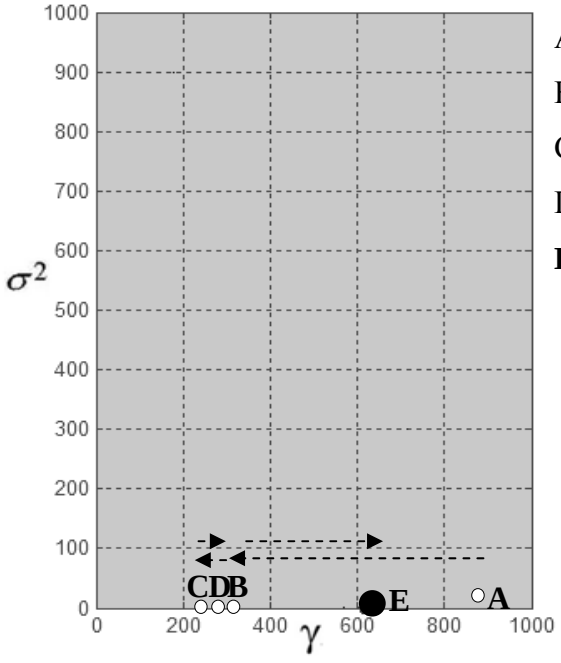


A [230.118, 2.1699]

B [515.43, 1.51067]

C [48.6424, 536.488] * Optimal γ and σ^2

Figure 6.15: Exploitation of Search Space by eABC-LSSVM in PN Price Prediction Data Set A



A [875.314, 15.2433]

B [346.392, 1]

C [292.406, 1]

D [304.863, 1]

E [623.934, 1] * Optimal γ and σ^2

Figure 6.16: Exploitation of Search Space by ABC-LSSVM in PN Price Prediction Data Set A

6.2.1.1.5 Results Analysis for Data Set A: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants

Based on the results obtained for CL, HO, HU and PN price prediction for Set A, it can be noted that the proposed technique, *e*ABC-LSSVM is capable to outperform the other identified techniques in all stated price prediction, regardless of statistical metrics utilized. The enhancement introduced in exploitation process which is based on Levy mutation has provided large step length in searching behavior in all price prediction reported earlier. A well-balanced EB and OB phase also assist the *e*ABC-LSSVM from falling to local minimum. Thus, more effective capability is possessed in searching for better solution. This feature which is made absents in ABC-LSSVM limited its exploitation process.

In addition, the adaptation of *cm*ABC assists the model from facing over fitting, as occurred in most cases of ABC-LSSVM, except in HO price prediction. This is indicated when one of LSSVM hyper-parameter fell to the minimum boundary value, which is $\sigma^2 = 1$. A small value of σ^2will cause the model to face with over fitting issue, and consequently affect the prediction accuracy obtained significantly (Wu, et al., 2007). Finally, the combination of *lv*ABC and *cm*ABC not only driven the model to achieve higher prediction accuracy and escape from local minimum, but also contributes in avoiding premature convergence situation (see Figure 6.2, 6.6, 6.10 and 6.14).

This experiment also indicated that the prediction accuracy obtained by *e*ABC-LSSVM for PN is 87.7220%, which is lower than the prediction accuracy of the other energy fuels price. However, this is not surprising since PN price possesses lower correlation with other input features as compared to price correlation possesses by CL, HO and HU (see Chapter 3).

6.2.1.2 *e*ABC-LSSVM vs Other Techniques

To further verify superior prediction and generalization performance of *e*ABC-LSSVM, it is further compared with the following hybrid and non-hybrid prediction models:

i. GA-LSSVM (Mustafa, et al., 2012)
ii. CV-LSSVM (Mellit, et al., 2013)
iii. DE-LSSVM (Feoktistov & Janaqi, 2004)
iv. PSO-LSSVM (Jiang & Zhao, 2013)
v. *e*PSO-LSSVM
vi. BPNN (Malliaris & Malliaris, 2008)
vii. SVM (Vapnik, 1995)

6.2.1.2.1 Crude Oil Prices

The predicted results of the optimal hyper-parameters by *e*ABC-LSSVM and other techniques are shown in Table 6.9. As can be seen, the proposed technique outperforms the other techniques when the prediction accuracy achieved was 94.6025%. From 30 run times, the average prediction accuracy obtained was 94.0394% while 93.3037% was recorded as minimum prediction accuracy by *e*ABC-LSSVM. Besides prediction accuracy, the error rates which are relative to MAPE, sMAPE, RMSPE and Theil's U of *e*ABC-LSSVM are also in a good agreement which indicate the superiority of the proposed technique. Among the hybrid LSSVM based techniques; *e*PSO-LSSVM produced the worst result, which is 93.4154% of prediction accuracy while BPNN performed the worst among all. Meanwhile, the SVM is ranked one place higher than BPNN when the prediction accuracy obtained is 93.9283%, which is 2.0659% higher than the prediction accuracy produced by BPNN. The T-test result for this case is also provided, which is reported in Table 6.10.

The visual results for comparisons in price prediction and convergence rate are provided in Figure 6. From the figure, it is depicted that *e*PSO-LSSVM experienced the worst in terms of convergence as compared to other hybrid LSSVM prediction technique.

Table 6.9: CL Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA(%)	Average PA (%)
***e*ABC-LSSVM**	15.2458	1.5832	0.6898	6.0579	**5.3802**	**5.3975**	**94.6025**	**5.3802**	**0.0681**	**0.0336**	93.3037	**94.0394**
GA-LSSVM	26.0286	17.7216	-	8.1275	6.1752	6.3512	93.6488	6.5168	0.0750	0.0393	93.3870	93.4806
PSO-LSSVM	352.3987	1	-	3.3815	5.6519	6.0178	93.9822	6.1512	0.0769	0.0392	92.9529	93.8793
***e*PSO-LSSVM**	806.8757	193.6208	-	8.4438	8.1839	6.5846	93.4154	6.7781	0.0779	0.0410	91.8703	92.932
CV-LSSVM	16.4610	18.1543	-	8.2733	8.0191	6.4917	93.5083	6.6787	0.0766	0.0403	93.5083	93.5083
DE-LSSVM	352.3987	1	-	3.3815	5.6519	6.0178	93.9822	6.1512	0.0769	0.0392	**93.9822**	93.9822
BPNN	-	-	-	**0.4765**	9.162	8.7676	91.2324	9.1162	0.1153	0.0585	46.8791	87.9176
SVM	-	-	-	8.5739	8.3958	6.7017	93.2983	6.8962	0.0788	0.0797	84.8034	90.816

Table 6.10: Significant Test for CL Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.030
eABC-LSSVM – PSO-LSSVM	.010
*e*ABC-LSSVM – *e*PSO-LSSVM	.001
*e*ABC-LSSVM – CV-LSSVM	.002
*e*ABC-LSSVM – DE-LSSVM	.010
*e*ABC-LSSVM – BPNN	.025
*e*ABC-LSSVM – SVM	.002

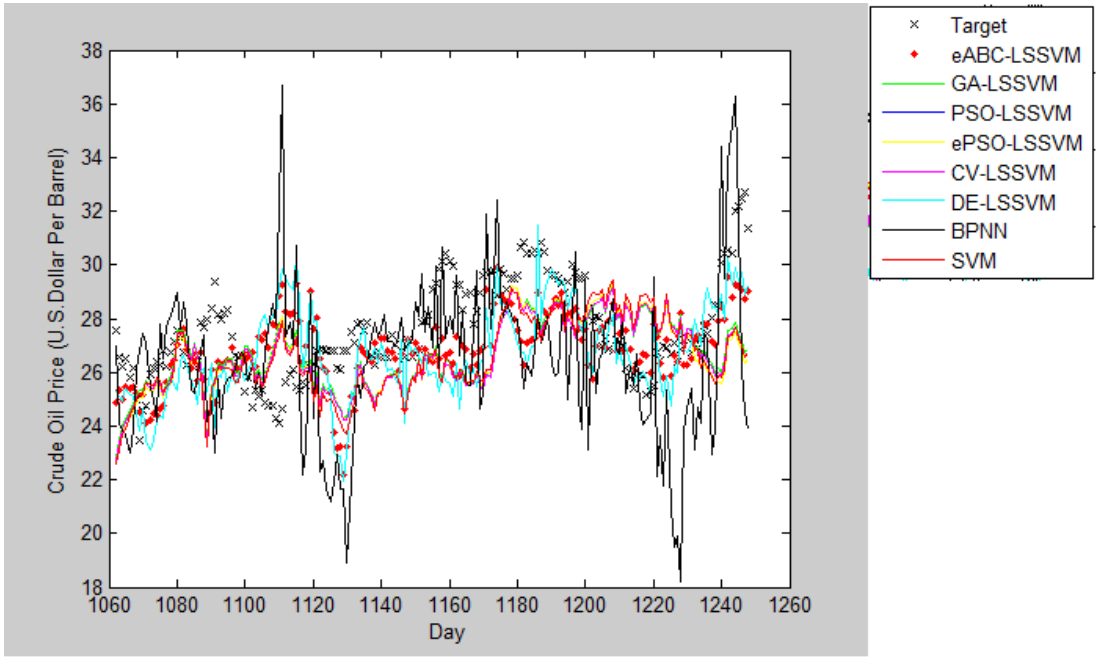


Figure 6.17: Comparison of CL price predictions Data Set A: eABC-LSSVM vs. Other Techniques

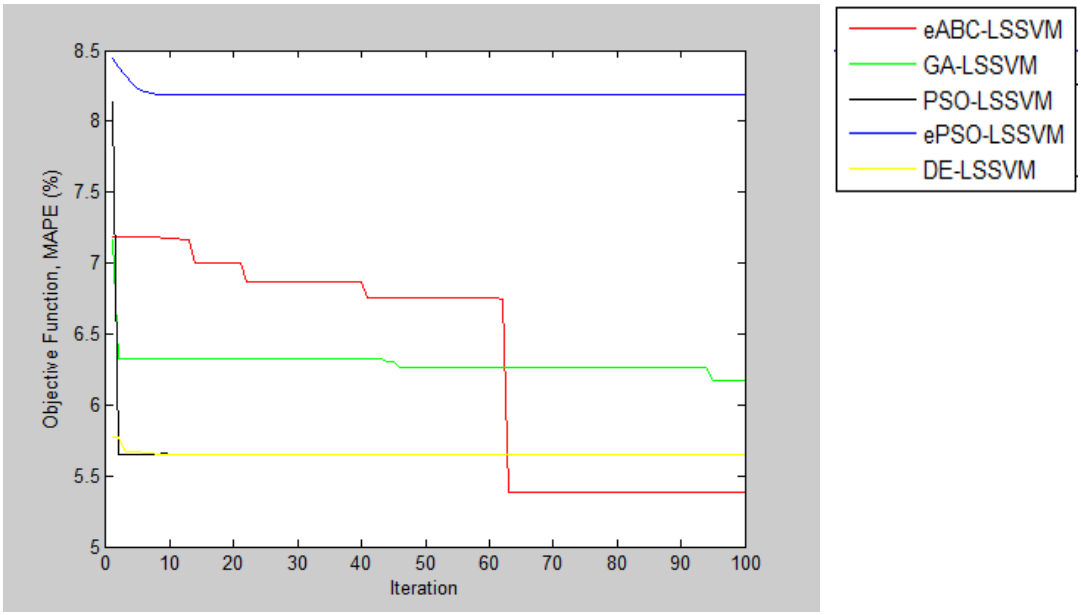


Figure 6.18: Comparison of Convergence for CL Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

6.2.1.2.2 Heating Oil Prices

Similar with previous cases, to demonstrate the effectiveness of *e*ABC-LSSVM for the time series data of interest, both table and graphical results are provided. As can be seen in Table 6.11, the *e*ABC-LSSVM is capable to produce lowest MAPE, sMAPE, RMSPE and Theil's U, which are 6.3310%, 6.2715%, 0.0801% and 0.0405 respectively. By referring to MAPE, this means that the prediction accuracy achieved is 93.6690%.The second place is recorded by SVM with 93.1570% of prediction accuracy, and this is closely followed by *e*PSO-LSSVM. The results also indicate that *e*ABC-LSSVM is not only superior in testing phase, but also outperforms the other techniques in validation set. From the table, it is recorded that *e*ABC-LSSVM is able to converge at 5.7943% of MAPE. The result from significant test is tabulated in Table 6.12. From the table, it clearly shows that good result is obtained for each pair of the comparison.

To visualize the comparison between actual and predicted values in price prediction, the graphical result is provided in Figure 6.19. From the figure, it is depicted that the values produced by BPNN are drifted widely from the actual price at most of the testing period. The following figure, viz. Figure 6.20 illustrates the comparison for convergence among all hybrid LSSVM based techniques. From the figure, the illustration of premature convergence situation that faced by the other techniques are clearly visualized, which finally affect their generalization performance.

Table 6.11: HO Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	80.2488	944.0822	1.8135	8.4098	**5.7943**	**6.3310**	**93.6690**	**6.2715**	**0.0801**	**0.0405**	**92.5443**	**92.8077**
GA-LSSVM	106.695	37.0208	-	7.8572	6.7774	7.6421	92.3579	7.3328	0.0976	0.0463	92.1818	92.3081
PSO-LSSVM	68.8673	1	-	4.6617	5.986	11.5140	88.4860	10.8196	0.1464	0.0678	88.486	88.486
***e*PSO-LSSVM**	54.1174	220.8441	-	8.2522	7.1790	6.875	93.1250	6.7079	0.0878	0.043	89.5130	92.4275
CV-LSSVM	28.3687	18.5478	-	7.8168	7.8684	7.6215	92.3785	7.3144	0.0973	0.0461	92.3785	92.3785
DE-LSSVM	68.8670	1	-	4.6617	5.9860	11.5140	88.4860	10.8196	0.1464	0.0678	88.4860	88.4860
BPNN	-	-	-	**1.2188**	9.1121	15.6724	84.3276	16.6512	0.0926	0.0978	38.3667	75.0055
SVM	-	-	-	7.7719	7.8536	6.8430	93.1570	6.7103	0.0862	0.0549	87.9134	90.2942

Table 6.12: Significant Test for HO Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.000
eABC-LSSVM – PSO-LSSVM	.000
*e*ABC-LSSVM – *e*PSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.000
*e*ABC-LSSVM – DE-LSSVM	.000
*e*ABC-LSSVM – BPNN	.000
*e*ABC-LSSVM – SVM	.000

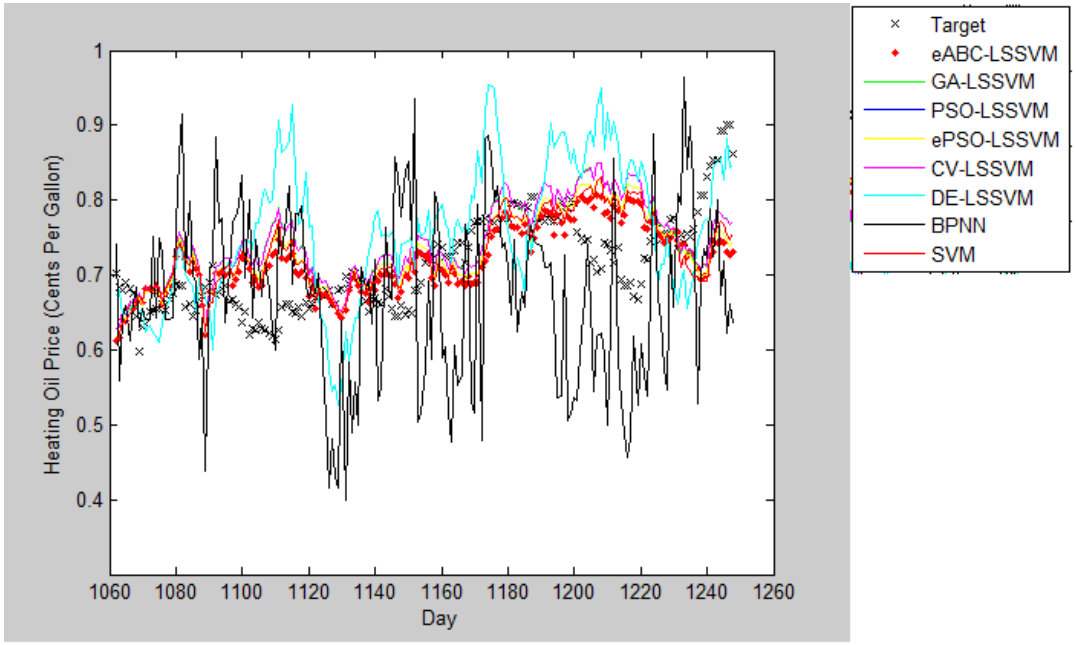


Figure 6.19: Comparison of HO price predictions Data Set A: eABC-LSSVM vs. Other Techniques

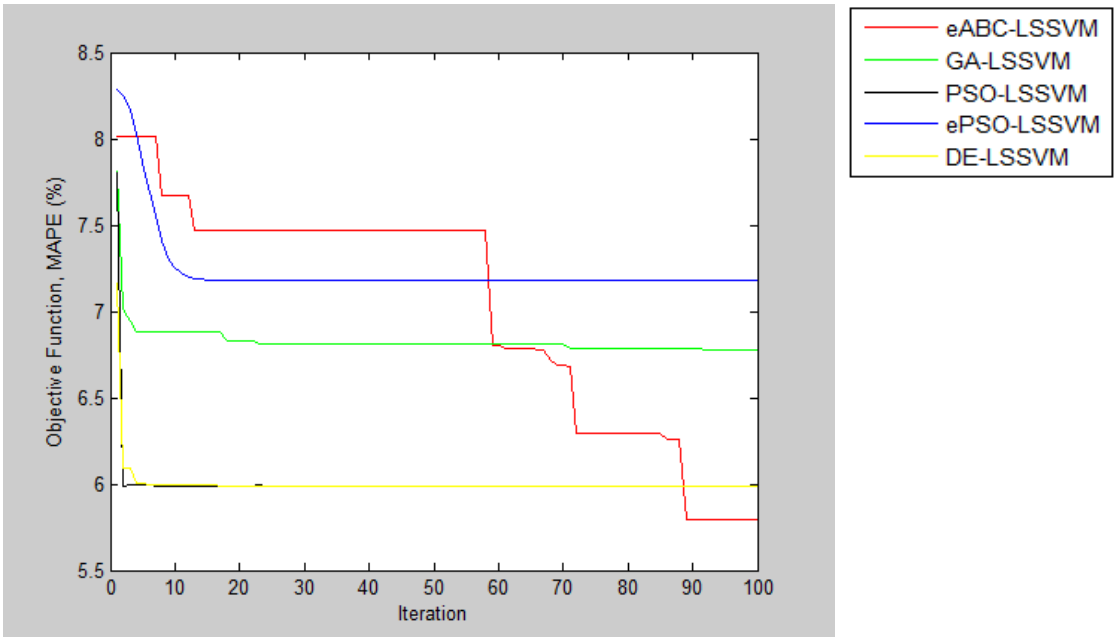


Figure 6.20: Comparison of Convergence for HO Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

6.2.1.2.3 Gasoline Prices

Comparison results for HU price prediction for Set A by all identified prediction techniques are tabulated in Table 6.13. The findings of this experiment revealed that the highest prediction accuracy was obtained by *e*ABC-LSSVM while the lowest was BPNN with 94.5376% and 87.6717% respectively. In terms of error rate, the MAPE produced by both techniques are 5.4624% and 12.3283% respectively. The second highest prediction accuracy was recorded by *e*PSO-LSSVM; differing 0.1034% from *e*ABC-LSSVM. Meanwhile, almost similar results were produced by GA-LSSVM and CV-LSSVM, which is around 94.3%. The T-test result for this case is as tabulated in Table 6.14.

The graphical results to illustrate the difference between actual and predicted HU prices for all techniques are shown in Figure 6.21 while the following figure, namely Figure 6.22 illustrates the comparison of convergence among the identified techniques. From Figure 6.21, the inaccurate results produced by BPNN are clearly illustrated since it is the only prediction technique that recorded prediction accuracy lower that 90%.

