ENHANCED ABC-LSSVM FOR ENERGY FUEL PRICE PREDICTION

Zuriani Mustaffa[1], Yuhanis Yusof[2] and Siti Sakira Kamaruddin[3]

School of Computing, UUM College of Arts and Sciences,
Universiti Utara Malaysia,
06010 Sintok, Kedah, Malaysia

Corresponding author: zuriani.m@gmail.com[1]

ABSTRACT

This paper presents an enhanced Artificial Bee Colony (*e*ABC) based on Lévy Probability Distribution (LPD) and conventional mutation. The purposes of enhancement are to enrich the searching behavior of the bees in the search space and prevent premature convergence. Such an approach is used to improve the performance of the original ABC in optimizing the embedded hyper-parameters of Least Squares Support Vector Machines (LSSVM). Later on, a procedure is put forward to serve as a prediction tool to solve prediction task. To evaluate the efficiency of the proposed model, crude oil prices data was employed as empirical data and a comparison against four approaches were conducted, which include standard ABC-LSSVM, Genetic Algorithm-LSSVM (GA-LSSVM), Cross Validation-LSSVM (CV-LSSVM), and conventional Back Propagation Neural Network (BPNN). From the experiment that was conducted, the proposed *e*ABC-LSSVM shows encouraging results in optimizing parameters of interest by producing higher prediction accuracy for employed time series data.

Keywords: Artificial bee colony, least squares support vector machines, levy probability distribution, prediction.

INTRODUCTION

Prediction is one of the critical data mining issues where the task of prediction is essential for planning and operation of organizations, including government agencies (Maimon, & Rokach, 2010). Prediction can be divided into two major types: Prediction of categorical class labels, or typically known as

classification, and numeric prediction which predicts continuous values (Han, Kamber, & Pei, 2012). In numeric prediction (or prediction), the application of Machine Learning (ML) techniques such as Artificial Neural Network (ANN), Support Vector Machines (SVM) and Least Squares Support Vector Machines (LSSVM) have contributed significantly in solving various prediction issues from different areas such as in the military system (Sun, Huang, & Zi, 2009), electrical engineering (Afshin, Sadeghian, & Raahemifar, 2007), finance (Van Gestel et al., 2001) and many more.

However, even though the above techniques come with good capability in learning and generalization, several embedded parameters that need to be optimized have limited their application, especially ANN which requires more control parameters as compared to SVM and LSSVM (Xiang & Jiang, 2009). With respect to this matter, the emergence of Evolutionary Computation (EC) in optimization community has contributed to the growth of scientific researches in dealing with various optimization issues, including in the optimization of ML control parameters. As a stochastic optimization approach, probabilistic rule is employed in the EC algorithm for the sake of improving a solution (Karaboga, & Basturk, 2007). Some forms of EC includes Evolutionary Algorithm (EA) and Swarm Intelligence (SI) (Karaboga, Gorkemli, Ozturk, & Karaboga, 2012). Prior to the emergence of SI techniques, EA approaches gained interest among the academia, especially Genetic Algorithm (GA) (Karaboga et al., 2012). As compared to other techniques under the same roof, such as Genetic Programming (GP) and Evolution Strategies (ES), GA is considered the most popular technique in terms of diversity in application(Yang, 2010). The good performance of GA has been proven in solving many optimization problems, including in prediction task, such as prediction of power flow tracing (Mustafa, Sulaiman, Shareef & Abd. Khalid, 2011) and stock market (Yu, Chen, Wang, & Lai, 2009). In both studies, GA is utilized to optimize hyper-parameters of LSSVM while in Wei and Jie (2008), GA-SVM is presented to predict electricity prices.

On the other hand, the emergence of Swarm Intelligence (SI) techniques has boosted the field of optimization. This method, which is also known as foraging algorithms (El-Abd, 2012) has been applied in a large number of optimization problems, which includes prediction. Interestingly, for prediction purposes, not only has the application of SI been applied as a hybrid method with ML but also utilized as a single approach. Examples of hybrid method between SI and ML are Particle Swarm Optimization (PSO) with LSSVM in predicting chaotic time series and water quality respectively (Ping, & Jian, 2009; Xiang, & Jiang, 2009) and financial prediction utilizing hybridization of

ABC-LSSVM (Mustaffa, & Yusof, 2011). Besides LSSVM, ABC is also used to optimize the weights area for training ANN in predicting the earthquake magnitude by Shah and Ghazali (2011).

In another case, the application of a single SI for prediction tasks has been proposed by Hadavandi, Ghanbari, and Abbasian-Naghneh (2010). By using gold price time series data, PSO was employed to predict the gold price of year 2005 to 2006. Besides PSO, ABC also gained interest in dealing with prediction tasks, especially in biotechnology, as presented in Bahamish, Abdullah, and Abdul Salam (2009) and Bahamish and Abdullah (2010).