Table 6.13:HU Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	121.3982	77.6900	1.3667	10.1283	**6.5947**	**5.4624**	**94.5376**	**5.5067**	**0.0693**	**0.0356**	94.2743	**94.4717**
GA-LSSVM	611.6890	88.6796	-	9.9076	8.2611	5.6161	94.3839	5.6360	0.0703	0.0359	93.9884	94.2857
PSO-LSSVM	378.2037	1	-	**4.0086**	6.8074	6.3782	93.6218	6.5790	0.0821	0.0423	92.7603	93.5931
***e*PSO-LSSVM**	485.7452	977.7914	-	10.4904	9.9817	5.5658	94.4342	5.6750	0.0706	0.0369	88.7528	94.0781
CV-LSSVM	29.0663	12.3566	-	9.5160	9.5165	5.6800	94.3200	5.6566	0.0712	0.0357	**94.3200**	94.3200
DE-LSSVM	378.2037	1	-	**4.0086**	6.8074	6.3782	93.6218	6.5790	0.0821	0.0423	93.6218	93.6218
BPNN	-	-	-	4.3353	7.3892	12.3283	87.6717	13.6021	0.1533	0.0835	75.7627	82.3248
SVM	-	-	-	10.3146	10.2114	6.1658	93.8342	6.3908	0.0748	0.0517	81.4876	89.5868

Table 6.14: Significant Test for HU Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.000
*e*ABC-LSSVM – PSO-LSSVM	.007
*e*ABC-LSSVM – ePSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.000
*e*ABC-LSSVM – DE-LSSVM	.007
*e*ABC-LSSVM – BPNN	.000
*e*ABC-LSSVM – SVM	.000

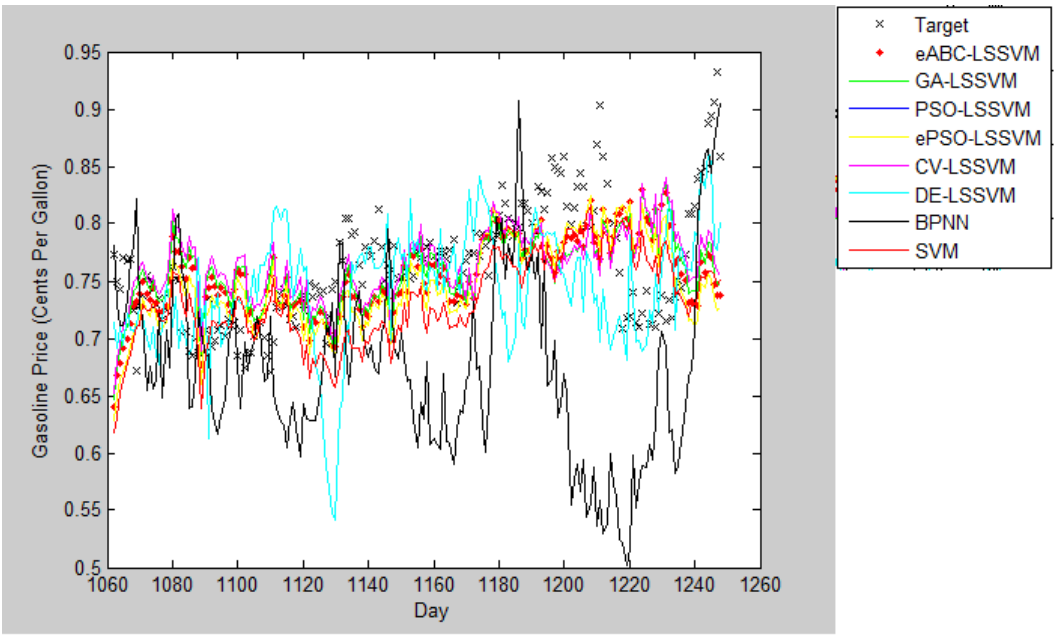


Figure 6.21: Comparison of HU Price Predictions Data Set A: eABC-LSSVM vs. Other Techniques

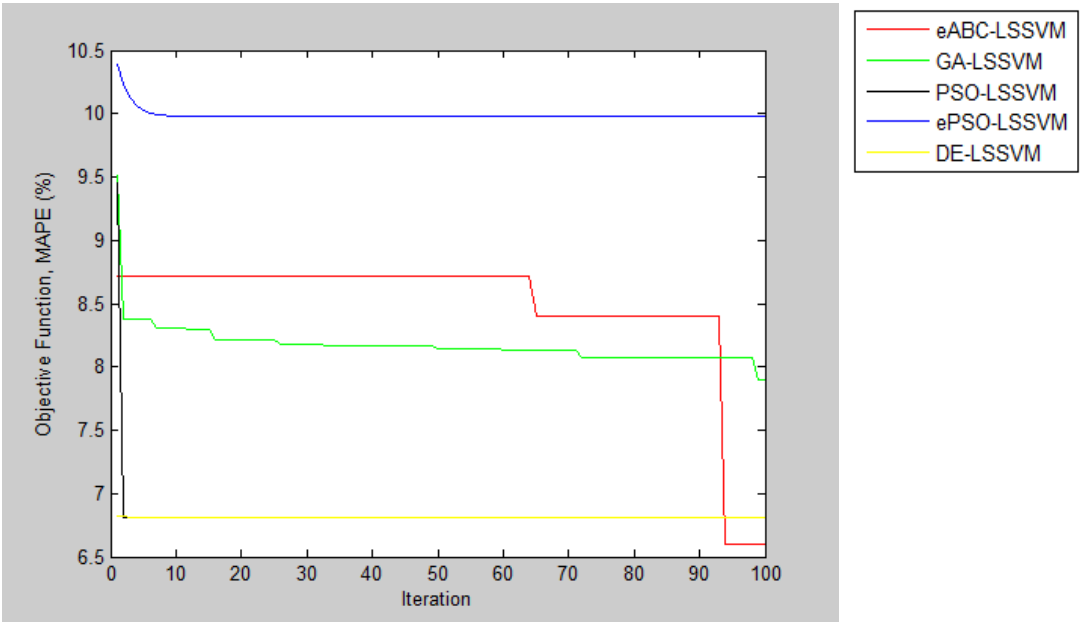


Figure 6.22: Comparison of Convergence for HU Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

6.2.1.2.4 Propane Prices

As shown in Table 6.15, the proposed *e*ABC-LSSVM consistently outperforms other techniques, regardless of prediction evaluation metrics. It can be clearly seen that the MAPE achieved by *e*ABC-LSSVM is 0.7015% lower from the closest competitor, which is *e*PSO-SSVM. As occurred in previous cases, BPNN yielded the highest MAPE, which is 19.4931%. The significant test for each difference with other techniques is provided in Table 6.16. For visualization purposes, the illustration of results obtained in Table 6.15 is depicted in Figure 6.23 while the convergence rate is shown in Figure 6.24. From Figure 6.24, it obviously illustrates the situation of GA-LSSVM, PSO-LSSVM, *e*PSO-LSSVM and DE-LSSVM which suffered from premature convergence, which is not the case in *e*ABC-LSSVM.

Table 6.15: PN Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	48.6424	536.4877	1.6454	12.2363	**6.3979**	**12.2780**	**87.7220**	**11.4125**	**0.1443**	**0.0631**	84.5807	**86.7378**
GA-LSSVM	69.4475	25.7831	-	11.4019	8.4047	14.3999	85.6001	13.1707	0.1684	0.0733	83.1855	84.8677
PSO-LSSVM	296.1088	1	-	3.9887	7.7232	18.1271	81.8729	16.1029	0.2298	0.0995	81.8729	81.8729
***e*PSO-LSSVM**	167.1616	624.6111	-	12.1766	12.152	12.9795	87.0205	12.005	0.1522	0.0666	80.6377	86.4587
CV-LSSVM	19.8276	10.3910	-	11.0659	11.6603	15.1721	84.8279	13.815	0.1768	0.0767	**84.8279**	84.8279
DE-LSSVM	296.1043	1	-	3.9888	7.7232	18.1271	81.8729	16.1029	0.2298	0.0995	81.8729	81.8729
BPNN	-	-	-	**3.5874**	8.8467	19.4931	80.5069	17.3279	0.2524	0.1096	60.0859	61.4279
SVM	-	-	-	11.6589	11.9021	15.0173	84.9827	13.7283	0.1742	0.0844	83.1624	86.5145

Table 6.16: Significant Test for PN Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.000
*e*ABC-LSSVM – PSO-LSSVM	.000
*e*ABC-LSSVM – *e*PSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.000
*e*ABC-LSSVM – DE-LSSVM	.000
*e*ABC-LSSVM – BPNN	.006
*e*ABC-LSSVM – SVM	.000

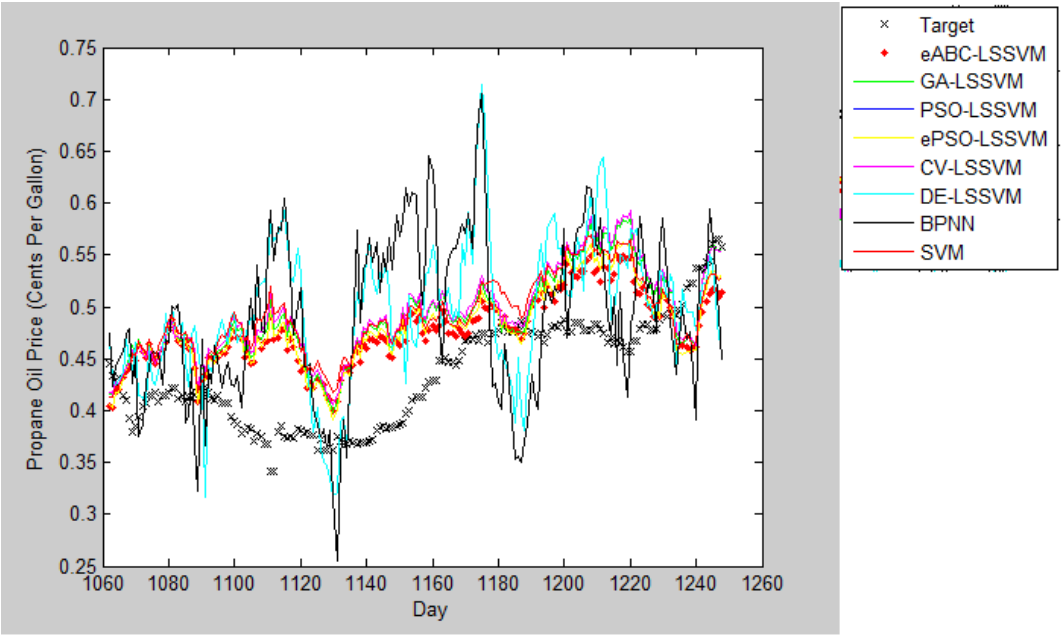


Figure 6.23: Comparison of PN Price Predictions Data Set A: eABC-LSSVM vs. Other Techniques

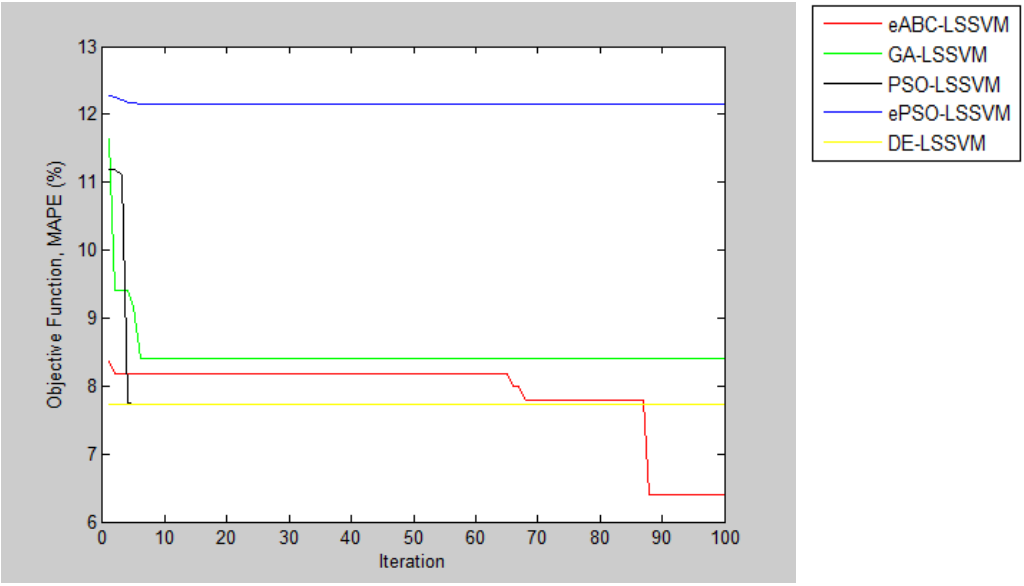


Figure 6.24: Comparison of convergence rate for PN Price Prediction Data Set A: eABC-LSSVM vs. Other Techniques

6.2.1.2.5 Results Analysis for Data Set A: *e*ABC-LSSVM vs. Other Techniques

Based on the results obtained (see Section 6.2.1.2), it is safe to conclude that *e*ABC-LSSVM is able to prove its superiority by outperforming the other identified prediction techniques. The efficiency of *e*ABC-LSSVM is proven by its capability in escaping premature convergence in all cases and producing good generalization in testing phase. Meanwhile, the over fitting problem experienced by PSO-LSSVM and DE-LSSVM in all prediction prices caused the model to suffer in unsatisfying generalization performance, especially in three last cases, viz. HO, HU and PN prices predictions (see sections 6.2.1.2.2, 6.2.1.2.3 and 6.2.1.2.4 respectively) . This situation is indicated by minimumσ^2 obtained. However, it is worth noting that minimum or maximum value of the parameter of interest does not necessary give the optimal solution (Lendasse, et al., 2005; Wu, et al., 2007). This explained the poor generalization performance of the technique.

Apart from PSO-LSSVM and DE-LSSVM, the BPNN also shown to be vulnerable with over fitting which is interpreted by good performance during training but poor upon new data is presented. This situation is expected due to the ERM principle employed in BPNN. Interestingly, this problem is addressed in SRM where it tends to minimize the generalization error rather than training error as offered in ERM. As stated in the previous chapter (see Chapter 1), the benefit of SRM is embodied in SVM, which is later inherited by LSSVM based techniques. Due to the difference in the principle governed by BPNN, it is also can be seen that, in all energy fuel price prediction, BPNN performed the worst as compared to the other SRM based

techniques. In addition, it is well documented that BPNN requires large historical data sets (up to 10 years daily data) to ensure it performs well in prediction task (Haidar & Wolff, 2011; Zhang, et al., 1998). Nonetheless, this is not the case in LSSVM based prediction technique (Xiang & Jiang, 2009).

6.2.2 Data Set B

Similar with experiments conducted in Data Set A, the prediction for each energy fuels price (CL, HO, HU and PN) was performed separately.

6.2.2.1 *e*ABC-LSSVM vs ABC-LSSVM and Its Variants

This section discusses obtained results by *e*ABC-LSSVM against variants ABC-LSSVM.

6.2.2.1.1 Crude Oil Prices

The empirical results for CL price prediction for Set B are shown in Table 6.17. From the table, the prediction accuracy produced by *e*ABC-LSSVM was 92.8728%, which is 6.8402% higher from ABC-LSSVM. In terms of sMAPE, RMSPE and Theil's U, the *e*ABC-LSSVM maintains its position by yielding 7.3761%, 0.0887% and 0.0518 respectively, which are the lowest among other techniques. Meanwhile, the prediction accuracy recorded by both *lv*ABC-LSSVM and *cm*ABC-LSSVM are 91.9582% and 88.0461% respectively. For the T-test results, as reported in CL price prediction Set A, positive results were also obtained in Set B (see Table 6.18). For graphical results, both comparisons in price prediction and convergence rate are illustrated in Figure 6.25 and 6.26 respectively. In Figure 6.25, it is clearly depicted that ABC-LSSVM, which is represented in blue line is incapable to tackle the CL price pattern, especially during the rising trend. Similar situation is also portrayed by *cm*ABC-LSSVM which is marked with green line. Meanwhile, in Figure 6.27, it can be seen that the exploitation process performed by *e*ABC-LSSVM covered wider search space before the best solution is found. On the other hand, the exploitation process carried out by ABC-LSSVM focuses on the narrower space which consequently resulted in the unsatisfied solutions obtained (see Figure 6.28).

Table 6.17: CL Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	602.5196	540.1780	440.6648	359.7949
σ^2	938.8506	251.8873	4.0518	1
α	1.0547	1.3945	-	-
MAPE Training (%)	6.2916	6.0509	3.6706	**1.8663**
MAPE Validation (%)	**3.8604**	3.9247	3.9532	3.9242
MAPE Testing (%)	**7.1272**	8.0418	11.9539	13.9674
PA (%)	**92.8728**	91.9582	88.0461	86.0326
sMAPE (%)	**7.3761**	8.5086	13.7635	16.0676
RMSPE (%)	**0.0887**	0.1042	0.1784	0.1952
Theil's U	**0.0518**	0.0628	0.1080	0.1202
Minimum PA (%)	**89.8010**	85.2335	82.7870	84.7656
Average PA (%)	**92.2791**	88.4277	86.4171	85.5033

Table 6.18: Significant Test for CL Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.000
*e*ABC-LSSVM – ABC-LSSVM	.000

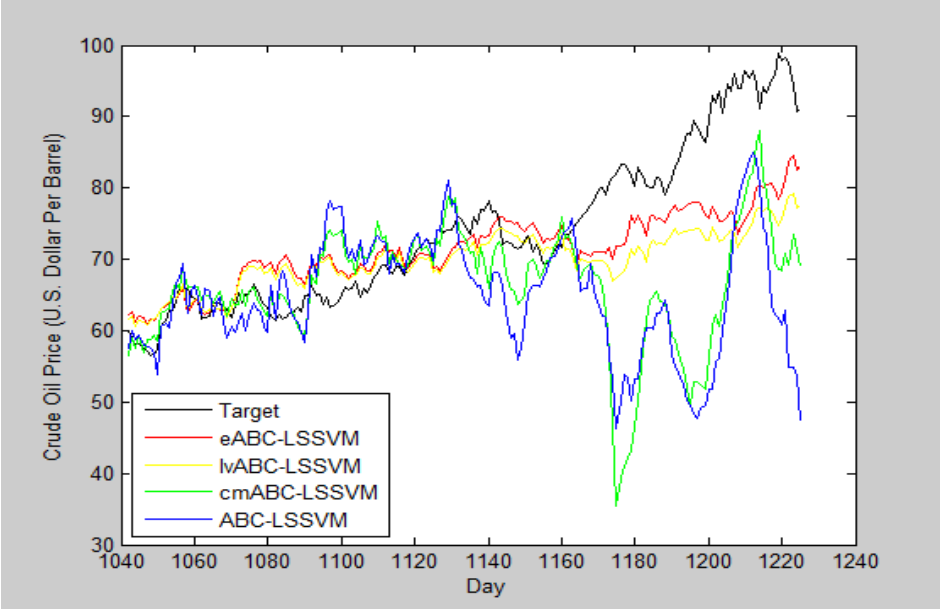


Figure 6.25: Comparison of CL Price Predictions Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

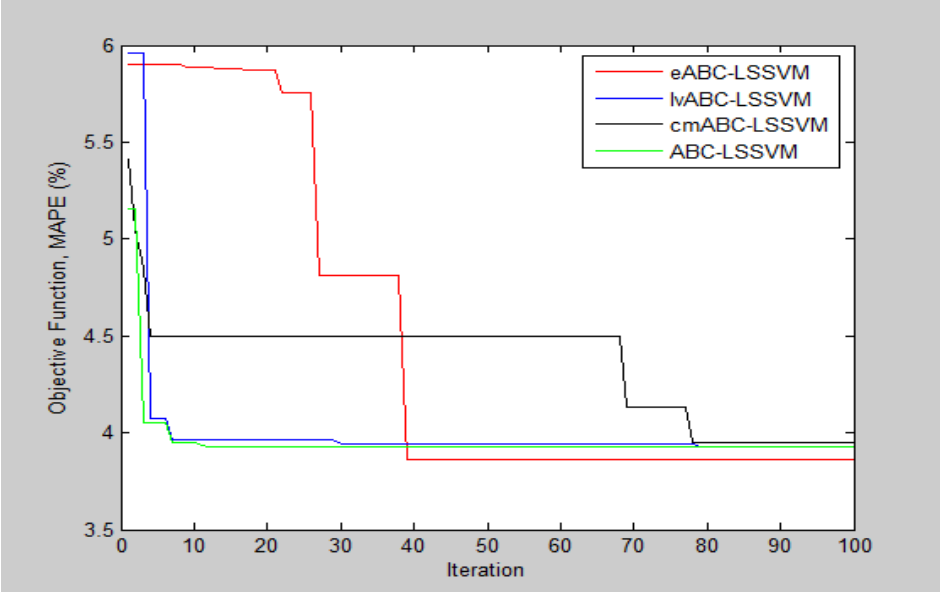
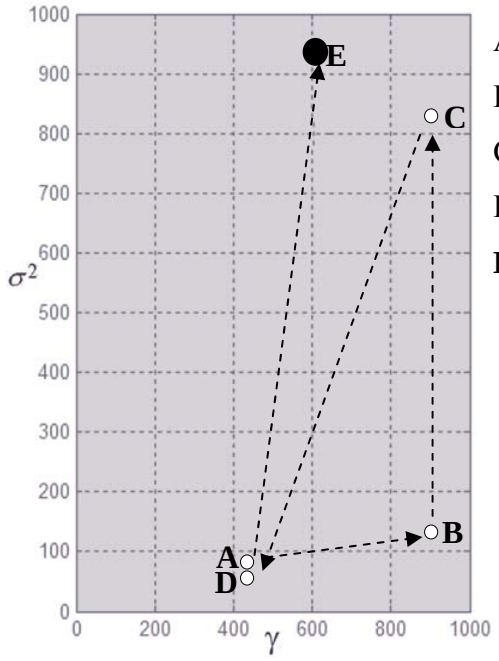


Figure 6.26: Comparison of Convergence for CL Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants



A [436.599, 95.0229]

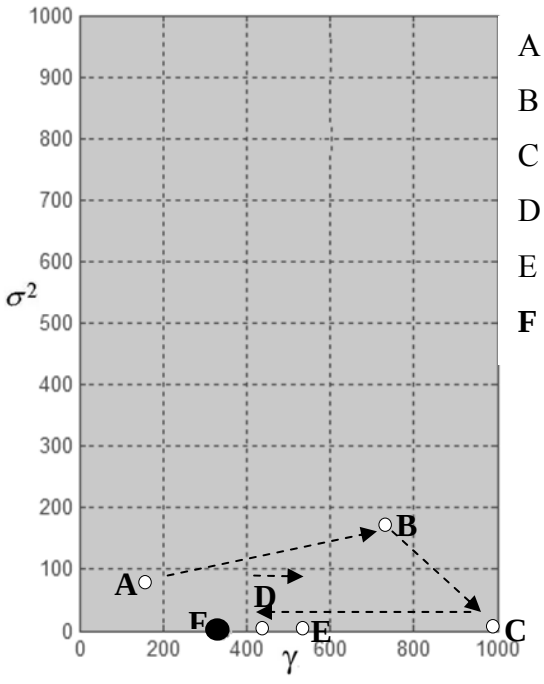
B [935.13, 133.053]

C [948.817, 812.331]

D [437.199, 87.0398]

E [602.5196, 938.8506] *Optimal γ and σ^2

Figure 6.27: Exploitation of Search Space by eABC-LSSVM in CL Price Prediction Data Set B



A [172.966, 80.76]

B [745.371, 165.792]

C [1000, 1]

D [450.512, 1]

E [525.877, 1]

F [359.795, 1]*Optimal γ and σ^2

Figure 6.28: Exploitation of Search Space by ABC-LSSVM in CL Price Prediction Data Set B

6.2.2.1.2 Heating Oil Prices

The performance comparison of four prediction techniques on five indices of HO prices is reported in Table 6.19. The prediction accuracy recorded by ABC-LSSVM was 86.9098% which reflects the MAPE obtained, which is 13.0902%. This is lesser than *e*ABC-LSSVM which was able to obtain 91.5382% of prediction accuracy. The final values of γ and σ^2 for both techniques are [235.4062, 1] and [260.8178, 768.3368] respectively. The closest competitor to *e*ABC-LSSVM was *lv*ABC-LSSVM with 90.8923% of prediction accuracy. Meanwhile, *cm*ABC-LSSVM was able to achieve 88.8399% of prediction accuracy, which is 1.9301% increment from the prediction accuracy recorded from its original form. Besides MAPE and prediction accuracy, consistent results were also recorded in terms of sMAPE, RMSPE and Theil's U in all experimented techniques.