As a relatively new optimization technique and besides having the advantage of less tuning parameters, ABC has been proven to be better than or at least comparable to other existing optimization techniques such as GA and PSO (El-Abd, 2012). With the main objective to improve the capability of standard ABC, this paper is devoted to developing an enhanced ABC (*eA*BC) algorithm by utilizing Lévy Probably Distribution(LPD) and conventional mutation. The new eABC is later utilized to optimize LSSVM's hyper parameters, namely regularization parameter, γ and kernel Radial Basis Function (RBF) parameter, σ^2, and as well as α, which belongs to LPD. Then prediction of crude oil price is done by LSSVM. Classified as strategic commodities, crude oil is one of the main commodities which possesses high volatility level (Hammoudeh, Sari, & Ewing, 2008). It plays a major role in the global economy as well as being A primary source of energy. Thus, the fluctuation of its price is the key element that influences government plans and policy (Jammazi, & Aloui, 2012).

REGRESSION USING LEAST SQUARES SUPPORT VECTOR MACHINES

Support Vector Machines (SVM) has been proven to be good in Machine Learning (ML) community (Deris, Zaki, & Chin, 2002); however, there is still a gap that can be filled in. This has led to the emerging of a new variation of SVM, which was first joined by Suykens and colleagues in 2002, namely Least Squares Support Vector Machines (LSSVM) (Suykens, Van Gestel, De Brabanter, De Moor, & Vandewalle, 2002). It is a modified form of SVM which offers a better solution strategy. In LSSVM, the solution is obtained by solving a linear set of equations rather than a hard Quadratic Programming (QP) solver in SVM, which in practice is harder to use as compared to linear equations (Suykens et al., 2002). The proposed approach simplifies the complex calculation which leads to an easier and faster training task. Thus,

the simpler optimization problem can be obtained. In addition, in terms of control parameters, LSSVM offers less control hyper-parameters, which is γ and $\sigma2$, as compared to three control parameters required in SVM (C, $\sigma2$ and ε) (Ou & Wang, 2009). The main theory of this approach is to map the non-linear input data from input space to high dimensional feature space. In high dimensional feature space, the linear separation can be done (see Figure 1). A brief introduction of LSSVM is presented here while further discussion can be found in Suykens et al., (2002).
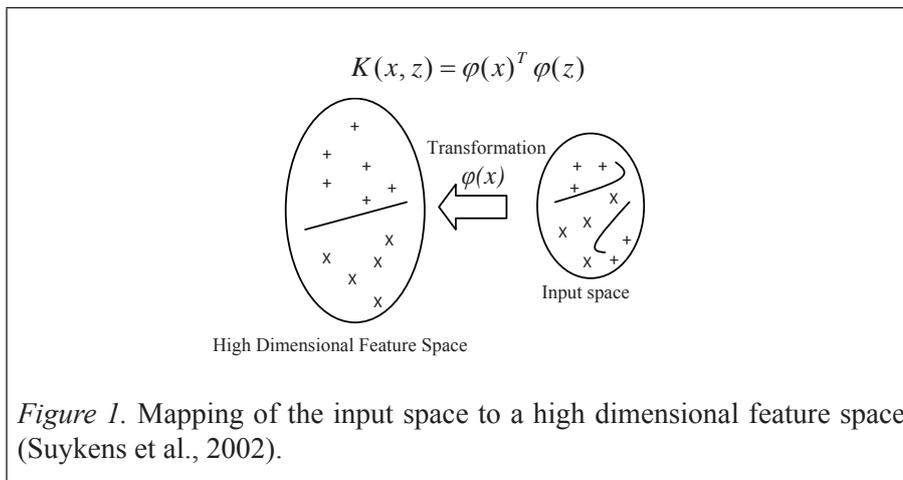

Figure 1. Mapping of the input space to a high dimensional feature space (Suykens et al., 2002).

The standard framework for LSSVM is based on the primal-dual formulation. Formally, given a training set$\{x_i, y_i\}^N i=1, 2, \ldots, N$, with the input vector 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 aim is to estimate a model of the following form (Suykens et al., 2002):

$$y(x) = w^T \phi(x_i) + b + e_i \tag{1}$$

where *w* is the weight vector, *b* denotes the bias and e_i is additive noise or error variable, $x \in R^n, y \in R,$ and $\phi(x_i): R^n \rightarrow R^{nh}$ is a mapping function that aims to map the input vectors *x* to a high dimensional feature space. The output *y(x)* is a daily spot price of crude oil from day 21 onwards while the input vector x_i consists of four input features. The coefficient vector w and bias term b are unknown and can be obtained through the optimization problem which is formulated as follows (Suykens et al., 2002):

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

Subject to the equality constraints

$$y_i = w^T \phi(x_i) + b + e_i \; , i=1, 2 \dots, N$$

With the application of Mercer's theorem (Vapnik, 1995), for the kernel matrix O as $\Omega_{ij} = K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, $i, j = 1,..,N$ it is not required to compute explicitly the nonlinear mapping $f(.)$ as this is done implicitly through the use of positive definite kernel functions K.