The results for T-test is tabulated in Table 6.20 while the graphical presentation of comparison between the actual prices of HO with predicted from day 1042 to 1225 is shown in Figure 6.29. The following figure, namely Figure 6.30 gives the illustration of the whole optimization process of MAPE in the time series data of interest. The last two figures, viz. Figure 6.31 and 6.32 depicted the graphical coordinate for both sets of *e*ABC-LSSVM and ABC-LSSVM in exploitation process. From Figure 6.31, it can be seen that the *e*ABC-LSSVM was able to find the optimal solution which is far from the initial position. On the other hand, in ABC-LSSVM, the searching process is limited in the small area, which eventually stuck in local minimum.

Table 6.19: HO Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	260.8178	162.9918	745.6910	235.4062
σ^2	768.3368	237.8871	3.6491	1
α	0.7813	1.1842	-	-
MAPE Training (%)	9.1302	8.9097	5.0993	**3.1480**
MAPE Validation (%)	**5.3012**	5.9705	6.1234	5.9532
MAPE Testing (%)	**8.4618**	9.1077	11.1601	13.0902
PA (%)	**91.5382**	90.8923	88.8399	86.9098
sMAPE (%)	**8.9332**	9.6475	12.2339	14.4654
RMSPE (%)	**0.0986**	0.1039	0.1413	0.1585
Theil's U	**0.0526**	0.0560	0.0746	0.0860
Minimum PA (%)	86.4657	**86.5037**	82.3658	84.3578
Average PA (%)	**89.0430**	88.0519	85.9887	86.7393

Table 6.20: Significant Test for HO Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.009
*e*ABC-LSSVM – ABC-LSSVM	.000

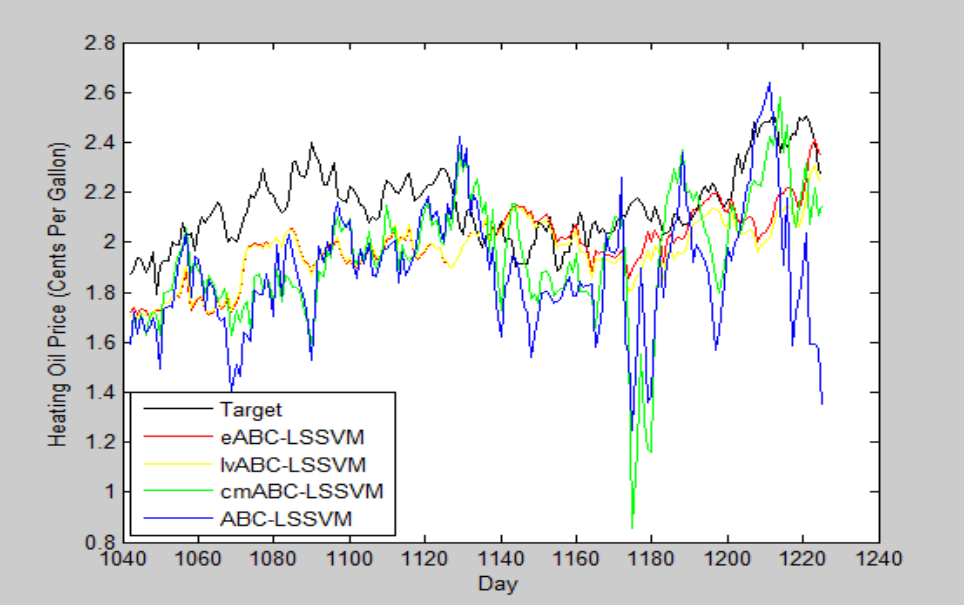


Figure 6.29: Comparison of HO Price Predictions Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

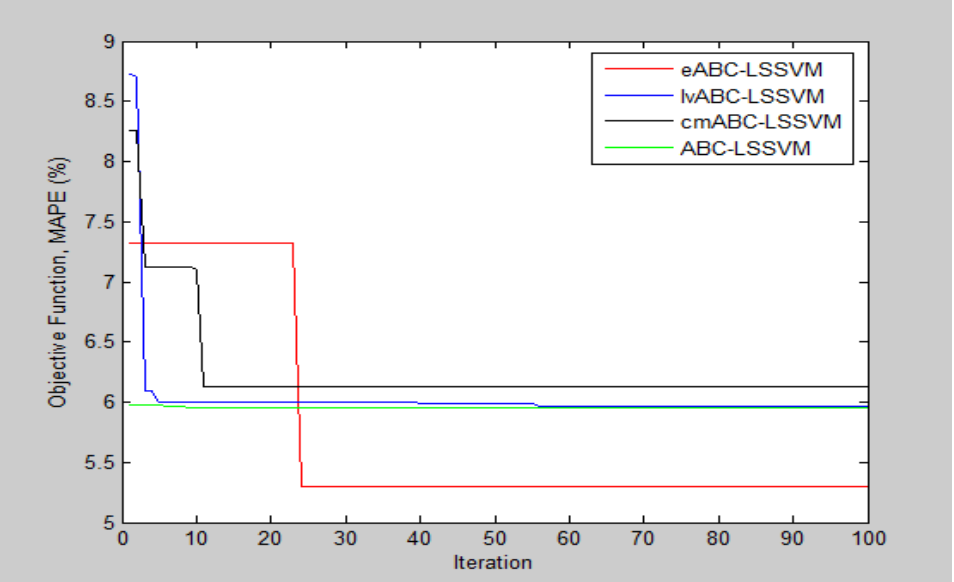
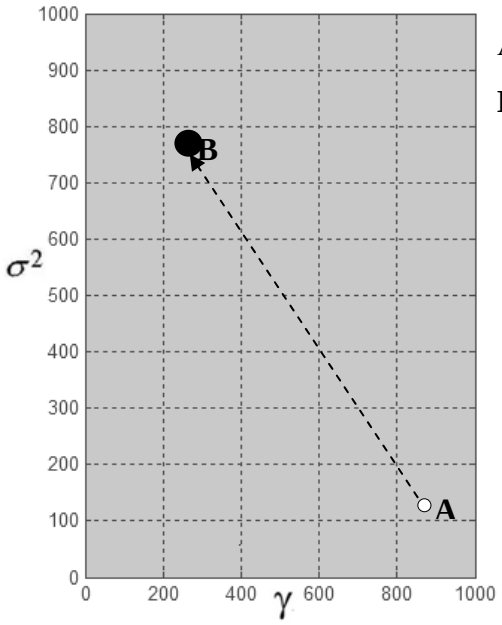


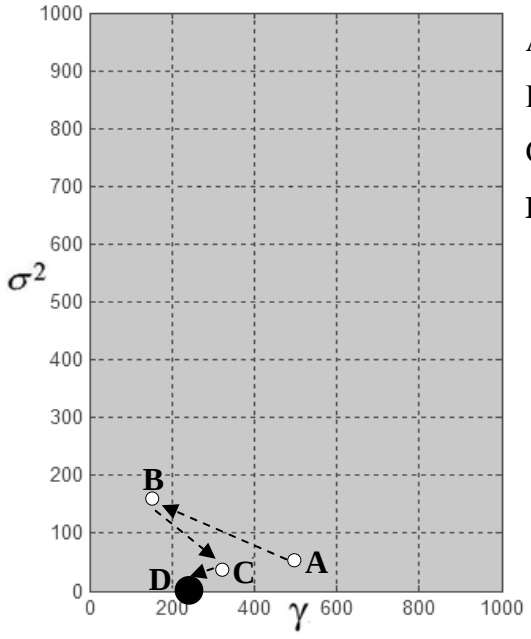
Figure 6.30: Comparison of Convergence for HO Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants



A [882.579, 10.9109]

B [260.818, 768.337] *Optimal γ and σ^2

Figure 6.31: Exploitation of Search Space by eABC-LSSVM in HO Price Prediction Data Set B



A [477.402, 50.0335]

B [181.891, 154.791]

C [326.459, 15.9157]

D [235.4062, 1] *Optimal γ and σ^2

Figure 6.32: Exploitation of Search Space by ABC-LSSVM in HO Price Prediction Data Set B

6.2.2.1.3 Gasoline Prices

Results depicted in Table 6.21 were recorded from HU price prediction for Set B. The lowest MAPE, sMAPE, RMSPE and Theil's U were achieved by *e*ABC-LSSVM with 7.5994%, 8.1161%, 0.09833% and 0.0585 respectively. This is followed by *lv*ABC-LSSVM with 7.9995% of MAPE while the other metrics indicated good agreement with MAPE recorded. These reflect the prediction accuracy obtained by both techniques, which are 92.4006% and 92.0005% respectively. Meanwhile, for *cm*ABC-LSSVM, prediction accuracy recorded was 90.6271%, which is 1.9503% higher than prediction accuracy obtained by ABC-LSSVM. The results from T-test are as tabulated in Table 6.22. To show the comparison between actual and predicted prices of HU by each prediction technique, the graphical result is provided in Figure 6.33. Meanwhile, in Figure 6.34, it can be seen that the *e*ABC-LSSVM was able to converge at the lowest MAPE at iteration 19.

For further analysis, the progress in exploitation process for both *e*ABC-LSSVM and ABC-LSSVM were visualized in Figure 6.35 and 6.36 respectively. By observing the figures, it is shown how the *e*ABC-LSSVM harness the search space before the optimal solution, which is C [145.313, 683.332] is found. On the other hand, in ABC-LSSVM, search space wastage occurred when the bees limited its searching process which resulted the technique to trap at E [247.216, 1].

Table 6.21: HU Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	145.3133	281.1954	987.4984	247.2162
σ^2	683.3321	374.4754	1.7843	1
α	1.0439	0.4821	-	-
MAPE Training (%)	7.5074	7.3384	2.6404	**2.4635**
MAPE Validation (%)	**4.3421**	4.4254	4.6067	4.4234
MAPE Testing (%)	**7.5994**	7.9995	9.3729	11.3232
PA (%)	**92.4006**	92.0005	90.6271	88.6768
sMAPE (%)	**8.1161**	8.5597	10.1592	12.5798
RMSPE (%)	**0.09833**	0.1021	0.1195	0.1487
Theil's U	**0.0585**	0.0605	0.0680	0.0876
Minimum PA (%)	**91.7003**	86.2916	85.8265	88.1739
Average PA (%)	**92.1654**	89.3517	88.5889	88.7281

Table 6.22: Significant Test for HU Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.004
*e*ABC-LSSVM – ABC-LSSVM	.000

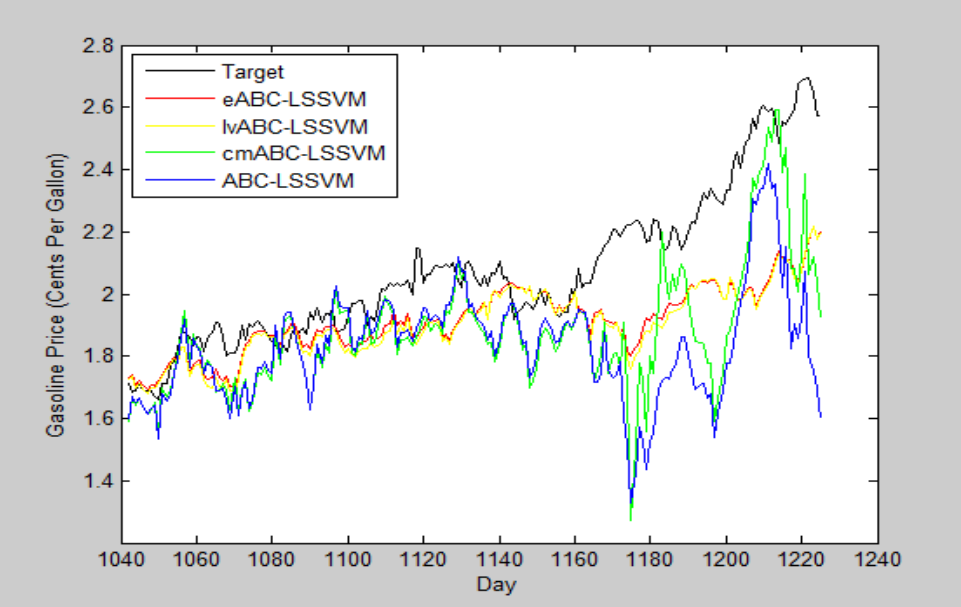


Figure 6.33: Comparison of HU Price Predictions Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

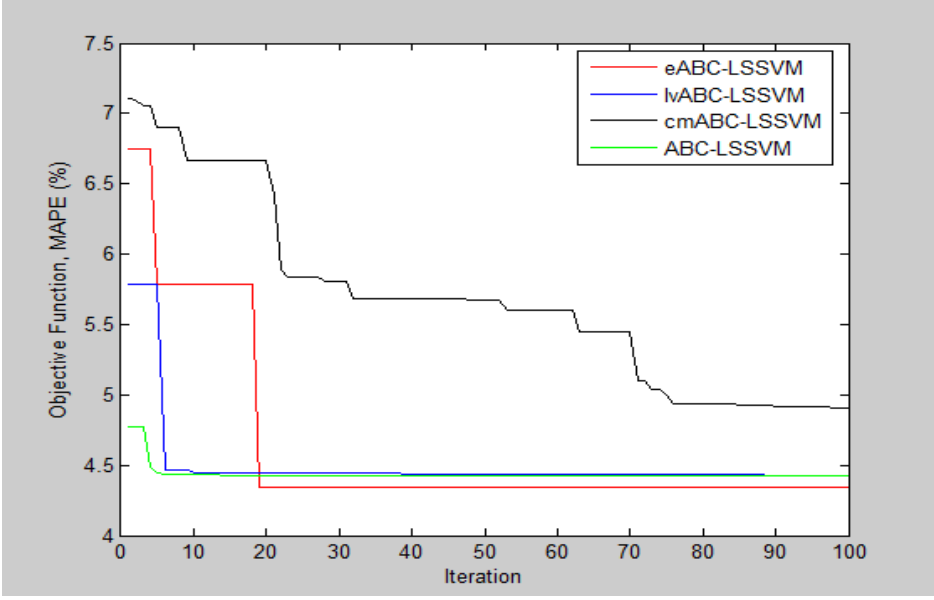
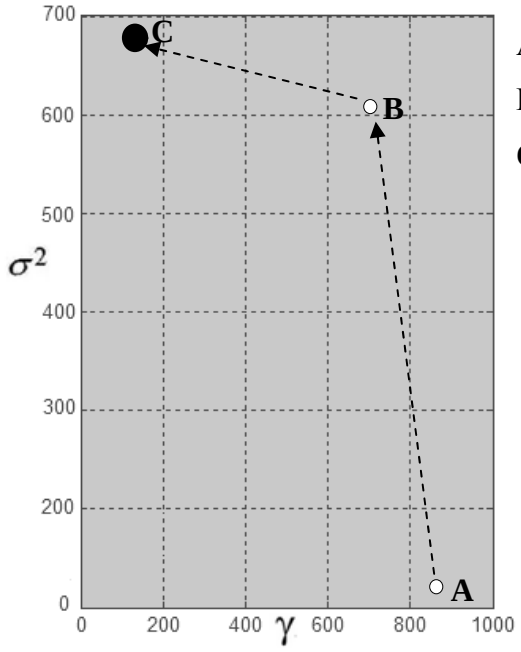


Figure 6.34: Comparison of Convergence for HU Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

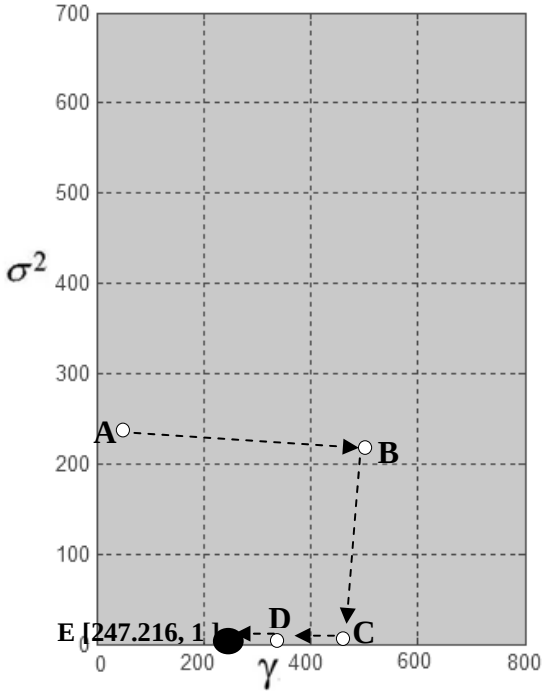


A [877.343, 65.8975]

B [715.714, 609.443]

C [145.313, 683.332] *Optimal γ and σ^2

Figure 6.35: Exploitation of Search Space by eABC-LSSVM in HU Price Prediction Data Set B



A [62.34, 237.693]

B [569.789, 207.897]

C [429.088, 1]

D [336.35, 1]

E [247.216, 1] *Optimal γ and σ^2

Figure 6.36: Exploitation of Search Space by ABC-LSSVM in HU Price Prediction Data Set B

6.2.2.1.4 Propane Prices

Based on Table 6.23, the lowest MAPE is obtained by *e*ABC-LSSVM, which is 9.0805% when the value of γ and σ^2 is set to 563.1288 and 673.5987 respectively. For *lv*ABC-LSSVM and *cm*ABC-LSSVM, both prediction techniques are capable to achieve 9.3492% and 9.7160% of MAPE respectively. However, for ABC-LSSVM, the result of MAPE is recorded at 17.8705%, which is higher from its variant. Nonetheless, this is expected when over fitting problem encountered in ABC-LSSVM where it is capable to perform the best in training set, but poor in testing set. This situation is indicated when a small value is set to σ^2. For significant test result, it can be seen in Table 6.24 while the graphical results for comparison between different techniques for both price prediction and convergence are depicted in Figure 6.37 and 6.38 respectively. From Figure 6.37, poor prediction performance by ABC-LSSVM was clearly depicted, which was caused by over fitting problem.

The visual progress of the bees in exploitation process for *e*ABC-LSSVM can be seen in Figure 6.39, while for ABC-LSSVM it is shown in Figure 6.40. From Figure 6.39, it is shown that after two small steps from position A to C by, the *e*ABC-LSSVM fully exploits the search space by performing larger step length and consequently is capable to obtain optimal value of γ and σ^2, which is at D [563.129, 673.599]. On the other hand, due to limited searching behavior in ABC-LSSVM, the optimal results is incapable to be achieved, instead, it is trapped in local minimum, which is at C [159.6119,1]. This finally causes this technique to face with premature convergence, as illustrate in Figure 6.38.