From the Lagrangian function which is corresponding to (3):

$$\zeta(w,b,e;\alpha) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha_i (w^T \phi(x_i) + b + e_i - y_i) \tag{3}$$

where $\alpha^i \in R$ $(i = 1, 2, \dots, N)$ are Lagrange multipliers called support values that can be positive or negative in the LSSVM formulation due to the equality constraints, γ is the regularization parameter which balances the complexity of the model with the training error. Differentiating (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 described in Karush-Kuhn-Tucker (KKT)(4):

$$\begin{cases} \dfrac{\partial \zeta}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\ \dfrac{\partial \zeta}{db} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ \dfrac{\partial \zeta}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma_i e_i \\ \dfrac{\partial \zeta}{\partial \alpha_i} = 0 \rightarrow y_i = w^T \varphi(x_i) + b + e_i \\ \qquad\qquad\qquad\qquad i=1,2,\dots,N \end{cases} \tag{4}$$

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

$$\begin{bmatrix} 0 & 1^T \\ y & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{5}$$

with $y = [y_1, \dots, y_N]^T$, $\alpha = [\alpha_1, \dots, \alpha_N]^T$. The resulting LS-SVM model for the regression in dual space becomes:

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{6}$$

where α_i and b are the solutions to the linear (5). In (6), several kernel functions $K(x, x_i)$, are available namely Gaussian Kernel or RBF Kernel, MLP Kernel and Quadratic Kernel are available. In this study, the RBF Kernel is used. Besides offerings less parameters to set, it tends to give a good performance under general smoothing assumptions (Ou & Wang, 2009), is able to reduce the computational complexity of the training process (Wu, Feng, & He, 2007), is suitable in dealing with non-linear cases (Bao, Zhang, Yu, Lai, & Wang, 2011) and is able to perform outstanding performance in many regression cases (Liao, Zheng, Grzybowski, & Yang, 2011). It is expressed as:

$$K(x, xi) = \exp(-\frac{\| x - x_i \|^2}{2\sigma^2})$$
(7)

where σ^2 is a tuning parameter which is associated with RBF function. Another tuning parameter, which is the regularization parameter, γ as can be seen in (2).

PARAMETERS OPTIMIZATION BASED ON ARTIFICIAL BEE COLONY

Introduced by Karaboga in 2005, Artificial Bee Colony (ABC) algorithm is a metaheuristic algorithm which emerged from observing of intelligent behavior of honey bees (Karaboga, 2005). Theoretically, an artificial bee swarm is formed by three groups of bees, namely Employed Bees (EB), Onlooker Bees (OB) and Scout Bees (SB). Both OB and SB are also known as unemployed bees (Karaboga et al., 2012). The divisions of colony members are equal, where half of the colony is composed of EBs and the other half comprised of OBs. The number of food sources/nectar sources is equal to the EBs, which means that one EB is associated with a specific nectar source. The aim of the whole colony is to maximize the amount of nectar. In ABC, the exploitation or intensification process is handled by EBs and OBs, while the exploration or diversification process is carried out by the SBs.

Generally, the duty of the EB is to search for new food sources (v_{ij}), which represent the possible solutions. Later, the amount of nectar (solutions' qualities/fitness value) is calculated. The information collected by the EBs is later passed on to the OBs which are waiting in the hive (dance area). Here, the OBs need to make A decision, whereby they have to exploit a nectar source depending on the information shared by the EBs. For this matter, OBs watch different dances before doing the selection of food source position according to the probability, which is proportional to the quality of that food source. The OBs also determine the source to be cast aside and allocate the associated

EB as SB. In the SB phase, the SB finds the new valuable food sources at which point it once again becomes the EB. They search the space near the hive randomly.

In ABC algorithm, initial food sources are produced randomly within the range of the boundaries of the parameters (Karaboga, 2005).

$$x_{ij} = x_j^{min} + rand(0,1)(x_j^{max} - x_j^{min})$$ (8)

where i=1,…, *SN*, j=1,…*D*. *SN* is the number of food sources and *D* is the number of parameters of interest. Meanwhile, x_j^{min} and x_j^{min} are the upper and lower bound of parameters of interest respectively. The fitness value of the randomly chosen site is defined as follows (Karaboga, 2005):

$$fit_i = \frac{1}{(1 + obj.Fun_i)}$$ (9)

where *obj.Func* is the objective function, the size of EB and OB are both equal to the number of food sources, denoted by *SN*. For each food source position, one EB is assigned to it. For each EB whose total numbers are equal to the number of the food sources, a new food source is obtained according to (10) (Karaboga, 2005):

$$v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj})$$ (10)

where;

$i = 1,2,…, SN$
$j = 1,2,…,D$
f = a random generalized real number within the range [-1, 1], k = is a randomly selected index number in the colony.

After producing the new solution, v_{ij}, its fitness is calculated and compared to the old solution, x_{ij}. Here, a greedy selection is applied, whereby, if the new solution is better than the previous one, the bee memorizes the new solution; otherwise she memorizes the previous solution. The OB selects a food source to exploit with the probability values related to the fitness values of the solution. This probability is calculated using the following equation (Karaboga, 2005):

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j}$$ (11)

where *fit*$_i$ is the fitness of the solution *v*. *SN* is the number of food source positions. Later, the OB searches for a new solution in the selected food source site using (10), in a similar way as exploited by the EBs. As the quality of the food increases, the number of OB visiting the food source increases too. Here, Roulette Wheel scheme is applied. The food source selection by OB represents the positive feedback in ABC algorithm. On the other hand, in the SB phase, if the fitness of a found food source hasn't been improved for a given number of trials (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 (12).

$$x_{id} = x_d^{min} + r\left(x_d^{max} - x_d^{min}\right) \tag{12}$$

where;

r = a random real number within the range [0.1]
x_d^{min} and x_d^{max} = the lower and upper borders in the *d*th dimension of the problems space.

Basic steps of ABC algorithm are as follows:

1. Initialize the food source positions (population)
2. Each EB is assigned to its food sources.
3. Each OB selects a source based on the quality of her solution, produces a new food source in the selected food source site and exploits the better source.
4. Decide the source to be abandoned and assigns it to the EB as SB for discovering new food sources.
5. Memorize the best food source (solution) found so far.
6. If the requirements are met, output the best solution, otherwise repeat steps 2-5 until the stopping criterion is met or maximum iteration is achieved.

Data representation of ABC is as tabulated in Table 1.

Table 1

Data Representation of ABC Algorithm

Data Representation	Variable
D	Parameters to be optimized, namely: i) γ ii) σ² iii) **α** (see section D(ii))

(continued)

Data Representation	Variable
New food source position, v_{ij}	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	Mean Absolute Percentage Error (MAPE) (See section E)

From the literature, it is revealed that the basic structure of ABC favors exploration more in comparison to exploitation (Gao, & Liu, 2012). However, by referring to Yang (2010), it is vital, to obtain an ideal combination of the exploitation and exploration process, for the sake of global optimal. nonetheless, in practice, both processes are opposed to each other. from the ABC algorithm, it can be seen that the exploitation process in both the EB and OB phases highly rely on a single equation, which represents the strategy applied. By applying a single equation for different phases, it has limited the searching process. Thus, to enrich the searching behavior of the bees, a modification in the exploitation process is introduced.

METHODOLOGY

This section discusses methodology utilized in the study. The discussion covers the research data and data preparation, data normalization, experiment setup, *e*ABC-LSSVM and the utilized performance evaluation metric.

Data acquisition and preprocessing

In this study, four correlated energy fuel price (time series data) were employed as input, namely crude oil (CL), heating oil (HO), gasoline (HU) and propane (PN). The time series data covered are from December 1997 to November 2002. This is similar to the one utilized by Malliaris and Malliaris (2008). The samples of datasets are as tabulated in Table 2.

Table 2

Samples of Original Input

Date	CL	HO	HU	PN
1-12-1997	18.63	0.5144	0.5338	0.3200
2-12-1997	18.70	0.5190	0.5316	0.3188

Date	CL	HO	HU	PN
3-12-1997	18.60	0.5220	0.5328	0.3213
4-12-1997	18.59	0.5218	0.5272	0.3198
5-12-1997	18.70	0.5276	0.5262	0.3200

Test for correlation

The correlation is considered as the most intuitive reflection to describe the relationship among the data (Wang, Kou, Hou, & Zhou, 2010). In this study, the correlation coefficient was tested using Pearson Product Moment Correlation Coefficient (Lomax, 2007), and it is defined as r=0.8-1.0 which indicates a high correlation (Ahmed Farid, & Salahudin, 2010). The higher the value of correlation coefficient reflects the level of significance between the prices of commodities with the input factors. Table 3 shows the correlation between energy fuel prices.

Table 3

Correlation between Energy Fuel Prices from December 1997 to November 2002

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

From Table 3 above, it clearly shows that the correlation coefficients for the above mentioned energy fuels prices are high in the sample of data, especially between CL-HO, CL-HU and HO-HU.