Table 6.23: PN Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	563.1288	568.8987	837.4621	159.6119
σ^2	673.5987	385.2584	333.9604	1
α	0.7700	1.6191	-	-
MAPE Training (%)	6.8133	6.7283	6.6598	**2.5982**
MAPE Validation (%)	**3.8136**	3.9216	5.0511	3.9195
MAPE Testing (%)	**9.0805**	9.3492	9.7160	17.8705
PA (%)	**90.9195**	90.6508	90.2840	82.1295
sMAPE (%)	**9.7614**	10.0947	10.5466	20.8082
RMSPE (%)	**0.1117**	0.1166	0.1226	0.2205
Theil's U	**0.0671**	0.0703	0.0741	0.1361
Minimum PA (%)	**87.2835**	79.9520	83.4929	79.8323
Average PA (%)	**89.9537**	87.1779	87.5765	80.3401

Table 6.24: Significant Test for PN Price Prediction Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.000
*e*ABC-LSSVM – ABC-LSSVM	.000

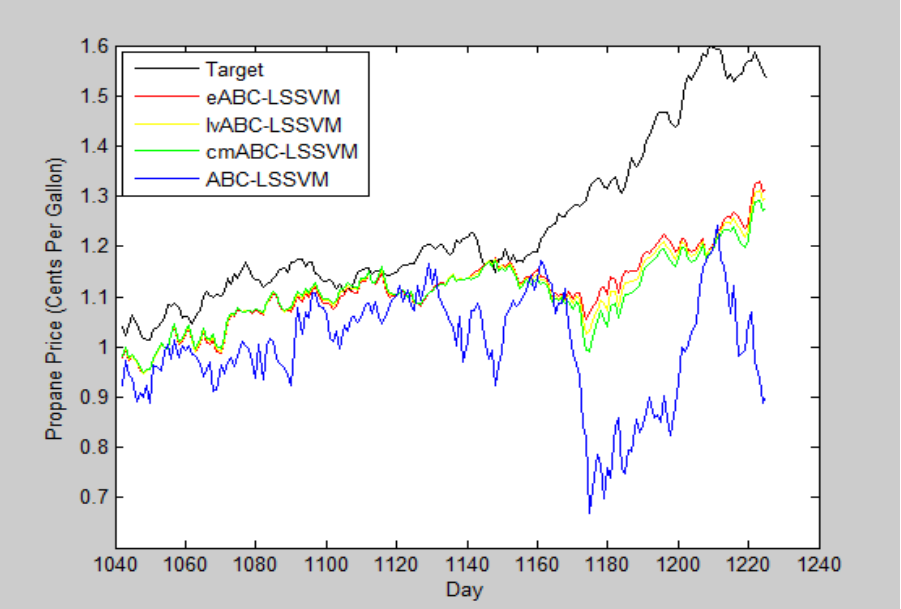


Figure 6.37: Comparison of PN Price Predictions Data Set B: eABC-LSSVM vs. ABC-LSSVM and Its Variants

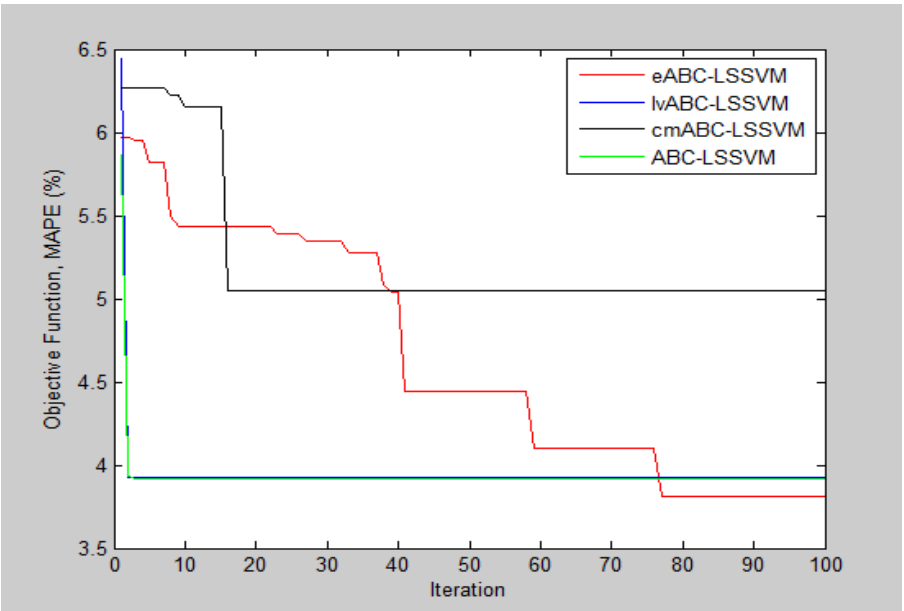
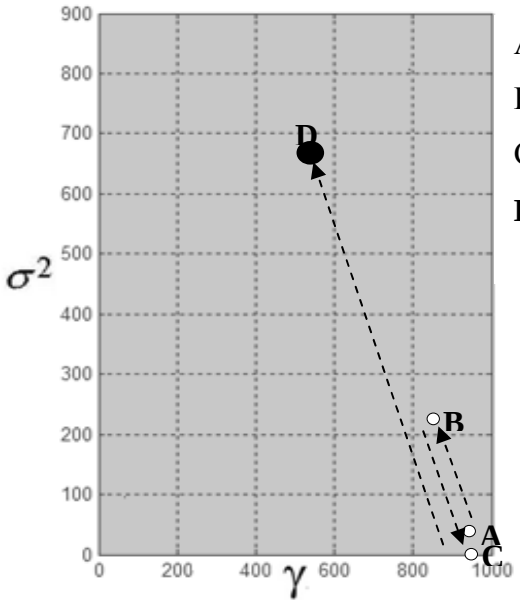


Figure 6.38: Comparison of Convergence Rate for PN Price Prediction Data Set B: eABC-LSSVM vsABC-LSSVM and Its Variants



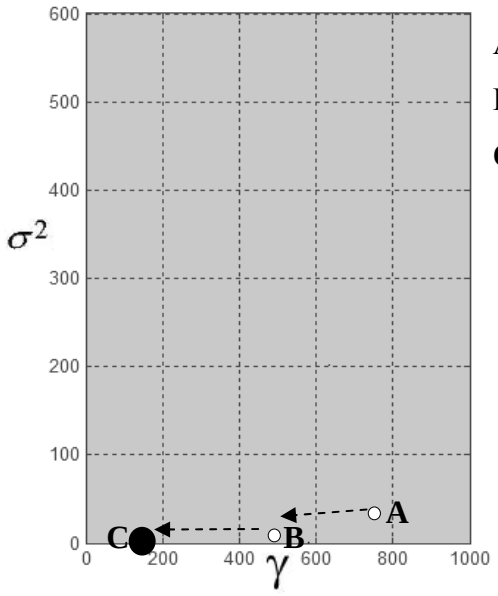
A [974.707, 33.9342]

B [857.677, 238.89]

C [976.56, 1.4885]

D [563.129, 673.599] *Optimal γ and σ^2

Figure 6.39: Exploitation of Search Space by eABC-LSSVM in PN Price Prediction Data Set B



A [797.996, 24.4519]

B [506.944, 1]

C [159.612, 1] *Optimal γ and σ^2

Figure 6.40: Exploitation of Search Space by ABC-LSSVM in PN Price Prediction Data Set B

6.2.2.1.5 Results Analysis for Data Set B: *e*ABC-LSSVM vs. ABC-LSSVM and Its Variants

Similarly with energy fuel prices prediction for Data Set A, the results obtained in Data Set B are also in favor to *e*ABC-LSSVM when it is able to produce the lowest prediction error, regardless of statistical metrics utilized. The superiority of *e*ABC-LSSVM is due to the employment of Levy mutation in assisting the model for better searching activity. This increases the chance of escaping from revising the same solution. Meanwhile, a good balance between EB and OB phase also facilitates the *e*ABC-LSSVM to seek for an optimal solution and at the same time reduce the risk of facing with the premature convergence.

On the other hand, the decision making process which involved conventional mutation helps the model from being exposed to over fitting, as occurred in all cases of ABC-LSSVM. Due to over fitting issue, ABC-LSSVM faced poor generalization, especially in PN price prediction which is clearly depicted in Figure 6.37. The shortcomings of ABC-LSSVM is also proven when it is able to outperform in training sets but incapable to sustain the performance when unseen data sets are presented. This situation is applied in all cases of ABC-LSSVM (see Table 6.17, 6.19, 6.21 and 6.23).

Upon completing the experiment for set B, it also can be seen that the prediction accuracy obtained by *e*ABC-LSSVM for PN was 90.9195%, which is higher that the results recorded for PN price prediction in Set A. This is expected due to higher

correlation possessed between PN and other input features in Set B as compared to in Set A (see Chapter 3).

The results also indicated that the addition of interest rate as input in Data Set B did not contribute to any increment in prediction accuracy. Even though the correlation between the IR and the energy fuels price is high enough (i.e 0.8 - 1) (see Chapter 3, section 3.3.1), the correlation among variables does not mean causation (Siew. A, personal communication, November 7, 2013; Taylor & Weerapana, 2012). This implies that even the inputs and interest rate are correlated, but it does not necessarily mean that one causes the other. However, it is noted that correlations are beneficial in prediction even though there is no causal relationship between the variables (Hyndman & Athanasopoulos, 2013)

6.2.2.2 *e*ABC-LSSVM vs Other Techniques

This section presents the results obtained by *e*ABC-LSSVM and other identified techniques namely GA-LSSVM, PSO-LSSVM, *e*PSO-LSSVM, DE-LSSVM, CV-LSSVM, SVM and BPNN for Data Set B.

6.2.2.2.1 Crude Oil Prices

The performance results of seven prediction techniques in CL price prediction from Set B are summarized in Table 6.25. From the table, the MAPE produced by *e*ABC-LSSVM was 7.1272%, and followed by *e*PSO-LSSVM with 7.2691%. This makes *e*PSO-LSSVM falls to second place. The highest MAPE was recorded by BPNN with 13.3650%. That is, the difference between MAPE obtained by *e*ABC-LSSM and BPNN is 6.2378%. Meanwhile, similar MAPE was achieved by GA-LSSVM and CV-LSSVM, which is around 9.77%. Among all the hybrid LSSVM based techniques, PSO-LSSVM yielded the poorest MAPE, which is 14.4834%. Besides MAPE, the results obtained by all prediction techniques were also measured using sMAPE, RMSPE and Theil's U. Based on the results recorded, all error rates are consistent.

Table 6.26 shows the statistical comparison of the *e*ABC-LSSVM over the seven techniques, using the two-tailed T-test at 0.05% level of significance. Results from the table indicated the positive results for all pairs. As presented in previous cases, in this experiment, the graphical results for comparison of CL price prediction between actual and predicted prices is also provided in Figure 6.41 while the visual results for comparison of the convergence rate is shown in Figure 6.42. From the figure, it can be seen that *e*ABC-LSSVM is able to converge at iteration 39 while the rest techniques experienced with premature convergence.

Table 6.25: CL Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	602.5196	938.8506	1.0547	6.2916	**3.8604**	**7.1272**	**92.8728**	**7.3761**	**0.0887**	**0.0518**	89.8010	**92.2791**
GA-LSSVM	114.5410	52.0205	-	5.8390	5.0576	9.7798	90.2202	10.6527	0.1319	0.0811	88.9993	89.4431
PSO-LSSVM	534.078	1	-	1.6657	3.9242	14.4834	85.5166	16.7516	0.202	0.125	84.5061	85.4156
***e*PSO-LSSVM**	332.3567	885.4780	-	6.3307	6.6291	7.2691	92.7309	7.5281	0.0903	0.0527	87.9531	91.6712
CV-LSSVM	130.4836	55.5437	-	5.8391	5.9276	9.7730	90.2270	10.6452	0.1318	0.0811	**90.2270**	90.2270
DE-LSSVM	72.1311	1	-	2.7409	5.9945	12.9991	87.0009	14.8833	0.1851	0.1130	87.0009	87.0009
BPNN	-	-	-	**0.3766**	4.9282	13.3650	86.6350	15.4333	0.1926	0.1212	38.7733	71.4063
SVM	-	-	-	6.2924	6.6010	8.4394	91.5606	8.8290	0.1045	0.1183	84.0880	89.3960

Table 6.26: Significant Test for CL Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.000
*e*ABC-LSSVM – PSO-LSSVM	.000
*e*ABC-LSSVM – *e*PSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.000
*e*ABC-LSSVM – DE-LSSVM	.000
*e*ABC-LSSVM – BPNN	.000
*e*ABC-LSSVM – SVM	.000

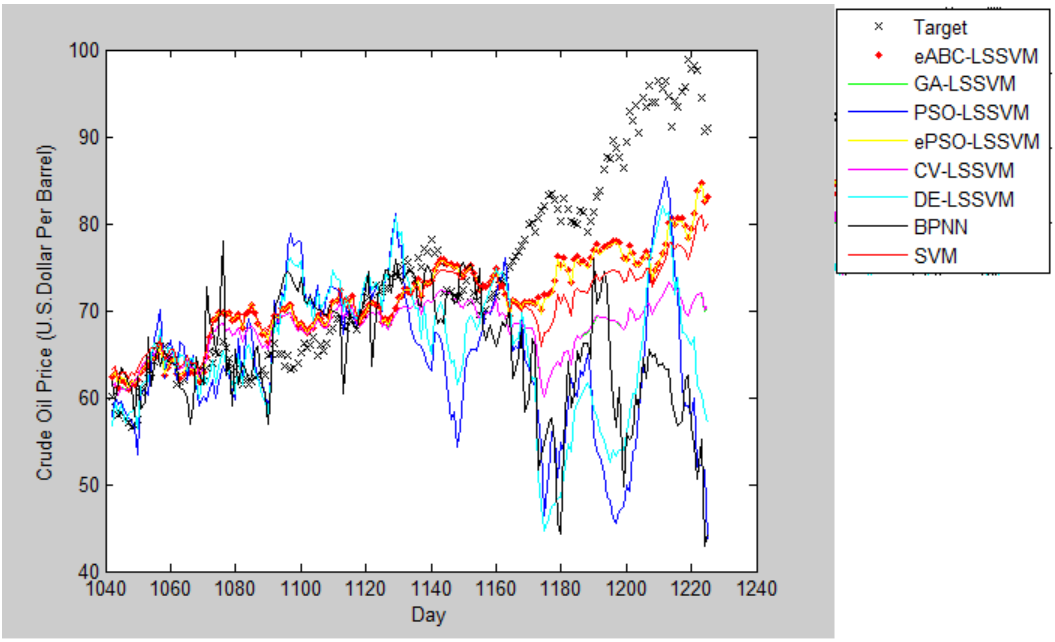


Figure 6.41: Comparison of CL Price Predictions Data Set B: eABC-LSSVM vs. Other Techniques

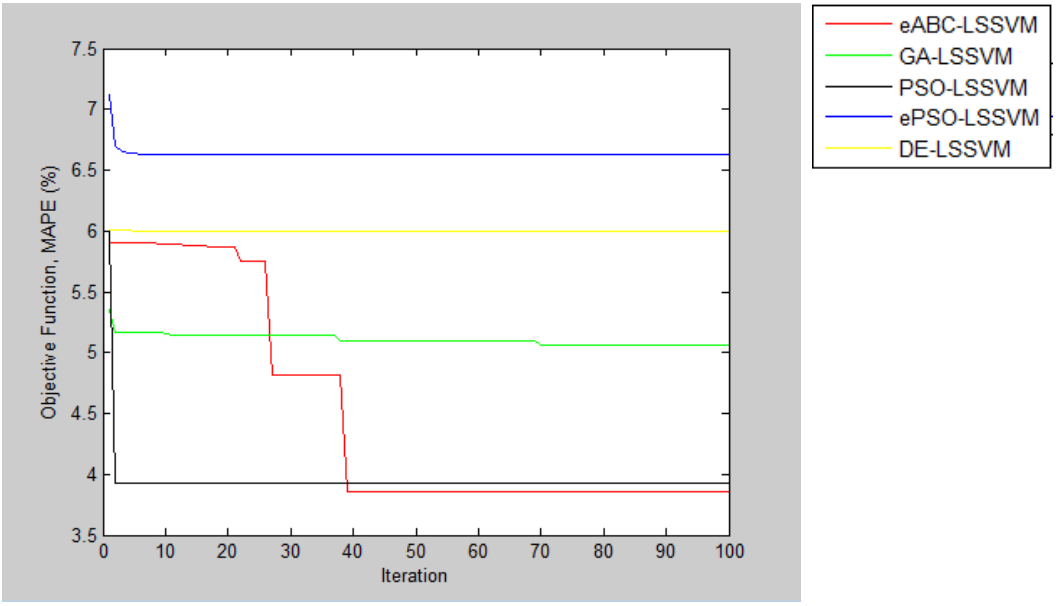


Figure 6.42: Comparison of Convergence for CL Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

6.2.2.2.2 Heating Oil Prices

In HO price prediction Set B (see Table 6.27), the minimum MAPE of *e*ABC-LSSVM in validation set, which is 5.3012% is obtained from the pair of hyper-parameters of interest, which is [260.8178, 768.8178]. The values of both parameters were consequently used in testing phase, which produced 8.4618% of MAPE. Thus, the highest prediction accuracy is achieved by the *e*ABC-LSSVM, which is 91.5382%. Similarly as previous cases, the BPNN still perform dismally when it is only capable to produce 74.4765% of prediction accuracy. This is lower from the results produced by the other techniques. The results for significant test between *e*ABC-LSSVM with other techniques are tabulated in Table 6.28, which shows the positive results for all cases.

The graphical results in Figure 6.43 clearly show the unsatisfying results yield by BPNN. Marked with a cyan line, the prediction values produced by BPNN are far from actual value, especially from day 130 to the last day of testing period. Meanwhile, the Figure 6.44 depicts the comparison of convergence among all prediction techniques. In the figure, it can be seen that *e*ABC-LSSVM is able to produce the lowest convergence value at iteration 24.

Table 6.27: HO Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	260.8178	768.3368	0.7813	9.1302	**5.3012**	**8.4618**	**91.5382**	**8.9332**	**0.0986**	**0.0526**	86.4657	**89.0430**
GA-LSSVM	46.8702	27.7520	-	8.2767	6.7470	14.0183	85.9817	15.4788	0.1615	0.0907	83.5563	85.0410
PSO-LSSVM	234.8068	1	-	3.15	5.9531	13.0899	86.9101	14.465	0.1585	0.0860	**86.9101**	86.9101
***e*PSO-LSSVM**	562.7499	519.832	-	8.9472	9.0128	8.8403	91.1597	9.3515	0.1019	0.0548	82.6457	88.5615
CV-LSSVM	28.3687	18.5478	-	8.1872	8.2836	14.4734	85.5266	16.0537	0.1674	0.0942	85.5266	85.5266
DE-LSSVM	234.8139	1	-	3.1500	5.9531	13.0899	86.9101	14.4650	0.1585	0.0860	**86.9101**	86.9101
BPNN	-	-	-	0.5964	7.4982	25.5235	74.4765	25.6494	0.4158	0.1986	34.5651	56.4254
SVM	-	-	-	9.0027	8.9610	8.8826	91.1174	9.3831	0.1003	0.0868	74.2200	86.9716

Table 6.28: Significant Test for HO Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.000
*e*ABC-LSSVM – PSO-LSSVM	.000
*e*ABC-LSSVM – *e*PSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.000
*e*ABC-LSSVM – DE-LSSVM	.000
*e*ABC-LSSVM – BPNN	.014
*e*ABC-LSSVM – SVM	.000

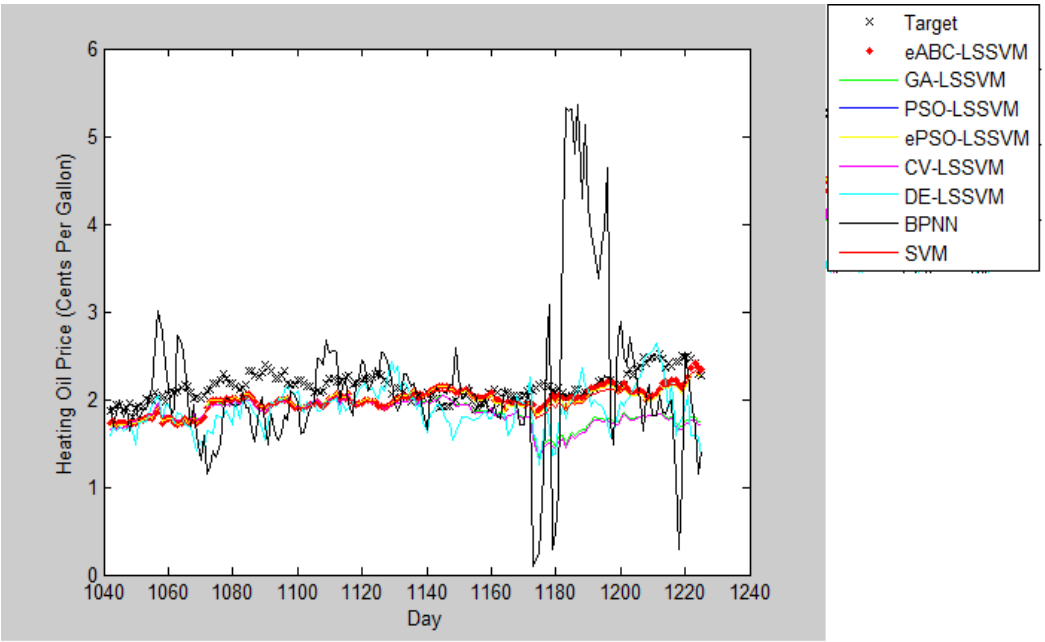


Figure 6.43: Comparison of HO Price Predictions Data Set B: eABC-LSSVM vs. Other Techniques

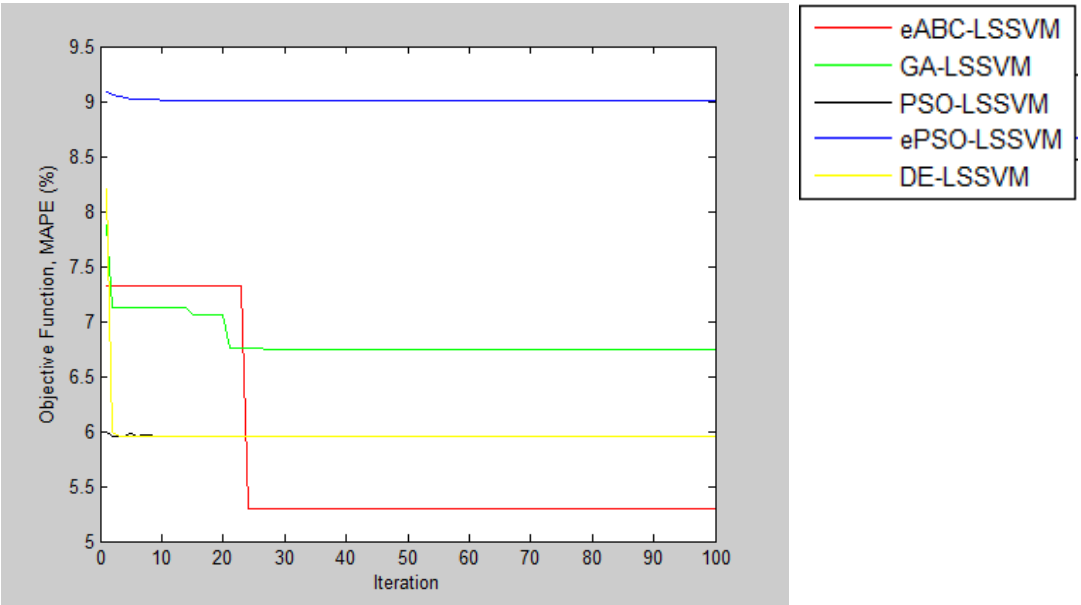


Figure 6.44: Comparison of Convergence HO Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

6.2.2.2.3 Gasoline Oil Prices

Based on the results of HU price prediction provided in Table 6.29, the lowest prediction error produced by *e*ABC-LSSVM which guided by MAPE, sMAPE, RMSPE and Theil's U. The smaller of the said metrics indicates a smaller discrepancy between the actual and predicted value. Hence, higher prediction accuracy is achieved by *e*ABC-LSSVM, which is 92.4006%. The worst prediction error was recorded by BPNN with 13.1276% of MAPE. For SVM and *e*PSO-LSSVM, similar MAPE were achieved by both techniques, which is 7.7%. However, SVM is capable to overcome *e*PSO-LSSVM with 0.0294% lower of MAPE. In Table 6.30, the T-test result indicated positive result obtained for *e*ABC-LSSVM in all pair of the comparison with other techniques, thus proving the significance of its superiority.

Figure 6.45 illustrates real and prediction values of different prediction techniques regarding HU price prediction while the comparison for convergence can be seen in Figure 6.46. The *e*ABC-LSSVM records 4.3421% of MAPE in validation set, thus is capable to escape from premature convergence at iteration 19.

Table 6.29: HU Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	145.3133	683.3321	1.0439	7.5074	**4.3421**	**7.5994**	**92.4006**	**8.1161**	**0.0983**	**0.0585**	**91.7003**	**92.1654**
GA-LSSVM	365.4990	99.4887	-	6.9053	5.8794	10.9045	89.0955	11.9770	0.1386	0.0831	88.3545	88.8192
PSO-LSSVM	248.4919	1	-	2.4604	4.4234	11.3246	88.6754	12.5812	0.1487	0.0876	88.6754	88.6754
***e*PSO-LSSVM**	334.133	545.5972	-	7.379	7.7263	7.7983	92.2017	8.3289	0.0996	0.0589	85.8498	90.8995
CV-LSSVM	208.0226	66.7834	-	6.8593	7.0761	11.1521	88.8479	12.2805	0.1419	0.0853	88.8479	88.8479
DE-LSSVM	248.4920	1	-	2.4604	4.4234	11.3246	88.6754	12.5812	0.1487	0.0876	88.6754	88.6754
BPNN	-	-	-	**0.4765**	7.6643	13.1276	86.8724	13.8622	0.1677	0.0866	30.8049	70.6541
SVM	-	-	-	7.6216	7.9386	7.7689	92.2311	8.3312	0.1024	0.0977	89.0153	90.2570

Table 6.30: Significant Test for HU Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.000
*e*ABC-LSSVM – PSO-LSSVM	.000
*e*ABC-LSSVM – *e*PSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.000
*e*ABC-LSSVM – DE-LSSVM	.000
*e*ABC-LSSVM – BPNN	.002
*e*ABC-LSSVM – SVM	.001

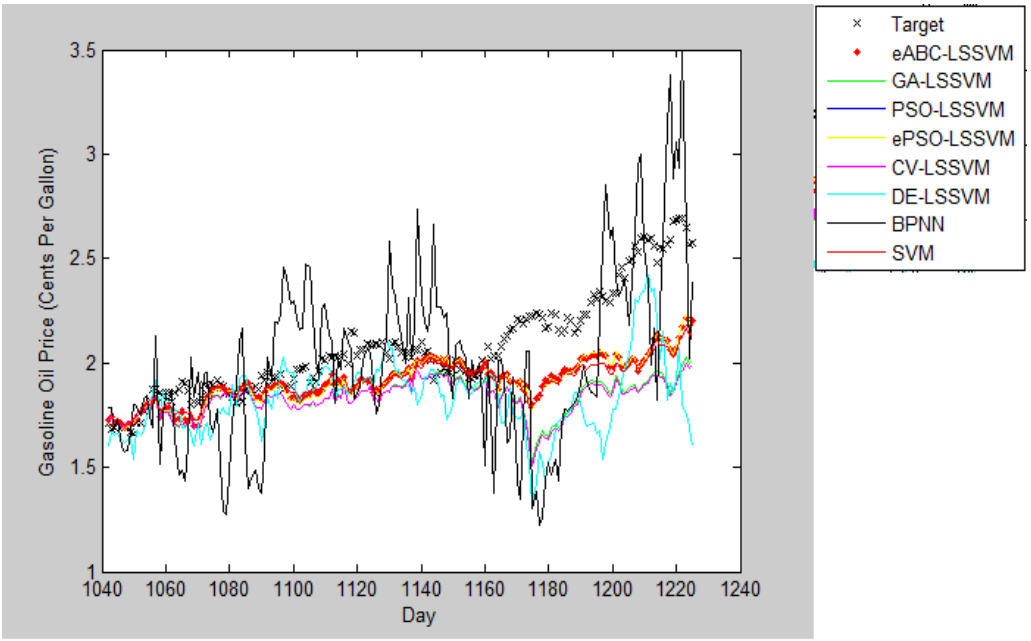


Figure 6.45: Comparison of HU Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

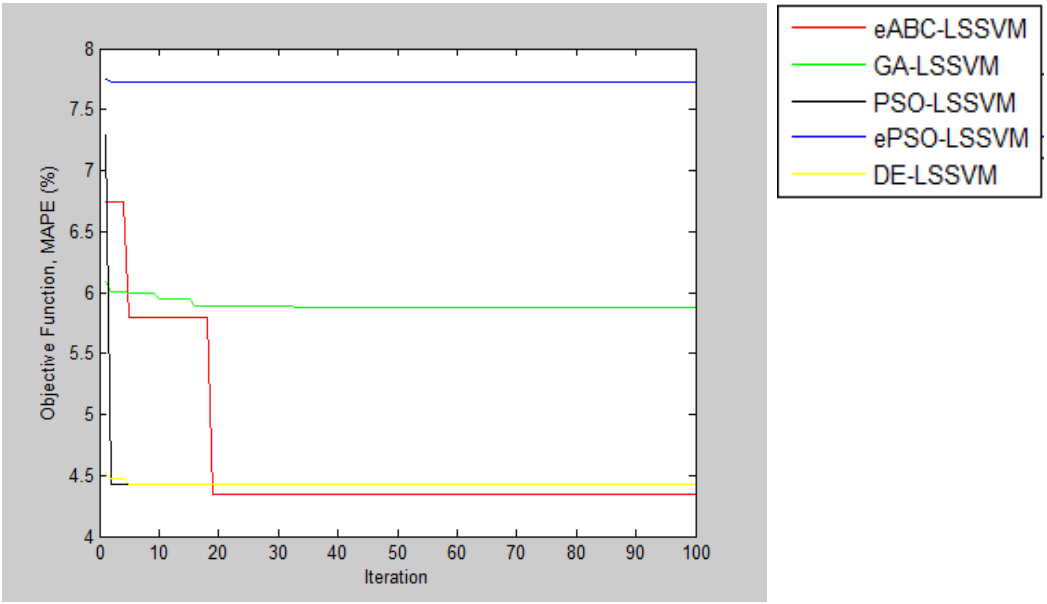


Figure 6.46: Comparison of Convergence for HU Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

6.2.2.2.4 Propane Prices

As shown in Table 6.31, the optimal values of γ and σ^2 for *e*ABC-LSSVM were 563.1288 and 673.5987 respectively. These yield 9.0805% of MAPE produced by the proposed technique. As compared to the other techniques, the MAPE recorded was the lowest, which interprets the highest prediction accuracy obtained that is 90.9195%. The second highest prediction accuracy was 90.7883%, which was achieved by *e*PSO-LSSVM while the other techniques are only capable to produce prediction accuracy below 90%. As shown in previous experiments, the significant test for this case is as tabulated in Table 6.32. From the table, it is evident that the significance level of the difference of the means between *e*ABC-LSSVM and other benchmarking techniques are significant.

Meanwhile, Figure 6.47 and 6.48 depicted the comparison between all identified techniques in PN price prediction from day 1042-1225 and the convergence rate in visual form respectively. From Figure 6.47, it is shown the obvious difference between actual and predicted value by BPNN which is represent with cyan line.

Table 6.31: PN Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	563.1288	673.5987	0.7700	6.8133	3.8136	9.0805	90.9195	9.7614	0.1117	0.0671	87.2835	89.9537
GA-LSSVM	373.091	67.5401	-	6.3775	5.0200	12.1718	87.8282	13.5721	0.1566	0.0958	87.5389	87.7090
PSO-LSSVM	499.6597	1	-	1.9206	3.9195	19.6808	80.3192	23.3614	0.2438	0.1533	79.4246	80.1999
***e*PSO-LSSVM**	475.1715	964.5081	-	6.8832	6.5748	9.2117	90.7883	9.895	0.1119	0.0673	82.1779	89.6831
CV-LSSVM	104.2350	33.5922	-	6.3688	6.2338	12.4217	87.5783	13.8490	0.1581	0.0969	87.5783	87.5783
DE-LSSVM	499.6595	1	-	1.9206	3.9195	19.6808	80.3192	23.3614	0.2438	0.3419	80.3192	80.3192
BPNN	-	-	-	0.3572	5.8864	15.1567	84.8433	17.8841	0.2108	0.1258	72.0764	73.8204
SVM	-	-	-	6.3948	6.0270	10.3407	89.6593	8.4496	0.1308	0.1147	78.8050	88.0686

Table 6.32: Significant Test for PN Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.000
*e*ABC-LSSVM – PSO-LSSVM	.000
*e*ABC-LSSVM – *e*PSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.000
*e*ABC-LSSVM – DE-LSSVM	.000
*e*ABC-LSSVM – BPNN	.002
*e*ABC-LSSVM – SVM	.000

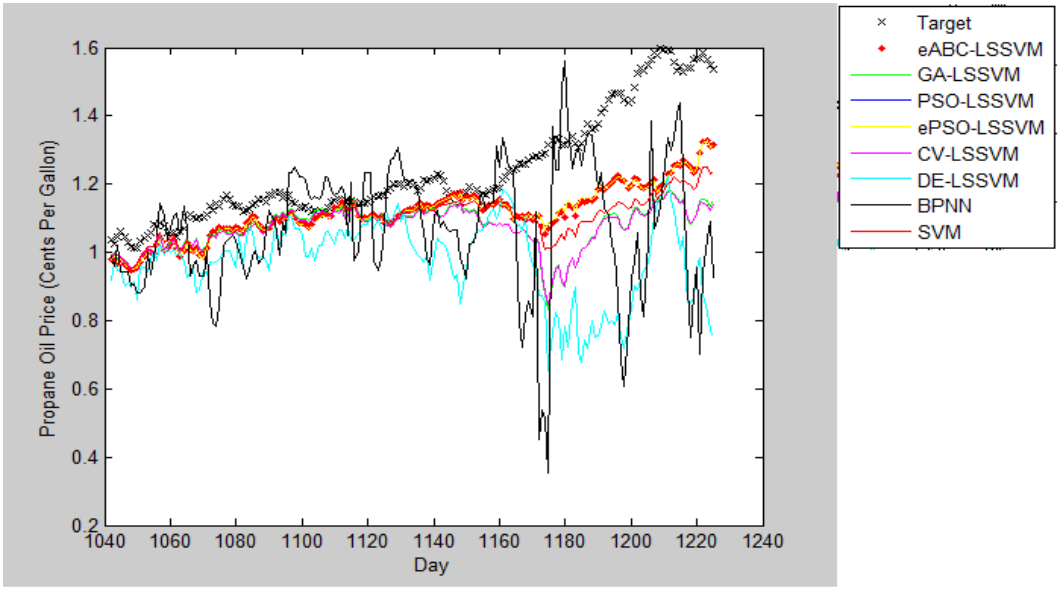


Figure 6.47: Comparison of PN Price Predictions Data Set B: eABC-LSSVM vs. Other Techniques

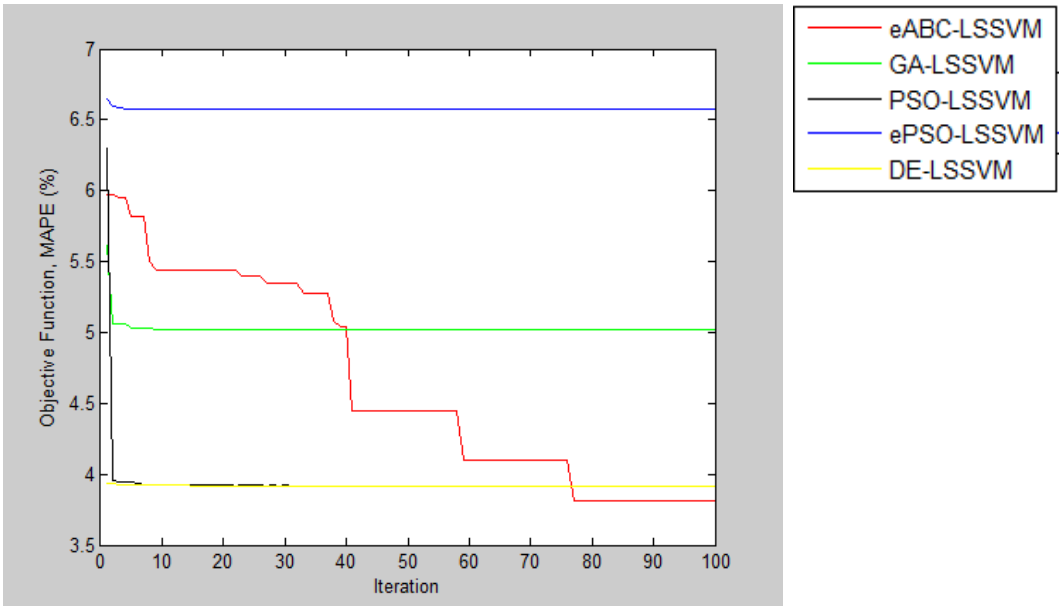


Figure 6.48: Comparison of Convergence for PN Price Prediction Data Set B: eABC-LSSVM vs. Other Techniques

6.2.2.2.5 Results Analysis for Data Set B: *e*ABC-LSSVM vs. Other Techniques

Based on the series of the experiments discussed previously, the efficiency of the proposed technique, viz. *e*ABC-LSSVM is proven when it is able to escape from local minimum in all cases. This is evident when optimal values of hyper parameter of interest are obtained, thus a good generalization performance is achieved. In addition, the *e*ABC-LSSVM is also capable to escape from premature convergence as occurred in all identified hybrid LSSVM techniques. Unlike in Set A, in this experiment, BPNN shared position with PSO-LSSVM as the poorest predictor where BPNN performed worst in HO and HU (see section 6.2.2.2.2 and 6.2.2.2.3 respectively) while the PSO-LSSVM in the CL and PN price prediction (see section 6.2.2.2.1 and 6.2.2.2.4 respectively).

It also can be seen that, in PN price prediction, the PSO-LSSVM also shared the rank with DE-LSSVM which produced similar result. The obvious over fitting issue that is not addressed by those techniques contribute to poor prediction accuracy produced. It can be seen that in the cases discussed, a good outcome was produced by those techniques in training set but when the new data are introduced, the prediction error becomes unpredictably larger. This consequently limited its generalization capability. For PSO-LSSVM and DE-LSSVM, this situation is expected when the minimum boundary value is set to σ^2. As analyzed previously (see section 6.2.1.1.5 and 6.2.1.2.5), over fitting is likely to occur when too small value is taken by σ^2.

6.3 *e*ABC-LSSVM on Metal

This section discusses the empirical results on gold price prediction.

6.3.1 Data Set C

This experiment is to predict the daily spot price one month into the future (21 trading days) of GC (see Chapter 3, section 3.5.1.1).

6.3.1.1 *e*ABC-LSSVM vs ABC-LSSVM and Its Variants

In this section, the empirical results between the *e*ABC-LSSVM and variants of ABC-LSSVM are presented.

6.3.1.1.1 Gold

The results from this experiment are as tabulated in Table 6.33. By producing lowest error rate relative to MAPE, sMAPE, RMSPE and Theil's U, which is 3.4934%, 3.5926%, 0.0462% and 0.0243 respectively, the *e*ABC-LSSVM is ranked first. It is closely followed by *cm*ABC-LSSVM which recorded 3.7608% of MAPE while the other metrics seems to be consistent. Meanwhile, the difference of MAPE between standard ABC-LSSVM and *e*ABC-LSSVM is 1.009%. This makes ABC-LSSVM ranked fourth, after *lv*ABC-LSSVM which recorded 3.8378% of MAPE. The significance of each improvement is shown in Table 6.34. In Figure 6.49, the figure illustrates the comparison between actual and predicted prices of each prediction techniques for the last 142 days. Meanwhile, Figure 6.50 illustrates the whole optimization process by all experimented prediction techniques, which is interpreted by MAPE.

As shown in energy fuel price prediction for both Data Set A and B, for Data Set C, the progress in exploitation process for *e*ABC-LSSVM and ABC-LSSVM are also given, which can be visualized in Figure 6.51 and 6.52 respectively. With the induction of Levy mutation, progressive searching behavior is presented in *e*ABC-LSSVM which assists the model to escape from local minimum. In addition, the *e*ABC-LSSVM is also capable to avoid from over fitting problem which is benefitted from conventional mutation. Unfortunately, the opposite situation is encountered in ABC-LSSVM.