Test for non-linearity (Raw Data)

The purpose of conducting the test for non-linearity is to support the justification of employing the non-linear prediction model (i.e. LSSVM). In this study, the BDS test (Brock, Scheinkman, Dechert, & LeBaron, 1996) is applied. The BDS tests is a two-tailed test, where the null hypothesis of independence and identical distribution will be rejected if Z>2 at a significance level of 5% (Vlad, 2010).

The BDS test results for the series of interest are summarized in Table 4-7. From the tables, it indicates that the z-Statistic are greater than the critical

values significantly. Thus, the null hyphothesis is rejected. This implies that time series of interest exhibits a significant non-linear behavior which proves the sutaibility of employing a non-linear model for further tasks. The results obtained also reinforce the facts provided by numerous literature that crude oil prices are governed by non-linear features (Abdullah, & Zeng, 2010; Bao et al., 2011; Haidar, & Wolff, 2011). The similar results also indicated by the other three variables are not surprising since HO, HU and PN are not only refining products of crude oil but also possesses high correlation with its prices (Malliaris, & Malliaris, 2008).

Table 4

The Values of BDS Test: CL

CL				
Dimension	BDS Statistic	Standard Error	z-Statistic	Prob.
2	0.195033	0.001242	157.0196	0.0000
3	0.331891	0.001962	169.1750	0.0000
4	0.426831	0.002320	183.9557	0.0000
5	0.492162	0.002401	204.9558	0.0000
6	0.536655	0.002299	233.4320	0.0000

Table 5

The Values of BDS Test: HO

HO				
Dimension	BDS Statistic	Standard Error	z-Statistic	Prob.
2	0.192315	0.001363	141.1323	0.0000
3	0.326891	151.8276	151.8276	0.0000
4	0.420029	164.8749	164.8749	0.0000
5	0.483535	183.3142	183.3142	0.0000
6	0.526276	208.2991	208.2991	0.0000

Table 6

 The Values of BDS Test: HU

HU				
Dimension	BDS Statistic	Standard Error	z-Statistic	Prob.
2	0.189073	0.001188	159.1354	0.0000
3	0.321582	0.001881	171.0004	0.0000

(continued)

HU				
Dimension	BDS Statistic	Standard Error	z-Statistic	Prob.
4	0.412931	0.002229	185.2621	0.0000
5	0.47998	0.002312	205.4883	0.0000
6	0.516509	0.002218	232.9067	0.0000

Table 7

The Values of BDS Test: PN

PN				
Dimension	BDS Statistic	Standard Error	z-Statistic	Prob.
2	0.193377	0.001716	112.7183	0.0000
3	0.327679	0.002714	120.7303	0.0000
4	0.419635	0.003216	130.4755	0.0000
5	0.481672	0.003335	144.4204	0.0000
6	0.522660	0.003200	163.3450	0.0000

Data normalization

Prior to training, all input and output were normalized using linear normalization, namely Min Max Normalization (Al-Shalabi et al., 2006). The goal of linear scaling is to independently normalize each feature component to the specified range. This ensures that the larger input value does not dominate the smaller input value. In addition, this approach may improve the prediction accuracy and data mining algorithm (Al-Shalabi et al., 2006). The definition of Min Max Normalization is as shown in the following equation:

$$v' = \left((v - \min_a)/(\max_a - \min_a)\right) * (new\max_a - new\min_a) + new\min_a \qquad (13)$$

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 [$new\max_a$, $new\min_a$] by solving the equation above. The normalized input for the sample in Table 2 is as tabulated in Table 8.

Table 8

Samples of Normalized Input

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

Experiment setup

Input and output arrangement

Table 9 indicates the variables assigned to the features involved. the input arrangement employed is as suggested by Malliaris and Malliaris (2008). the output was the daily spot price of crude oil from day 21 onwards(cl21).

Table 9

Assigning Input and Output Variables

Input	Variable	Output
Daily closing price of crude oil, heating oil, gasoline and propane	CL, HO, HU, PN	
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	CL21
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	

The daily spot price will help the model to fix to current price location while the purpose of including the derivation input is to help the model to learn the underlying relationship that is constant over time (Malliaris & Malliaris, 2008).

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 overfitting) and testing set (for real assessment of how well the model generalize) (Marsland, 2009). The proportion for each subset is classified into three classes as tabulated in Table 10. For each class, the data arrangement is based on the arithmetic progression formula (Tan, 2010). The purpose of having three different kinds of data proportion is to choose the best data proportion with respect to the generalization.