Table 6.33: Data Set C Price Prediction: eABC-LSSVM vs. ABC-LSSVM and Its Variants

	***e*ABC-LSSVM**	***lv*ABC-LSSVM**	***cm*ABC-LSSVM**	**ABC-LSSVM**
γ	1.3774	459.4879	897.5449	462.6987
σ^2	1.1313	180.1307	11.3653	1
α	1.4806	1.4001	-	-
MAPE Training (%)	2.8251	3.4356	2.3208	**0.7867**
MAPE Validation (%)	1.9939	1.9696	1.9928	**1.9691**
MAPE Testing (%)	**3.4934**	3.8378	3.7608	4.5024
PA (%)	**96.5066**	96.1622	96.2393	95.4976
sMAPE (%)	**3.5926**	3.9569	3.8440	4.5366
RMSPE (%)	**0.0462**	0.0500	0.0473	0.0549
Theil's U	**0.0243**	0.0262	0.0246	0.0281
Minimum PA (%)	**96.2546**	95.4789	95.3979	95.4757
Average PA (%)	**96.4201**	95.9168	95.6737	95.4802

Table 6.34: Significant Test for Data Set C Price Prediction: eABC-LSSVM vs. ABC-LSSVM and Its Variants

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – *lv*ABC-LSSVM	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	.000
*e*ABC-LSSVM – ABC-LSSVM	.000

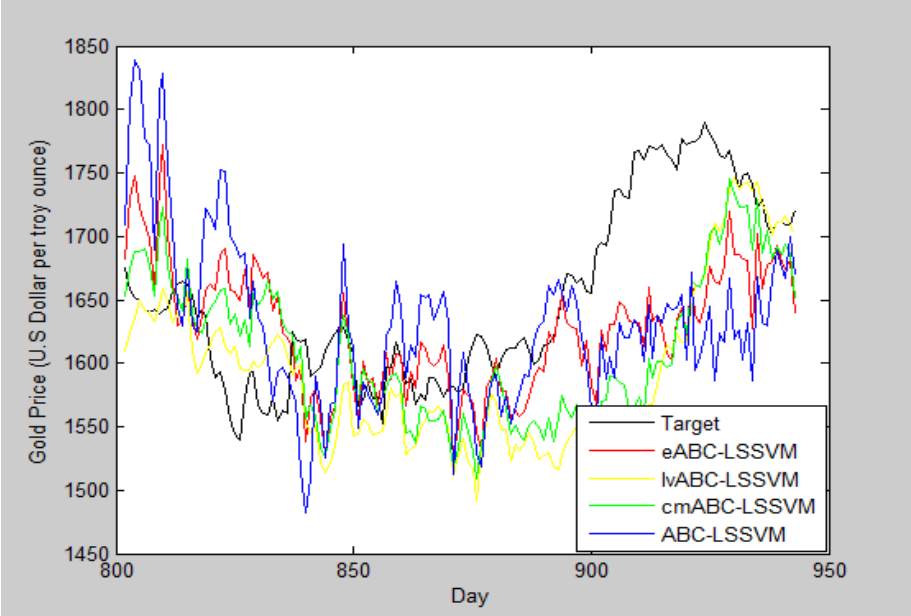


Figure 6.49: Comparison of Data Set C price prediction: eABC-LSSVM vs. ABC-LSSVM and Its Variants

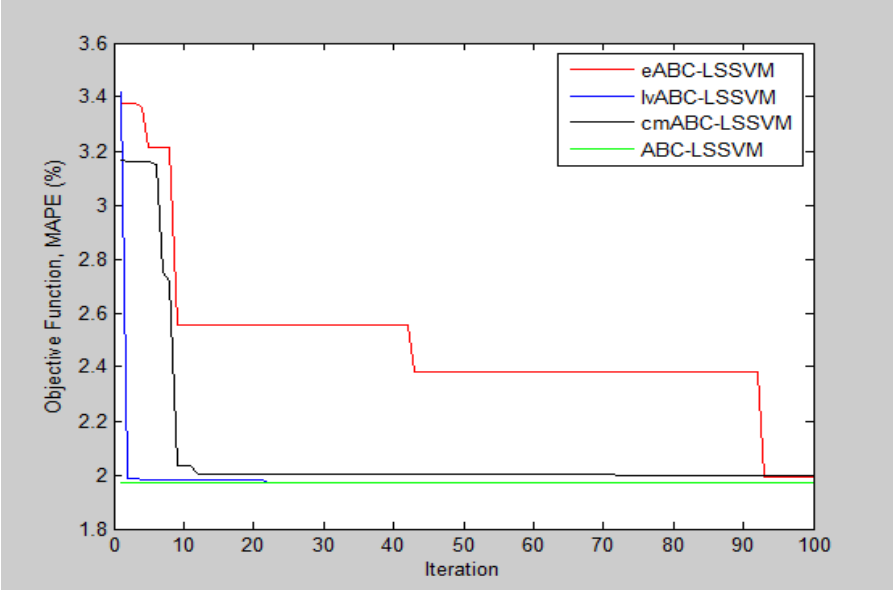
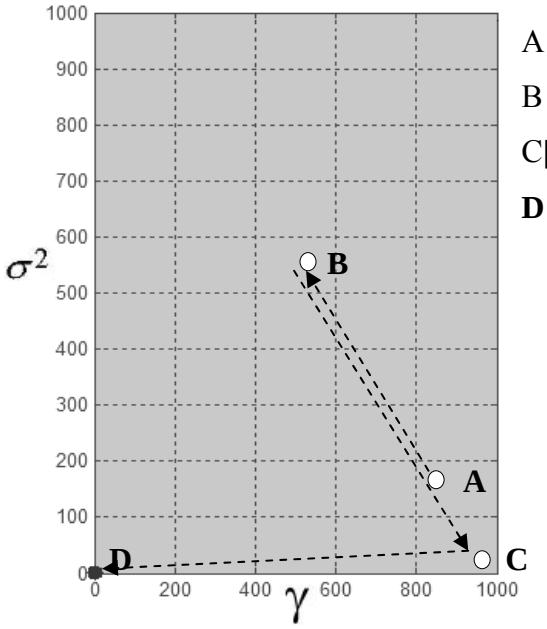


Figure 6.50: Comparison of convergence for Data Set C Price Prediction: eABC-LSSVM vs. ABC-LSSVM and Its Variants



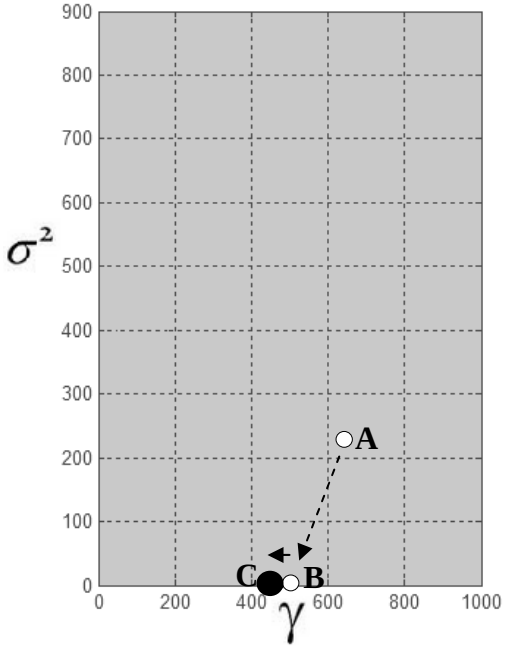
A [845.681, 179.33]

B [52.082, 552.323]

C[989.288, 4.5599]

D [1.3774, 1.1313]***Optimal γ and σ^2**

Figure 6.51: Exploitation of Search Space by eABC-LSSVM in Data Set C Price Prediction



A [605.208, 1]

B [482.811, 1]

C [462.699, 1]*Optimal γ and σ^2

Figure 6.52: Exploitation of search space by ABC-LSSVM in Data Set CPrice Prediction

6.3.1.1.2 Results and Analysis for Data Set C: *e*ABC-LSSVM vs ABC-LSSVM and Its Variants

Similar as previous cases, over fitting problem arise in ABC-LSSVM when the σ^2 takes a small value, this is 1. This situation can be seen when ABC-LSSVM is capable to outperform in training phase but presents contradictory performance in testing phase (see Table 6.33). It is worth noting that the training phase is for model fitting purposes while the real evaluation on generalization performance is based on testing phase (Williams, 2011). However, the situation performed in ABC-LSSVM is incapable to prove its superiority in generalization performance, as offered in *e*ABC-LSSVM.

6.3.1.2 *e*ABC-LSSVM vs Other Techniques

In this section, the results obtained by *e*ABC-LSSVM in Data Set C is compared against other identified techniques, namely GA-LSSVM, PSO-LSSVM, *e*PSO-LSSVM, DE-LSSVM, CV-LSSVM, SVM and BPNN.

6.3.1.2.1 Gold

For GC price prediction (see Table 6.35), the best prediction performance of the *e*ABC-LSSVM is obtained at γ= 1.3774 and σ^2= 1.1313, which is 3.4934% of MAPE. With that, the prediction accuracy achieved was 96.5066%, which is the highest among all the experimented techniques. The difference between average of prediction accuracy in 30 run times and lowest prediction accuracy are not much different, which indicates the consistency of the proposed technique for this experiment. CV-LSSVM and GA-LSSVM ranked second and third when the prediction accuracy obtained were 96.3863% and 96.3094% respectively.

Similarly reported in previous cases, the worst prediction performance was yielded by BPNN with 5.8138% of MAPE. The result for significant test is provided in Table 6.36. The graphical result which visualizes the comparison among prediction techniques in GC price prediction and convergence are depicted in Figure 6.53 and 6.54 respectively. From Figure 6.53, good agreements between predicted and experimental values are visualized by *e*ABC-LSSVM which is represented by red dotted mark.

Table 6.35:Data Set C Price Prediction: eABC-LSSVM vs. Other Techniques

	γ	σ^2	α	MAPE Train. (%)	MAPE Val. (%)	MAPE Test. (%)	PA (%)	sMAPE (%)	RMSPE (%)	Theil's U	Minimum PA (%)	Average PA (%)
***e*ABC-LSSVM**	1.3774	1.1313	1.4806	2.8251	1.9939	**3.4934**	**96.5066**	**3.5926**	**0.0462**	**0.0243**	96.2546	**96.4201**
GA-LSSVM	145.5840	43.1870	-	3.2668	2.4365	3.6906	96.3094	3.7954	0.0480	0.0251	96.2302	96.2731
PSO-LSSVM	478.8643	1	-	0.7779	**1.9691**	4.5205	95.4795	4.5551	0.0551	0.0283	95.0788	95.4261
***e*PSO-LSSVM**	916.0726	274.5093	-	3.4424	3.4414	3.7814	96.2186	3.8966	0.0493	0.0259	83.4303	94.7440
CV-LSSVM	174.8014	36.0605	-	3.1891	3.1849	3.6137	96.3863	3.7115	0.0471	0.0246	**96.3863**	96.3863
DE-LSSVM	478.8644	1	-	0.7779	**1.9691**	4.5205	95.4795	4.5551	0.0551	0.0283	95.4795	95.4795
BPNN	-	-	-	**0.0800**	4.4031	5.8138	94.1862	5.7938	0.0712	0.0351	27.6259	86.2470
SVM	-	-	-	3.1525	3.1603	4.1406	95.8594	4.0170	0.0493	0.0508	82.2340	88.6548

Table 6.36: Significant Test for Data Set C Price Prediction: eABC-LSSVM vs. Other Techniques

Methods	**Sig. (2-tailed)**
*e*ABC-LSSVM – GA-LSSVM	.040
*e*ABC-LSSVM – PSO-LSSVM	.000
*e*ABC-LSSVM –*e*PSO-LSSVM	.000
*e*ABC-LSSVM – CV-LSSVM	.971
*e*ABC-LSSVM – DE-LSSVM	.000
*e*ABC-LSSVM – BPNN	.000
*e*ABC-LSSVM – SVM	.000

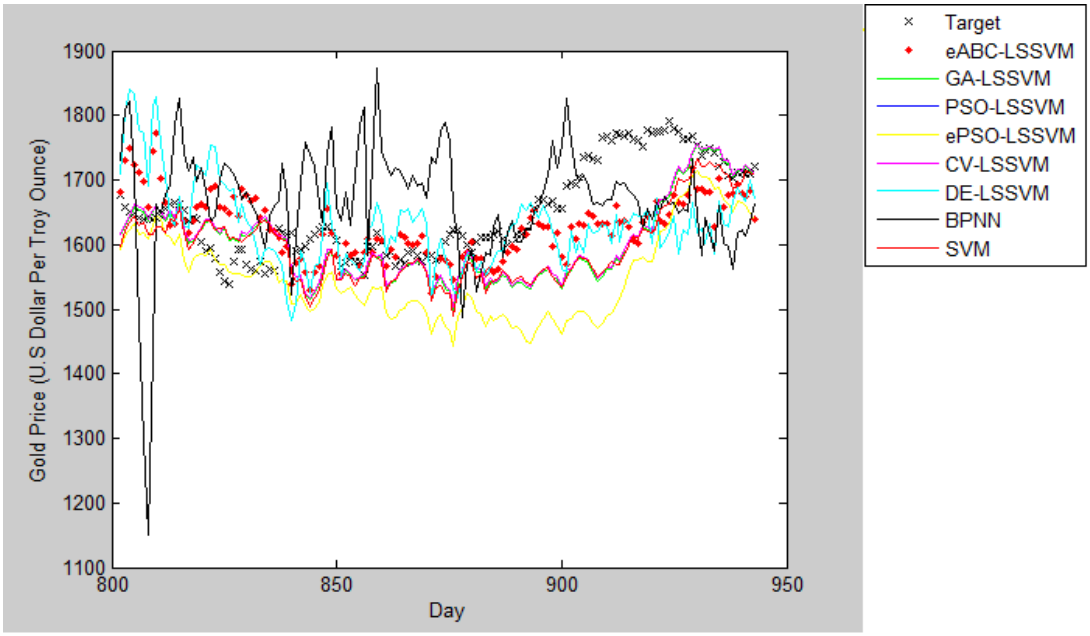


Figure 6.53: Comparison of Data Set C price predictions: eABC-LSSVM vs. Other Techniques

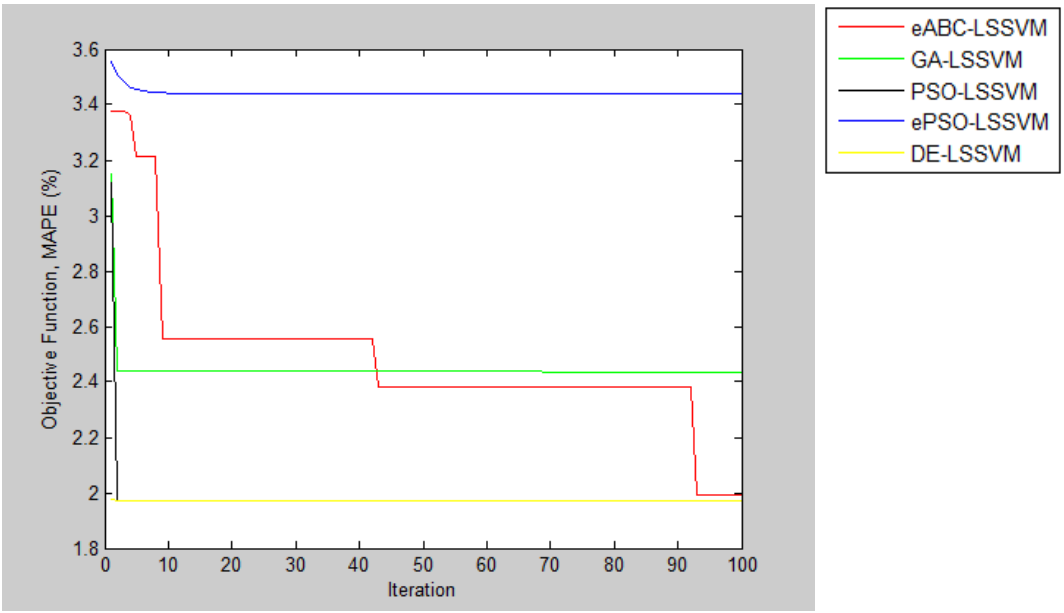


Figure 6.54: Comparison of Convergence for Data Set C Price Prediction: eABC-LSSVM vs. Other Techniques

6.3.1.2.2 Results Analysis for Data Set C: *e*ABC-LSSVM vs. Other Techniques

In GC price prediction, the comparison between *e*ABC-LSSVM with seven others experimented techniques are in favor of the *e*ABC-LSSVM. Among the hybrid LSSVM based technique, PSO-LSSVM and DE-LSSVM shared the same reputation as the poorest prediction technique in GC price prediction while BPNN was marked as the worst among all the competitors chosen. This is due to over fitting. This is portrayed by smaller MAPE produced during training, especially the one produced by BPNN. Unfortunately, the good performance recorded is incapable to be sustained in testing set, which played a main role in evaluating the generalization performance of prediction technique. It is also proven that, when the prediction techniques are capable to escape from falling to the boundary value, the over fitting problem is unlikely to occur. This situation is demonstrated by all hybrid LSSVM based prediction techniques except PSO-LSSVM and DE-LSSVM.

6.4 Summary

This chapter discussed the results obtained upon completing a series of experiment on energy fuels and metal price. The results obtained were judged based on five criteria, namely MAPE, prediction accuracy, sMAPE, RMSPE and Theil's U and compared against state of the art algorithms namely ABC-LSSVM, GA-LSSVM, PSO-LSSVM, *e*PSO-LSSVM, DE-LSSVM, CV-LSSVM, SVM and BPNN. Based on the conducted experiments, the empirical results proven that the *e*ABC-LSSVM posses a capability to fully exploit the search space and perform large step length which finally increased the chance of the algorithm to escape from local minimum.

Besides, the *e*ABC-LSSVM is also able to prevent from over fitting and under fitting problem, thus better results are achieved. This is indicated by the *e*ABC-LSSVM posses a capability to avoid the hyper parameters from falling to the boundary values in all cases. In addition, the empirical results also indicated that it is an advantage when the global optimal is sufficiently away from the current solution. Thus, it is not only facilitates the technique to converge in acceptable manner, but also contributes in avoiding the technique from falling to local minimum (Wang & Li, 2008). This situations are clearly visualized in the exploitation of the search space performed by *e*ABC-LSSVM. Consequently, Findings of the study indicated that the proposed *e*ABC-LSSVM does not only possess higher generalization performance, but also possess the advantage in escaping the premature convergence.

In the next chapter, conclusion of the study and contributions are presented. Besides, some recommendation for future works are also suggested.

CHAPTER SEVEN

CONCLUSIONS AND RECOMMENDATION FOR FUTURE WORKS

7.1 Conclusion

Prediction of non-renewable natural resources has experienced major changes for past decades. Starting from conventional statistical techniques to CI approach, this issue has never failed to attract both academic and practitioners community. As highlighted in Chapter 1, this is due to the prices of such data sets played a crucial role not only to the world economy but also to the human individually. With respect to that matter, the avalanche of studies regarding non-renewable natural resources price prediction can be seen encouraging and still ongoing. This situation has been presented in Chapter 2. From the literature review that has been conducted, the gaps that exist in the existing prediction technique have led to the proposed *e*ABC-LSSVM. This is essential as the prediction accuracy of the said time series data is vital.

By using real data set of energy fuel prices (Data Set A and B) and one data set of metal price (Data Set C), the ABC and LSSVM are hybridized for the purpose of optimizing the LSSVM hyper-parameters. This is to achieve the objective 1 (see Chapter 1, section 1.5). As to address the highlighted problem (see Chapter 1, section 1.3), enhancements that are introduced to standard ABC-LSSVM are proven to be beneficial to the problem under study. Firstly, by employing *lv*ABC in EB and

OB phases, the searching behavior is enriched (see Chapter 3, section 3.3.2). The implementation of Levy mutation which offers large variation of mutation in the search space give an advantage in terms of providing more distinct values in seeking for the optimal solution. Meanwhile, the introduction of different strategies in both phases produced a good balance in exploitation process which assisted in obtaining good solution and avoiding local minimum. Thus, the objective to solve the local minimum problem, which is defined in objective 2 (see Chapter 1, section 1.5) is achieved.

Secondly, the *cm*ABC that was introduced in decision making process facilitates the model from facing with over fitting and under fitting problem (see Chapter 3, section 3.3.3). A small and large value of hyper-parameters of interest may lead to poor generalization performance since upper and lower level of the boundary does not promise the optimal value (Fu, et al., 2010, Lendasse, et al., 2005; Wu, et al., 2007). With such an enhancement, the objective to address the problem of over fitting and under fitting, which is stated in objective 2 (see Chapter 1, section 1.5) is achieved.

Thirdly, empirical results proved that the combined implementation of different types of mutation at different stages of the algorithm led to better predictive power to *e*ABC-LSSVM. This can be seen when *lv*ABC-LSSVM is unable to outperform ABC-LSSVM in CL price prediction Data Set A (see Chapter 6, section 6.2.1.1) while similar situation is also experienced by *cm*ABC-LSSVM in HO price prediction Data Set A (see Chapter 6, section 6.2.1.1). Interestingly, when both

*lv*ABC-LSSVM and *cm*ABC-LSSVM are integrated, the results obtained for both price predictions are much better and the problem of premature convergence could be avoided. Thus, the objective 4 is achieved.

7.2 Contribution

The contributions of this study are as follow:

7.2.1 Knowledge Contribution

In this study, two major issues have been addressed. First is the limitation of searching behavior of the artificial bees in the search space is enriched where this would directly impact the global search ability. By inducing Levy mutation and different strategies, the performance of standard ABC is enhanced. This contributes the algorithm in avoiding local minimum, hence global minimum is achieved.

Secondly, the algorithm is also enhanced by using conventional mutation which assist the algorithm from falling to boundary values which consequently would prevent the LSSVM from facing with data fitting problem. By using this approach, the algorithm is induced to explore another region in order to search for global minimum rather than local minimum.

With the combination of these features, the *e*ABC-LSSVM is introduced which is not only contributes to better generalization but also capability to escape from premature convergence.

7.2.2 Practical Contribution

Based on the obtained results, the *e*ABC-LSSVM have been proven capable to predict the change of trend which finally contribute to give high accuracy in price prediction. By following the signal given by *e*ABC-LSSVM, it will be beneficial for the commodities traders and investors in reducing risk projecting their profit.