Table 10

Data Proportion for Training, Validation and Testing

	Training	Validation	Testing	Description of Data Arrangement
A	60	20	20	Training (2,3,4), (6,7,8), …, N where N = Total number of training + validation Validation = 1+ (n-1) x d, where d=4, n=1,2,3, …, N Testing(1000-1248)
B	70	15	15	Training = (2,3,4,5,6), (8,9,10,11,12), …, N Validation = (1+ (n-1) x d, where d=6, n=1,2,3, …, N Testing (1062-1248)
C	80	10	10	Training (2,3,4,5,6,7,8,9), (11,12,13,14,15,16,17,18), …, N where N = Total number of training + validation Validation = 1+ (n-1) x d, where d=9, n=1,2,3, …, N Testing (1125-1248)

*e*ABC-LSSVM

In the proposed model, the modification involved 2 major components and two sub-components. Firstly, the major components of ABC and LSSVM are hybridized, termed as ABC-LSSVM. Secondly, mutation based on LPD is applied in the EB and OB phases (sub-component A), termed as *lv*ABC. The purpose is to enhance the searching process of the bees for the sake of global optima. Finally, to prevent premature convergence, conventional mutation is introduced (sub component B). Both sub components A and B are embedded in the ABC-LSSVM, to form the *e*ABC-LSSVM. The simplified form of proposed methodology as illustrated in Figure 2.

Figure 2. *e*ABC-LSSVM Model.
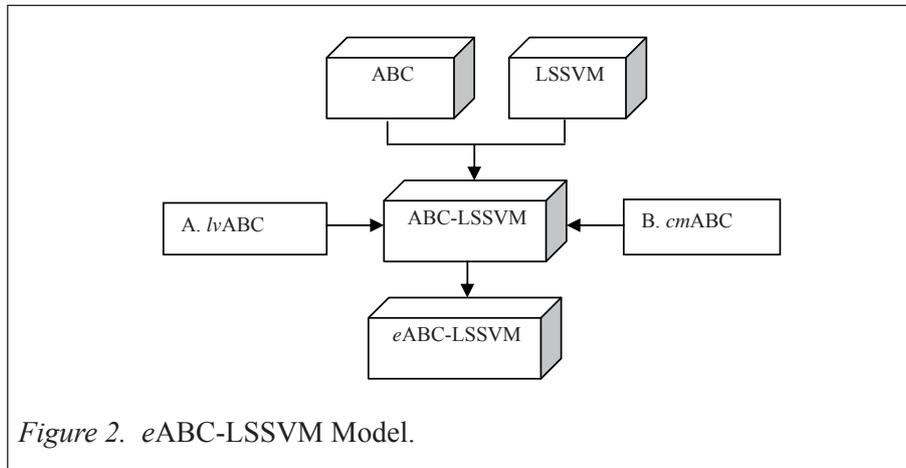
Hybrid ABC-LSSVM

In this study, there are three adjustable parameters of interest: regularization parameter, γ, kernel RBF parameter, σ^2 and α. The γ and σ^2 are associated to LSSVM hyper parameters where γ is responsible to determine the tradeoff between Empirical Risk Minimization (ERM) and Structural Risk Minimization (SRM) while σ^2 controls the value of function regression error (Wu et al., 2007). On the other hand, α which belongs to LPD, is responsible to control the shape of the distribution. All of these parameters are subject to A predetermined constraint where [1,1000] for γ and σ^2, while for α, it is subject to $0 < \alpha < 2$. The goal is to find the ideal values of all parameters that will generate the minimum (lowest) value of the objective function, whereby in this study is the Mean Absolute Percentage Error (MAPE).

To describe how the proposed method works, LSSVM will be embedded in the ABC algorithm as fitness function evaluation, so the optimized value of parameters of interest can be obtained after a maximum number of iterations has been achieved. The objective function to be minimized is MAPE, where the lower the MAPE, the better the prediction accuracy.

Levy Mutation for improving exploitation process of bees in search space (*lv*ABC)

To enhance the exploitation process in the standard ABC, Lévy Mutation in ABC (*lv*ABC) is applied in generating new food sources, which involves the EB and OB phases. The *lv*ABC is based on Lévy Probability Distribution (LPD) which was first introduced by P. Levy in 1930 (Lee, & Yao, 2004). The advantage of LPD relies on its tuning parameter, namely α where different

87

value of α will yield different shapes of probability distribution. The smaller the α, the longer the tail of the distribution. This characteristic provides an effective strategy for searching since it provides A wider search space and reduces the possibility of revisiting a similar location (Lee, & Yao, 2004).

$$L_{\alpha,\theta}(y) = \frac{1}{\pi} \int_0^\infty e^{-\theta q^\alpha} \cos(qy) dq \, y \in R \tag{14}$$

From (14), the distribution is symmetric with respect to $y=0$ and has two parameters, α and θ. α controls the shape of the distribution, requiring $0 < \alpha < 2$ while θ is the scaling factor satisfying $\theta > 0$. Since the analytic form of (14) is unknown generally, an algorithm to generate Lévy random number is commonly utilized (Lee, & Yao, 2004; 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 (14), the following relation can be obtained:

$$L_{\alpha,\theta}(by) = \frac{1}{b} L_{\alpha,\theta'}(y) \tag{15}$$

where $\theta = \theta b^{-\alpha}$. In particular, by setting $\theta' = 1$, (15) becomes:

$$L_{\alpha,\theta}(y') = \gamma^{\frac{-1}{\theta}} L_{\theta,1}(y) \tag{16}$$

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).

In the EB phase, instead of applying (10), the following equation is introduced:

$$v_{ij} = x_{ij} + \left(x_{ij} - x_{kj}\right) + L \tag{17}$$

On the other hand, in the OB phase, the following equation is used to replace (10):

$$v_{ij} = x_{ij} + L \tag{18}$$

Eqs. (17) and (18) are designed through experimental approach. Here, L is represented by a random number generated by LPD. The main objective is to enrich the searching process by applying different strategies in both phases. In this study, all parameters involved are automatically tuned by eABC-LSSVM.

Conventional mutation for preventing premature convergence (*cm*ABC)

In this study, the adaptation of conventional mutation is as a precautionary step in preventing the model from converging too quickly where this situation may lead the model to be trapped in local minima. By applying a mutation strategy, the model is induced to explore other areas in order to look for global minima rather than local minima (Haupt, & Haupt, 1998).

In the standard ABC-LSSVM (Mustaffa, & Yusof, 2011), if the generated parameter value is out of boundaries, it is automatically shifted onto the boundaries. Nonetheless, in the proposed *cm*ABC, instead of forcing the parameter value to the boundary, a mutation strategy is applied. This operation is executed by multiplying the generated random number with the range of boundary that has been determined. In this study, the boundaries are set to the range of between [1, 1000].

$$new\_param = (ub - lb) * rand\_num \tag{19}$$

where *new_param* = new parameter
rand_num = random number
ub = upper bound
lb = lower bound

The pseudo code of the *e*ABC-LSSVM is as given in Figure 3.

```
Initialize the food source positions (solutions) xᵢ, i = 1,…,SN
Evaluate the nectar amount (fitness function fitᵢ) of food
sources with LSSVM based on training and validation sets
Cycle = 1
REPEAT
EMPLOYED BEE (EB) PHASE
For each EB
    Produce new food source positions (solutions)using
    Eq.(17)(lvABC)
    Evaluate solutions vᵢⱼ and xᵢⱼ
IfVᵢⱼ is better than xᵢⱼ
    Greedy selection
If solution is out of boundary, apply Eq. (19) (cmABC)
    Calculate the value fitᵢand evaluate (train and validate)
    using LSSVM
    Memorized best solution
End For
Calculate the probability values pᵢ for the best solution (11)
Select a solution xᵢⱼ depending on pᵢ
ONLOOKER BEE (OB) PHASE
For each OB
```

(continued)

```
    Choose the food source depending on p_i
    Produce  new  food  source  positions  (solutions)using
    Eq.(18)(lvABC)
    Evaluate solutions v_ij and x_ij
IfV_ij is better than x_ij
    Greedy selection
    If solution is out of boundary, apply Eq. (19) (cmABC)
    Calculate the value fit_iand evaluate (train and validate)
    using LSSVM
    Memorized best solution
End For
SCOUT BEE (SB) PHASE
    Abandon its food source
    Search for new ones to replace the abandoned food
    source
Memorized the best solution
Cycle = Cycle + 1
Until Maximum Cycle Number
```

*Figure 3. e*ABC-LSSVM Algorithm

Performance evaluation metric

Choosing an appropriate performance evaluation metric is crucial in order to justify the results obtained from the conducted experiment. In this study, the evaluation of the prediction model is guided by five statistical metrics, namely the Mean Absolute Percentage Error (MAPE), symmetric MAPE (sMAPE), Root Mean Square Percentage Error (RMSPE) (Hyndman, & Koehler, 2006), Prediction Accuracy (PA) and Theil's U(Armstrong, 2001). These metrics interpret the learning and generalization capabilities of the prediction model. The formula for the above mentioned metrics are shown in (20) - (24).

$$MAPE = \frac{1}{N}\left[\sum_{n=1}^{N}\left|\frac{y_n - y(x_n)}{y_n}\right|\right] \qquad (20)$$

The interpretation of results produced by MAPE is as tabulated in Table 11 (Yorucu, 2003):

Table 11

Range of Percentage by MAPE

Percentage	Interpretation
10% <	Highly accurate prediction
10-20%	Good prediction

(continued)