The main contributions of this study can be summarized as follows:

i. The problem on optimization of LSSVM hyper-parameter in prediction task is addressed by using ABC-LSSVM (see Chapter 3, section 3.2.2).By using ABC-LSSVM, the hyper-parameters are automatically optimized. This is defined as objective 1 (see Chapter 1, section 1.5).

ii. By using the *lv*ABC-LSSVM (see Chapter 3, section 3.3.2), the tendency of ABC-LSSVM to fall into local minimum is addressed. This is achieved by introducing Levy mutation and different strategies for exploitation process, which involves EB and OB phases. The features of Levy mutation that exist in *lv*ABC-LSSVM is capable to assist the model to produce variation in mutation while different strategies applied bring out a good balance in the exploitation process. This is defined as objective 2 (see Chapter 1, section 1.5).

iii. The tendency of ABC-LSSVM to suffer from over fitting and under fitting problem is addressed by making use of *cm*ABC (see Chapter 3, section

3.3.3).By using *cm*ABC, the algorithm is encouraged to discover for a global optimum. This is defined as objective 3 (see Chapter 1, section 1.5).

iv. The combination of (ii) and (iii) in standard ABC-LSSVM does not only give the *e*ABC-LSSVM better predictive power but also advantage to escape from premature convergence. This is defined as objective 4 (see Chapter 1, section 1.5).

7.3 Recommendation for Future Works

Based on the results and discussion presented in Chapter 6, the proposed *e*ABC-LSSVM is proven to be more superior relative to the metrics utilized as compared to the other identified prediction techniques. This indicates that the *e*ABC-LSSVM possess significant implication to the problem of interest. However, there is always possible potential applications that can be explored in the future and improve its limitation.

Firstly, agreeing with Gao, et al., (2012), where based on the series of reports and discussion presented earlier (see Chapter 6), to achieve both speed in convergence and avoiding local minimum are obviously a challenging task. However, it is interesting to be considered for future works as to improve the efficiency of the algorithm.

Secondly, in this study, the experiments that have been conducted involved energy fuels prices and gold prices, where both types of data sets are categorized as non-

renewable natural resources. Thus, it would be interesting to test the efficiency and applicability of *e*ABC-LSSVM on renewable commodities data sets such as cereal, currencies and others.

Thirdly, as to enable the learning algorithm to operate effectively, feature selection task is also considered for future works. With the advantage of ABC algorithm, the feature selection based on ABC can be embedded in the learning algorithm which functions to remove the irrelevant inputs for the sake of higher prediction accuracy.

Fourthly, to reduce the burden of a single computer in running the algorithm, grid processing approach for ABC algorithm can be introduced. By using this approach, the jobs between the agents in ABC algorithm are distributed to different computer resource. This may lead to faster and efficient computation.

Lastly, as the ABC algorithm allows one parameter to be changed in each iteration, it would be interesting to explore the possibility of changing more than one parameters at each iteration which may contribute in producing faster and better solution.

.

REFERENCES

Abdullah, S. N., & Zeng, X. (2010, July 18-23). *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 (IJCNN), Barcelona, Spain.

Afshin, M. (2007). *Application of Least Squares Support Vector Machines in Medium-term Load Forecasting* (Master's Thesis). Retrieved from ProQuest Dissertations & Theses Database (UMI No. 304701942).

Afshin, M., Sadeghian, A., & Raahemifar, K. (2007, June 24-28). *On Efficient Tuning of LS-SVM Hyper-Parameters in Short-Term Load Forecasting: A Comparative Study.* Proceedings of the IEEE Power Engineering Society General Meeting (PES), Tampa, Florida, USA.

Ahmad, A. S., Hassan, M. Y., & Majid, M. S. (2012, December 2-5). *Application of Hybrid GMDH and Least Squares Support Vector Machines in Energy Consumption Forecasting.* Proceedings of the IEEE International Conference on Power and Energy (PECon), Kota Kinabalu, Sabah, Malaysia.

Ahmed Farid, J., & Salahudin, F. (2010). *Risk Frameworks and Applications* (2nd ed.). Karachi: Alchemy Technologies Pvt. Ltd.

Aiken, M. (1999). Using Neural Network to Forecast Inflation. *Industrial Management and Data System, 99*(7), 296-301.

Akram, Q. F. (2009). Commodity Prices, Interest Rates and the Dollar. *Energy Economics, 31*, 838-851.

Al-Shalabi, L., Shaaban, Z., & Kasasbeh, B. (2006). Data Mining: A Preprocessing Engine. *Journal of Computer Science, 2*(9), 735-739.

Alizadeh, A., Moghaddam, M., Khakzad, M., & Ebrahimipour, V. (2012). A Flexible Neural Network-Fuzzy Mathematical Programming Algorithm for Improvement of Oil Price Estimation and Forecasting. *Computers and Industrial Engineering, 62*, 421-430.

Armstrong, J. S. (2001). *A Handbook for Researchers and Practitioners*. New York: Springer.

Babayigit, B., & Ozdemir, R. (2012, July 1-4). *A Modified Artificial Bee Colony Algorithm for Numerical Function Optimization.* Proceedings of the IEEE Symposium on Computers and Communications (ISCC), Cappadocia, Turkey.

Bao, Y., Yang, Y., Xiong, T., & Zhang, J. (2011, April 15-19). *A Comparative Study of Multi-step-ahead Prediction for Crude Oil Price with Support Vector Regression.* Proceedings of the Fourth International Joint Conference on Computational Sciences and Optimization (CSO), Lijiang City, China.

Bao, Y., Zhang, X., Yu, L., Lai, K. K., & Wang, S. (2011). An Integrated Model Using Wavelet Decomposition and Least Squares Support Vector Machines for Monthly Crude Oil Prices Forecasting. *New Mathematics and Natural Computation, 7*(2), 299-311.

Barchart. (2011). *Historical Quote*. Retrieved December 25, 2011, from http://www.barchart.com/historicalquote.php?

Belaire, J., & Contreras, D. (2002). How to compute the BDS test: A Software Comparison. *Journal of Applied Econometrics, 17*(6), 691-699.

Bessedik, S. A., & Hadi, H. (2013). Prediction of Flashover Voltage of Insulators using Least Squares Support Vector Machine with Particle Swarm Optimization. *Electric Power Systems Research, 104*, 87-92.

Bhar, R., & Hammoudeh, S. (2011). Commodities and Financial Variables: Analyzing Relationships in a Changing Regime Environment. *International Review of Economics and Finance, 20*, 469-484.

Board of Governors of the Federal Reserve System. (2012). *Selected Interest Rates (Daily) - H.15.* Retrieved February 17, 2012, from http://www.federalreserve.gov/releases/h15/data.htm

Bolaji, A. L. A., Khader, A. T., Al-Betar, M. A., & A. Awadallah, M. A (2013). Artificial Bee Colony Algorithm, its Variants and Applications: A Survey. *Journal of Theoritical and Applied Information Technology, 47*(2), 434-459.

Bonabeau, E., & Meyer, C. (2001, May). Swarm Intelligence: A Whole New Way To Think About Business. *Harvard Business Review, 106-114.*

Brajevic, I., Tuba, M., & Subotic, M. (2010, August 20-22). *Improved Artificial Bee Colony Algorithm for Constrained Problems.* Proceedings of the 11th WSEAS International Conference on Neural Networks, Taipei, Taiwan.

Brock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A Test for Independence on the Correlation Dimension. *Econometrics Reviews, 15*(3), 197-235.

Cao, D. Z., Pang, S. L., & Bai, Y. H. (2005, August 18-21). *Forecasting Exchange Rate Using Support Vector Machines.* Proceedings of the Fourth International Conference on Machine Learning and Cybernetics (ICMLC), Guangzhou, China.

Chen, Q., Wu, Y., & Chen, X. (2008, August 3-5). *Research on Customers Demand Forecasting for E-business Web Site Based on LS-SVM.* Proceedings of the International Symposium on Electronic Commerce and Security (ISECS), Guangzhou, China.

Chen, X., Wang, J., Sun, D., & Liang, J. (2008, October 18-20). *Time Series Forecasting Based on Novel Support Vector Machine Using Artificial Fish Swarm Algorithm.* Proceedings of the Fourth International Conference on Natural Computation (ICNC), Jinan, China.

Cheng, G., Guo, R., & Wu, W. (2010, January 22-24). *Petroleum Lithology Discrimination Based on PSO-LSSVM Classification Model.* Proceedings of the Second International Conference on Computer Modeling and Simulation (ICCMS), Sanya, China.

Cheng, J., Qian, J. S., & Guo, Y. N. (2006). Least Squares Support Vector Machines for Gas Concentration Forecasting in Coal Mine. *International Journal of Computer Science and Network Security, 6*(6), 125-129.

Colman, A. M., & Pulford, B. D. (2008). *A Crash Course in SPSS for Windows: Updated for Versions 14, 15 and 16* (4th ed.). West Sussex, United Kingdom: Wiley Blackwell.

Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. United Kingdom: Cambridge University Press.

Dooley, G., & Lenihan, H. (2005). An Assesment of Time Series Methods in Metal Price Forecasting. *Resources Policy, 30*, 208-217.

Dorigo, M., & Stutzle, T. (2004). *Ant Colony Optimization*. Cambridge: MIT Press

Dos Santos, G. S., Luvizotto, L. G. J., Mariani, V. C., & Dos Santos Coelho, L. (2012). Least Squares Support Vector Machines with Tuning based on Differential Evolution Approach Applied to the Identification of a Thermal Process. *Expert Systems with Applications, 39*, 4805-4812.

Dunsby, A., Eckstein, J., Gaspar, J., & Mulholland, S. (2008). *Commodity Investing: Maximizing Returns Through Fundamental Analysis*. New Jersey: John Wiley & Sons.

El-Abd, M. (2012). Performance Assessment of Foraging Algorithms vs. Evolutionary Algorithms. *Information Sciences, 182*, 243-263.

Elish, K. O., & Elish, M. O. (2008). Predicting defect-prone software modules using Support Vector Machines. *The Journal of Systems and Software, 81*, 649-660.

Fang, X. Y., & Bai, T. (2009). *Share Price Prediction using Wavelet Transform and Ant Colony Optimization for Parameters Optimization in SVM.* Proceedings of the Global Congress on Intelligent System (GCIS), Xiamen, China.

Feoktistov, V., & Janaqi, S. (2004, January 8-9). *Hybridization of Differential Evolution with Least-Square Support Vector Machines.* Proceedings of the Fifteenth Dutch-Belgian Conference on Machine Learning (BENELEARN), Brussels, Belgium.

Fontanills, G. A. (2007). *Getting Started in Commodities*. New Jersey: John Wiley & Sons.

Frush, S. (2008). *Commodities Demystified*. New York: McGraw-Hill.

Fu, H., Liu, S., & Sun, F. (2010, August 4-7). *Ship Motion Prediction Based on AGA-LSSVM.* Proceedings of the International Conference on Mechatronics and Automation (ICMA), Xi'an, China.

Fu, X., & Yu, J. (2009, August 14-16). *A Hybrid Algorithm Based on Extremal Optimization with Adaptive Levy Mutation and Differential Evolution and Application.* Proceedings of the Fifth International Conference on Natural Computation (ICNC), Tianjin, China.

Gao, W., & Liu, S. (2012). A Modified Artificial Bee Colony. *Computers & Operations Research, 39*, 687-697.

Gao, W., Liu, S., & Huang, L. (2012). A Global Best Artificial Bee Colony Algorithm for Global Optimization. *Journal of Computational and Applied Mathematics, 236*, 2741-2753.

Gao, W., Liu, S., & Huang, L. (2013). A novel artificial bee colony algorithm with Powell's method. *Applied Science Computing, 13*, 3763-3775.

Gao, X., Wang, P., Qi, Y., Yan, A., Zhang, H., & Gong, Y. (2009, August 14-16). *Comparison Studies of LS-SVM and SVM on Modelling for Fermentation Processes.* Proceedings of the Fifth International Conference on Natural Computation (ICNC), Tianjian, China.

Gencoglu, M. T., & Uyar, M. (2009). Prediction of Flashover Voltage of Insulators using Least Squares Support Vector Machines. *Expert Systems with Applications, 36*, 10789-10798.

Hadavandi, E., Ghanbari, A., & Abbasian-Naghneh, S. (2010, August 13-15). *Developing a Time Series Model Based on Particle Swarm Optimization for Gold Price Forecasting.* Proceedings of the Third International Conference on Business Intelligence and Financial Engineering (BIFE), Hong Kong, China.

Haidar, I., & Wolff, R. C. (2011, October 9-12). *Forecasting of Crude Oil Price (Revisited).* Proceedings of the 30th United States Association for Energy Economics (USAEE), Washington D.C.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). *The Weka Data Mining Software: An Update; SIGKDD Explorations*. Retrieved from http://www.cs.waikato.ac.nz/ml/weka/.

Hammoudeh, S., Sari, R., & Ewing, B. T. (2008). Relationships Among Commodities and with Financial Variables: A New Look. *Contemporary Economic Policy, 27*(2), 251-264.

Hammoudeh, S., & Yuan, Y. (2008). Metal Volatility in Presence of Oil and Interest Rate Shocks. *Energy Economics, 30*, 606-620.

Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms* (2nd ed.). New York: A John Wiley & Sons, Inc., Publication.

He, R. J., & Yang, Z. Y. (2012). Differential Evolution with Adaptive Mutation and Parameter Control using Levy Probability Distribution. *Journal of Computer Science and Technology, 27*(5), 1035-1055.

Hegazy, O., Soliman, O. S., & Abdul Salam, M. (2013). A Machine Learning Model for Stock Market Prediction. *International Journal of Computer Science and Telecommunications, 4*(12), 17-23.

Hoorfar, A., Lakhani, S., & Jamnejad, V. (2007, September 17-21). *Application of Levy Mutation Operator in Evolutionary Programming Optimization of Antennas.* Proceedings of the International Conference on Electromagnetics in Advanced Applications (ICEAA), Torino, Italy.

Hsieh, T. J., Hsiao, H. F., & Yeh, W. C. (2011). Forecasting Stock Markets Using Wavelet Transforms and Recurrent Neural Networks: An Integrated System Based on Artificial Bee Colony Algorithm *Applied Soft Computing, 11*, 2510-2525.

Hsieh, T. J., Hsiao, H. F., & Yeh, W. C. (2012). Mining Financial Distress Trend Data using Penalty Guided Support Vector Machines based on Hybrid Particle Swarm Optimization and Artificial Bee Colony Algorithm. *Neurocomputing, 82*, 196-206.

Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 42* (2), 513-529.

Hussein, S. F.M, Shah, M. B.N, Jalal, M. R.A, & Abdullah, S. S. (2011, April 19-21). *Gold Price Prediction Using Radial Basis Function Neural Network.* Proceedings of the 4th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), Kuala Lumpur, Malaysia.

Hyndman, R. J., & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. Retrieved from https://www.otexts.org/fpp.

Hyndman, R. J., & Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy.*International Journal of Forecasting, 22*, 679-688.

Irani, R., & Nasimi, R. (2011). Application of Artificial Bee Colony-Based Neural Network in Bottom Hole Pressure Prediction in Underbalanced Drilling. *Journal of Petroleum Science and Engineering, 78*, 6-12.

Ismail, S., Shabri, A., & Samsudin, R. (2011). A Hybrid Model of Self-Organizing Maps (SOM) and Least Squares Support Vector Machines (LSSVM) for Time Series Forecasting. *Expert Systems with Applications, 38,* 10574-10578.

Jahjouh, M. M. (2012). *Design Optimization of Reinforced Concrete Frames using Artificial Bee Colony Algorithm.* (Master's Thesis, The Islamic University of Gaza, Palestine). Retrieved from library.iugaza.edu.ps/thesis/103769.pdf

Jain, R. K., Smith, K. M., Culligan, P. J., & Taylor, J. E. (2014). Forecasting Energy Consumption of Multi-family Residential Buildings using Support Vector Regression: Investigating the Impact of Temporal and Spatial Monitoring Granularity on Performance Accuracy. *Applied Energy, 123*, 168-178.

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

Jiang, B. T., & Zhao, F. Y. (2013). Particle Swarm Optimization-based Least Squares Support Vector Regression for Critical Heat Flux Prediction. *Annals of Nuclear Energy, 53*, 69-81.

Kao, L. J., Chiu, C.-C., Lu, C. J., & Ya\ng, J. L. (2013). Integration of Nonlinear Independent Component Analysis and Support Vector Regression for Stock Price Forecasting. *Neurocomputing, 99*, 534-542.

Karaboga, D. (2005). *An Idea Base on Honey Bee Swarm for Numerical Optimization* (Technical Report-TR06). Retrieved from http://mf.erciyes.edu.tr/abc/pub/tr06_2005.pdf.

Karaboga, D., & Akay, B. (2009). A Comparative Study of Artificial Bee Colony. *Applied Mathematics and Computation, 214*, 108-132.

Karaboga, D., & Basturk, B. (2008). On the Performance of Artificial Bee Colony (ABC) Algorithm. *Applied Soft Computing, 8*, 687-697.

Karaboga, D., Gorkemli, B., Ozturk, C., & Karaboga, N. (2012). A Comprehensive Survey: Artificial Bee Colony (ABC) Algorithm and Applications. *Artificial Intelligence Review*, 1-37. doi: 10.1007/s10462-012-9328-0

Kemp, D. D. (2004). *Exploring Environmental Issues: An Integrated Approach.* London: Routledge.

Kennedy, & Eberhart, R. (1995). *Particle Swarm Optimization.* Proceedings of the IEEE International Conference on Neural Networks (ICNN), Perth, Australia.

Khalifa, A. A. A., Hong, M., & Ramchander, S. (2010). Return Distributions and Volatility Forecasting in Metal Futures Markets: Evidence from Gold, Silver and Copper. *The Journal of Futures Markets, 31*(1), 55-80.

Khashman, A., & Nwulu, N. I. (2011, January 27-29). *Intelligent Prediction of Crude Oil Price Using Support Vector Machines.* Proceedings of the 9th

IEEE International Symposium on Applied Machine Intelligence and Informatics (SAMI), Smolenice, Slovakia.

Khazem, H. A. (2008). *Using Artificial Neural Networks to Forecast the Futures Prices of Crude Oil.* (D.B.A. dissertation). Retrieved from ProQuest Dissertations & Theses database (UMI No. 3295968).

Khemchandani, R., Jayadeva, & Chandra, S. (2009). Regularized Least Squares Fuzzy Support Vector Regression for Financial Time Series Forecasting. *Expert Systems with Applications, 36*, 132-138.

Kohzadi, N. (1994). *Neural Networks versus Time Series Models for Forecasting Commodity Prices.* (Ph.D dissertation). Retrieved from ProQuest Dissertations & Theses database (UMI No. 250231218).

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data Preprocessing for Supervised Learning. *International Journal of Computer Science, 1*(2), 1306-4428.

Krebs, R. E. (2003). *The Basics of Earth Science*. Connecticut: Greenwood Publishing Group, Inc.

Kulkarni, S., & Haidar, I. (2009). Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices. *International Journal of Computer Science and Information Security, 2*(1), 1-8.

Kumar, M., & Thenmozhi, M. (2007). Support Vector Machines Approach to Predict the S&P CNX NIFTY Index Returns. *Journal of Academy of Business and Economics, 7*(1), 1-19**.**

Labys, W. C. (2006). *Modelling and Forecasting Primary Commodity Prices*. Hampshire: Ashgate Publishing Limited.

Lai, L. K. C., & Liu, J. N. K. (2010, July 11-14). *Stock Forecasting Using Support Vector Machines.* Proceedings of the Ninth International Conference on Machine Learning and Cybernetics (ICMLC), Qingdao, China.

Lee, C. Y., & Yao, X. (2001, May 27-30). *Evolutionary algorithms with adaptive Levy mutations.* Proceedings of the Congress on Evolutionary Computation (ICEC), Seoul, South Korea.

Lee, C. Y., & Yao, X. (2004). Evolutionary programming using mutations based on the Levy probability distribution. *IEEE Transactions on Evolutionary Computation, 8*(1), 1-13.

Lee, W. P., & Cai, W. T. (2011, July 26-28). *A Novel Artificial Bee Colony Algorithm with Diversity Strategy.* Proceedings of the Seventh International Conference on Natural Computation (ICNC), Shanghai, China.

Lendasse, A., Ji, Y., Reyhani, N., & Verleysen, M. (2005, September 11-15). *LS-SVM Hyperparameter Selection with a Nonparametric Noise Estimator.* Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications (ICANN), Warsaw, Poland.

Levy, P. (1937). *Theorie de l'Addition des Veriables Aleatoires*. Paris, France: Gauthier-Villars.

Li, G. D., Masuda, S., & Nagai, M. (2013). An Optimal Prediction Model for Time Series Prediction in Manufcaturing Systems. *International Journal of Advance Manufacturing Technology, 67*, 2343-2349.

Li, H., Guo, S., Zhao, H., Su, C., & Wang, B. (2012). Annual Electric Load Forecasting by a Least Squares Support Vector Machines with a Fruit Fly Optimization Algorithm. *Energies, 5,* 4430-4445.

Li, X., Shao, Z., & Qian, J. (2002). An Optimizing Method base on Autonomous Animates: Fish Swarm Algorithm. *Systems Engineering Theory and Practice, 22*, 32-38.

Li, Y. (2009, November 8-12). *Short-Term Load Forecasting Based on LS-SVM Optimized by BCC Algorithm.* Proceedings of the 15th International Conference on Intelligent System Applications to Power Systems (ISAP), Curitiba, Brazil.

Liao, R., Zheng, H., Grzybowski, S., & Yang, L. (2011). Particle Swarm Optimization-Least Squares Support Vector Regression based Forecasting model on Dissolved Gases in Oil-Filled Power Transformers *Electric Power Systems Research, 81*, 2074-2080.

Liao, W., & Balzen, Z. (2013). LSSVM Network Flow Prediction Based on the Self-adaptive Genetic Algorithm Optimization. *Journal of Networks, 8*(2), 507-512.

Lin, C. S., Chiu, S. H., & Lin, T. Y. (2012). Empirical Mode Decomposition-based Least Squares Support Vector Regression for Foreign Exchange Rate Forecasting. *Economic Modelling, 29*, 2583-2590.

Liu, C. (2009, September 20-22). *Price Forecast for Gold Futures Based on GA-BP Neural Network.* Proceedings of the International Conference on Management and Service Science (MASS), Wuhan, China.

Liu, H. X., Yao, X. J., Zhang, R. S., Liu, M. C., Hu, Z. D., & Fan, B. T. (2005). Prediction of Tissue/Blood Partition Coefficients of Organic Compounds based on the Molecular Structure using Least Squares Support Vector Machines. *Journal of Computer-Aided Molecular Design, 19*, 499-508

Liu, L., & Wang, W. (2008, December 12-14). *Exchange Rates Forecasting with Least Squares Support Vector Machine.* Proceedings of the International Conference on Computer Science and Software Engineering (JCSSE), Kanchanaburi, Thailand.

Liu, P., & Yao, J. (2009, November 20-22). *Application of Least Square Support Vector Machine based on Particle Swarm Optimization to Chaotic Time Series Prediction.* Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), Shanghai, China.

Lomax, R. G. (2007). *An Introduction to Statistical Concepts*. New Jersey: Lawrence Erlbaum Associates, Inc. Publisher.

Lu, S., Cai, Z. J., & Zhang, X. B. (2009, August 8-11). *Application of GA-SVM time series prediction in tax forecasting.* Proceedings of the 2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Beijing, China.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting, 16*, 451-476.

Malliaris, M. E., & Malliaris, S. G. (2008). Forecasting Inter-Related Energy Product Prices. *The European Journal of Finance, 14*(6), 453-468.

Marsland, S. (2009). *Machine Learning An Algorithm Perspective*. Boca Raton, Florida: A Chapman & Hall Book.

McCulloch, J. H. (1996). *J. Huston McCulloch*. Retrieved March 15, 2012, from http://www.econ.ohio-state.edu/jhm/jhm.html

Mellit, A., Massi Pavan, A., & Benghanem, M. (2013). Least Squares Support Vector Machine for Short-Term Prediction of Meteorological Time Series. *Theor Appl Climatol, 111*, 297-307.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. New Jersey: John, Wiley & Sons.

Mustafa, M. W., Sulaiman, M. H., Shareef, H., & Abd. Khalid, S. N. (2011). Transmission Loss Allocation in Deregulated Power System Using the Hybrid Genetic Algorithm-Support Vector Machine Technique. *Cyber Journals: Multidiscplinary Journals in Science and Technology, Journal of Selected Areas in Renewable and Sustainable Energy (JRSE)*, 10-18.

Mustafa, M. W., Sulaiman, M. H., Shareef, H., & Abd. Khalid, S. N. (2012). Reactive Power Tracing in Pool-Based Power System utilising the Hybrid Genetic Algorithm and Least Squares Support Vector Machine. *IET, Generation, Transmission & Distribution, 6*(2), 133-141.

Nieto, P. J. G., Fernandez, J. R. A., Juez, F. J. d. C., Lasheras, F. S., & Muniz, C. D. (2013). Hybrid modelling based on support vector regression with genetic algorithms in forecasting the cyanotoxins presence in the Trasona reservoir (Northern Spain). *Environmental Research, 122*, 1-10.

Nourani, E., Rahmani, A. M., & Navin, A. H. (2012, May 21-22). *Forecasting Stock Prices using a Hybrid Artificial Bee Colony base Neural Networks.* Proceedings of the International Conference on Innovation, Management and Technology Research (ICIMTR), Malacca, Malaysia.

OANDA (2012). *Currency Converter.* Retrieved January 30, 2012, from http://www.oanda.com/currency/converter/

Ou, P., & Wang, H. (2009). Prediction of Stock Market Index Movement by Ten Data Mining Techniques. *Modern Applied Science, 3*(12), 28-42.

Pahasa, J., & Ngamroo, I. (2011). A heuristic training-based least squares support vector machines for power system stabilization by SMES. *Expert Systems with Applications, 38*, 13987-13993.

Pan, W. T. (2012). A New Fruit Fly Optimization Algorithm: Taking the Financial Distress Model as an Example. *Knowledge Based System, 26*, 69-74.

Papadimitriou, T., Gogas, P., & Stathakis, E. (2014). Forecasting Energy Markets using Suppoer Vector Machines. *Energy Economics, 44*, 135-142.

Parisi, A., Parisi, F., & Diaz, D. (2008). Forecasting Gold Price Changes: Rolling and Recursive Neural Network Models. *Journal of Multi. Finance Management, 18*, 477-487.

Park, J. B., Jeong, Y.W., Shin, J. R., & Lee, K. Y. (2010). An Improved Particle Swarm Optimization for Nonconvex Economic Dispatch Problems. *IEEE Transactions on Power Systems, 25*(1), 156-165.

Pelkmans, K., Suykens, J. A. K., Gestel, T. V., Brabanter, J. D., Lukas, L., Hamer, B., Vandewalle, J. (2002). *LS-SVMlab: A Matlab/C Toolbox for Least*

Squares Support Vector Machines. Retrieved 22 January, 2010, from http://www.esat.kuleuven.be/sista/lssvmlab/

Peng, Y. (2011). An Improved Artificial Fish Swarm Algorithm for Optimal Operation of Cascade Reservoirs. *Journal of Computers, 6*(4), 740-746.

Qi, J., Hu, J., Peng, Y. H., & Ren, Q. (2011). Electrical evoked potentials prediction model in visual prostheses based on support vector regression with multiple weights. *Applied Soft Computing, 11*, 5230-5242.

Radetzki, M. (2008). *A Handbook of Primary Commodities in the Global Economy*. United Kingdom: Cambridge University Press.

Rajasekhar, A., Abraham, A., & Pant, M. (2011, October 9-12). *Levy Mutated Artificial Bee Colony Algorithm for Global Optimization.* Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Anchorage, Alaska.

Rao, Y., Xu, S., & Xiong, L. (2011, March 26-28). *Time Series Prediction of Heavy Metal Contamination in Mining Areas Based on Exponential Smoothing Mode.* Proceedings of the International Conference on Information Science and Technology, Nanjing, Jiangsu, China.

Rashedi, E., Nezahabadi-pour, H., & Saryazdi, S. (2009). GSA: A Gravitational Search Algorithm. *Information Sciences, 179*, 2232-2248.

Richer, T. J., & Blackwell, T. M. (2006, July 16-21). *The Levy Particle Swarm.* Proceedings of the IEEE Congress on Evolutionary Computation (CEC), Vancouver, Canada.

Roy, R., & Jadhav, H. T. (2015). Optimal Power Flow Solution of Power System Incorporating Stochastic Wind power using Gbest Guided Artificial Bee Colony Algorithm. *Electrical Power and Energy Systems, 64*, 562-578.

Selvi, V., & Umarani, R. (2010). Comparative Analysisi of Ant Colony and Particle Swarm Optimization Techniques. *International Journal of Computer Applications, 5*(4), 1-5.

Shadbolt, J., & Taylor, J. G. (2002). *Neural Networks and the Financial Markets*. London: Springer-Verlag.

Sharma, T. K., Pant, M., & Bansal, J. C. (2012, June 10-15). *Some Modifications to Enhance the Performance of Artificial Bee Colony.* Proceedings of the IEEE World Congress on Computational Intelligence (WCCI), Brisbane, Australia.

Sharma, T. K., Pant, M., & Bhardwaj, T. (2011, December 4-7). *PSO Ingrained Artificial Bee Colony Algorithms for Solving Continuous Optimization Problems.* Proceedings of the International Conference on Computer Applications and Industrial Electronics (ICCAIE), Penang, Malaysia.

Shen, W., Zhang, Y., & Ma, X. (2009, July 24-26). *Stock Return Forecast with LS-SVM and Particle Swarm Optimization.* Proceedings of the International Conference on Business Intelligence and Financial Engineering (BIFE), Beijing, China.

Shi, J. Y., Zou, X. B., Huang, X. W., Zhao, J. W., Li, Y., Hao, L., & Zhang, J. (2013). Rapid Detecting Total Acid Content and Classifying Different Types of Vinegear based on Near Infrared Spectroscopy and Least Squares Support Vector Machines. *Food Chemistry, 138*, 192-199.

Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An Application of Support Vector Machines in Bankruptcy Prediction Model. *Expert Systems with Applications, 28*, 128-135.

Shrivastava, N. A., Ch, S., & Panigrahi, B. K. (2011, December 28-30). *Price Forecasting using Computational Intelligence Techniques: A Comparative Analysis.* Proceedings of the International Conference on Energy, Automation, and Signal (ICEAS), Bhubaneswar, Odisha, India.

Storn, R., & Price, K. (1997). Differential Evolution- ASimple and Efficient Heuristic for Global Optimization Over Continuous Spaces. *Journal of Global Optimization, 11*, 341-359.

Subotic, M. (2011, April 28-30). *Artificial Bee Colony Algorithm with Multiple Onlookers for Constrained Optimization Problems.* Proceedings of the European Computing Conference (ECC), Paris, France.

Sulaiman, M. H., Mustafa, M. W., Shareef, H., & Abd. Khalid, S. N. (2012). An Application of Artificial Bee Colony Algorithm with Least Squares Support Vector Machine for Real and Reactive Power Tracing in Deregulated Power System. *Electrical Power and Energy Systems, 37*(1), 67-77.

Sun, W., & Zhang, J. (2008, January 28-29). *Forecasting Day Ahead Spot Electricity Prices Based on GASVM.* Proceedings of the International Conference on Internet Computing in Science and Engineering (ICICSE), Harbin, China.

Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least Squares Support Vector Machines*. Leuven, Belgium: World Scientific Publishing Co. Pte. Ltd.

Tan, S. T. (2010). *Applied Mathematics for the Managerial, Life, and Social Sciences* (6th ed.). Belmont, USA: Richard Stratton.

Tang, L., & Hammoudeh, S. (2002). An Empirical Exploration of the World Oil Price Under the Target Zone Model. *Energy Economics, 24*, 577-596.

Tarhouni, M., Laabidi, K., Zidi, S., & Ksouri-Lahmari, M. (2011, March 22-25). *A Nonlinear MIMO Systems Identification Based on Improved Multi-Kernel Least Squares Support Vector Machines (Improved Multi-Kernel LSSVM).* Proceedings of the 8th International Multi-Conference on Systems, Signals and Devices (SSD), Sousse, Tunisia.

Taylor, J. B., & Weerapana, A. (2012). *Principles of Economics* (7th ed.). Mason, USA: South Western, Cengage Learning.

Tehrani, R., & Khodayar, F. (2011). A hybrid optimized Artificial Intelligent Model to Forecast Crude Oil using Genetic Algorithm. *African Journal of Bussiness Management, 5*(34), 13130-13135.

Tsai, P. W., Pan, J. S., Liao, B. Y., & Chu, S. C. (2009). Enhanced Artificial Bee Colony Optimization. *International Journal of Innovative Computing, Information and Control, 5*(12), 1-12.

U.S. Energy Information Administration. (2013). *Energy and Financial Merkets: What Drives Crude Oil Prices?* Retrieved February 22, 2013, from http://www.eia.gov/finance/markets/

Van Gestel, T., A. K. Suykens, J., Baestaens, D. E., Lambrechts, A., Lanckriet, G., Vandaele, B., Vandewalle, J. (2001). Finance Time Series Prediction Using

Least Squares Support Vector Machines Within the Evidence Framework. *IEEE Transactions on Neural Networks, 12*(4), 809-821.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory* (2nd ed.). New York: Springer-Verlag.

Varahrami, V. (2011). Recognition of Good Prediction of Gold Price Between MLFF and GMDH Neural Network. *Journal of Economics and International Finance, 3*(4), 204-210.

Vlad, S. (2010). Investigation of Chaotic Behavior in Euro-Leu Exchange Rate. *Journal of Applied Computer Science and Mathematics, 8*(4), 67-71.

Wang, D. C., Wang, C. X., Xie, Y. H., & Zhu, T. Y. (2010, December 13-15). *Air Temperature Prediction Based on EMD and LSSVM.* Proceedings of the Fourth International Conference on Genetic and Evolutionary Computing (ICGEC), Shenzhen, China.

Wang, J., Kou, L., Hou, X., & Zhou, Z. (2010, June 25-27). *Empirical Analysis on Co-movement of Stock Price of Gold Mine Enterprises and the International Gold Price.* Proceedings of the International Conference on Electrical and Control Engineering (ICECE), Wuhan, China.

Wang, K. M., & Lee, Y. M. (2010). Could Gold Serve as an Exchange Rate Hedge in Japan. *Inzinerine Ekonomika-Engineering Economics, 21*(2), 160-170.

Wang, S., Yu, L., & Lai, K. K. (2005). Crude Oil Price Forecasting with TEI@I Methodology. *Journal of Systems Science and Complexity, 18*(2), 145-166.

Wang, X., & Li, H. (2009, October 16-18). *One-Month Ahead Prediction of Wind Speed and Output Power Based on EMD and LSSVM.* Proceedings of the International Conference on Energy and Environment Technology (ICEET), Guilin, China.

Wang, X., Zhang, H., Zhang, C., Cai, X., Wang, J., & Wang, J. (2005, December 5-8). *Prediction of Chaotic Time Series using LS-SVM with Automatic Parameter Selection.* Proceedings of the Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), Dalian, China.

Wang, Y., & Li, B. (2008, June 1-6). *A Restart Univariate Estimation of Distribution Algorithm: Sampling Under Mixed Gaussian and levy Probability Distribution.* Proceedings of the Proceedings of the IEEE World Congress on Computational Intelligence, Hong Kong, China.

Wang, Y., & Li, Y. (2010, November 7-9). *Applying LS-SVM to Predict Primary Energy Consumption.* Proceedings of the International Conference on E-Product E-Service and E-Entertainment (ICEEE), Henan, China.

Wang, Y. S., & Chueh, Y. L. (2013). Dynamic Transmission Effects between the Interest Rate, the US Dollar, and Gold and Crude Oil Prices. *Economic Modelling, 30*, 792-798.

Williams, G. (2011). *Data Mining with Rattle and R The Art of Excavating Data for Knowledge Discovery*. New York: Springer.

Wu, D., Feng, S., & He, Y. (2007). Infrared Spectroscopy Technique for the Nondestructive measurement of Fat Content in Milk Powder. *Journal Dairy Science, 90*, 3613-3619.

Wu, J., & Niu, D. (2009, October 28-30). *Short-Term Power Load Forecasting Using Least Squares Support Vector Machines(LS-SVM).* Proceedings of the Second International Workshop on Computer Science and Engineering (WCSE), Qingdao, China.

Xiang, Y., & Jiang, L. (2009, January 23-25). *Water Quality Prediction Using LS-SVM and Particle Swarm Optimization.* Proceedings of the Second International Workshop on Knowledge Discovery and Data Mining (WKDD), Moscow, Russia.

Xiao, J., He, C., & Wang, S. (2012, August 18-21). *Crude Oil Price Forecasting: A Transfer Learning based Analog Complexing Model.* Proceedings of the Fifth International Conference on Bussiness Intelligence and Financial Engineering (BIFE), Lanzhou, Gansu, China.

Xie, C., Shao, C., & Zhao, D. (2011). Parameters Optimization of Least Squares Support Vector Machines and Its Application. *Journal of Computers, 6*(9), 1935-1941.

Xie, H. L., Li, N., Lu, F. C., & Xie, Q. (2009, March 28-30). *Application of LS-SVM by GA for Dissolved Gas Concentration Forecasting in Power Transformer Oil.* Proceedings of the Asia-Pacific Power and Energy Engineering Conference (APPEEC), Wuhan, China.

Yang, J. H., & Dou, W. (2012, November 17-18). *Prediction of Gold Price based on WT-SVR and EMD-SVR model.* Proceedings of the Eighth International Conference on Computational Intelligence and Security, Guangzhou, China.

Yang, X. S. (2010). *Nature-Inspired Metaheuristic Algorithms* (2nd ed.). United Kingdom: Luniver Press.

Yang, Z., Gu, X. S., Liang, X. Y., & Ling, L. C. (2010). Genetic Algorithm-Least Squares Support Vector Regression based Predicting and Optimizing Model on Carbon Fiber Composite Integrated Conductivity. *Materials and Design, 31*, 1042-1049.

Yeh, C. C., Chi, D. J., & Hsu, M. F. (2010). A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications, 37*, 1535-1541.

Yorucu, V. (2003). The Analysis of Forecasting Performance by Using Time Series Data for Two Mediterranean Islands. *Review of Social, Economic & Business Studies, 2*, 175-196.

Yu, L., Chen, H., Wang, S., & Lai, K. K. (2009). Evolving Least Squares Support Vector Machines for Stock Market Trend Mining. *IEEE Transactions on Evolutionary Computation, 13*(1), 87-102.

Yu, L., Wang, S., & Lai, K. K. (2007). An Online Learning Algorithm with Adaptive Forgetting Factors for Feedforward Neural Networks in Financial Time Series Forecasting. *Nonlinear Dynamics and Systems Theory, 7*(1), 51-66.

Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting Crude Oil Price with an EMD-based Neural Network Ensemble Learning Paradigm. *Energy Economics, 30*, 2623-2635.

Yu, L., Wang, S., Wen, B., & Lai, K. K. (2008, June 18-20). *An AI-Agent-Based Trapezoidal Fuzzy Ensemble Forecasting Model for Crude Oil Price*

Prediction. Proceedings of the Third International Conference on Innovative Computing Information and Control (ICICIC), Dalian, China.

Yuan, Y., Zhuang, X. T., Liu, Z. Y., & Huang, W.Q. (2014). Analysis of the temporal properties of price shock sequences in crude oil markets. *Physica A, 394*, 235-246.

Yusof, N., A. Rashid, R., & Mohamed, Z. (2010, December 5-7). *Malaysia Crude Oil Production Estimation: an Application of ARIMA Model.* Proceedings of the International Conference on Science and Social Research (CSSR), Kuala Lumpur, Malaysia.

Zhai, J., & Huang, L. (2013). Marketing Prediction Based on Time Series Prediction Algorithm of Least Squares Support Vector Machines. *Journal of Convergence Information Technology, 8*(3), 245-250.

Zhang, D., Guan, X., Tang, Y., & Tang, Y. (2011, May 28-29). *Modified Artificial Bee Colony for Numerical Optimization.* Proceedings of the 3rd International Workshop on Intelligent Systems and Applications (ISA), Wuhan, China.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with Artificial Neural Network: The State of Art. *International Journal of Forecasting, 14*, 35-62.

Zhang, S., Wang, F., & He, D. (2012). Real-time product quality control for batch processes based on stacked least-squares support vector regression models. *Computers and Chemical Engineering, 36*, 217-226.

Zhang, T., & Fukushige, A. (2002, May 12-17). *Forecasting Time Series by Bayesian Neural Networks.* Proceedings of the International Joint Conference on Neural Networks (IJCNN), Honolulu, Hawaii.

Zhang, W., Niu, P., Li, G., & Li, P. (2013). Forecasting of Turbine Heat Rate with Online Least Squares Support Vector Machine based on Gravitational Search Algorithm *Knowledge-Based Systems, 39*, 34-44

Zhang, X., Wu, Q., & Zhang, J. (2010, October 20-21). *Crude Oil Price Forecasting Using Fuzzy Time Series.* Proceedings of the 3rd International Symposium on Knowledge Acquisition and Modeling (KAM), Wuhan, China.

Zhang, Y. J., & Wei, Y. M. (2010). The Crude Oil Market and the Gold Market: Evidence for Cointegration, Causality and Price Discovery. *Resources Policy, 35*(3), 168-177.

Zhou, B., & Shi, A. (2010, August 13-15). *LSSVM and Hybrid Particle Swarm Optimization for Ship Motion Prediction.* Proceedings of the International Conference on Intelligent Control and Information Processing (ICICIP), Dalian, China.

Zhou, S., & Lai, K. K. (2011). An Improved EMD Online Learning-Based Model for Gold Market Forecasting. In J. Watada et. al (Eds.), *Intelligent Decision Technologies Smart Innovation, Systems and Technologies* (pp. 75-84). Berlin Heidelberg: Springer-Verlag

Zhou, S., Lai, K. K., & Yen, J. (2012). A Dynamic Meta-Learning Rate-Based Model for Gold Market Forecasting. *Expert Systems with Applications, 39*, 6168-6173.

Zhu, B., & Wei, Y. (2013). Carbon Price Forecasting with a Novel Hybrid ARIMA and Least Squares Support Vector Machines Methodology. *Omega, 41*, 517-524.

Zhu, G., & Kwong, S. (2010). Gbest-guided artificial bee colony algorithm for numerical function optimization. *Applied Mathematics and Computation, 217*, 3166-3173.

Zou, M. (2009, July 12-15). *Condition Prediction of Hydroturbine Generating Units using Least Squares Support Vector Regression with Genetic Algorithm.* Proceedings of the Eighth International Conference on Machine Learning and Cybernetics (ICMLC), Baoding, China.