Percentage	Interpretation
20-50%	Reasonable prediction
>50%	Inaccurate prediction

$$sMAPE = mean(200 \mid y_n - y(x_n) \mid /(y_n + y(x_n))) \tag{21}$$

$$RMSPE = \sqrt{\left(\frac{\sum_{n=1}^{N}\left(\frac{y_n - y(x_n)}{y_n}\right)^2}{N}\right)} \tag{22}$$

$$PA = 100\% - (MAPE \times 100) \tag{23}$$

$$Theil's U = \frac{\sqrt{\frac{1}{N}\Sigma_{n=1}^{N}(y_n - y(x_n))^2}}{\sqrt{\frac{1}{N}\Sigma_{n=1}^{N}(y_n)^2} + \sqrt{\frac{1}{N}\Sigma_{n=1}^{N}(y(x_n))^2}} \tag{24}$$

The interpretation of results produced by Theil's U is as tabulated in Table 12 (Frechtling, 2001):

Table 12

Range of Theil's U

Theil's U	Interpretation
<1	The prediction technique is better than Random Walk. The smaller the Theil's U is, the better its prediction technique is.
1	The prediction technique is about as good as Random Walk
>1	The prediction technique is worse than Random Walk.

When the value of Theil's U is equal to zero, it means that the predicted value is equal to the actual/target value. Thus, it is better for Theil's U value to be close to zero (Tseng & Lee, 2006).

Where $n = 1, 2, ..., x$

y_n = actual values

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

N = Number of test data

EMPIRICAL RESULTS AND DISCUSSION

All experiments were performed on MATLAB software and the LS-SVMlab Toolbox (Pelkmans et al*., 2002) was applied to derive all the hybrid LSSVM models. The computer configuration is as tabulated in Table 13. Prior to experiment execution, the properties of the proposed technique was set as tabulated in Table 14. Four approaches (ABC-LSSVM, GA-LSSVM, BPNN and CV-LSSVM) are selected for the purpose of comparison against the proposed technique. Brief reviews of competitors chosen under study are as follows:

i.   ABC-LSSVM (Mustaffa, & Yusof, 2011)
     The proposed technique was developed based on this approach. Details of ABC and LSSVM are as presented in Section 2 and 3.

ii. ANN/BPNN
     Inspired by observation of the human brain, ANN is designed to learn in a similar way to the human brain (Zhang, Patuwo, & Hu, 1998). As a prominent ML technique, BPNN is always chosen as a comparable benchmark to LSSVM (Chen, Wu, & Chen, 2008; Tan, Yan, Gao, & Yang, 2012) . The BPNN consists of an input layer, a hidden layer and an output layer. In determining the property of BPNN such as Learning rate and momentum constant, the man-made-choice approach was utilized.

iii. GA-LSSVM (Mustafa et al*., 2011)
     GA-LSSVM is one of the recent state of the art algorithms. GA is pioneered by John Holland, and later popularized by David Goldberg (Haupt, & Haupt, 1998). This technique is built based on three fundamental operators, namely selection, crossover and mutation. The popularity of GA makes it the most dominant technique as compared to the other approaches under the EA group.

iv. CV-LSSVM
     It is common for CV to be utilized as an optimization technique in LSSVM (Fu, Liu, & Sun, 2010), and has been presented by many researchers including in Afshin et al., (2007), Wang and Li (2010) and many others. Generally, in CV, the average accuracy of the N(N=Number of partitions) is used to estimate the prediction accuracy that will be obtained if the produced parameter values was used (Afshin et al*., (2007).

At the first stage, by using three different classes of data proportion, namely A, B and C (see Table 10), the experiments to measure the performance of *e*ABC-LSSVM were conducted and the results from this experiment are tabulated in Table 15. The parameters with the minimum MAPE of validation are chosen as the ideal value for parameters of interest. From the table, the *e*ABC-LSSVM has the superior capability in producing lower MAPE by using data proportion from set B. The best MAPE is obtained at $\gamma=353.2191$, $\sigma^2=1.9226$ and $\alpha=0.976$, which is 94.4608%. This is followed by data proportion from set C with 0.5114% lesser and finally data proportion set A with 92.0922% of prediction accuracy.

By employing the data proportion from set B, the same experimental procedure was performed for the four identified approaches, including ABC-LSSVM, GA-LSSVM, CV-LSSVM, BPNN and with no exception of *lv*ABC-LSSVM and *cm*ABC-LSSVM which were embedded in *e*ABC-LSSVM. The empirical results clearly show that, with the increment of 0.3871% of PA from ABC-LSSVM, the values of performance evaluations metrics are in favor of *e*ABC-LSSVM. Both *lv*ABC-LSSVM and *cm*ABC-LSSVM recorded an increment of 0.1086% and 0.2182% of PA as compared to the PA produced by ABC-LSSVM, respectively.

On the other hand, MAPE offered by GA-LSSVM is 6.3512%, which is slightly lower than MAPE produced by CV-LSSVM, which is 6.4917%. For BPNN, a significant difference of MAPE was recorded, which is more than 10%. This makes it the performance of BPNN far lower than other techniques. The significant difference in terms of MAPE produced by BPNN with the rest of the prediction models maybe because of the other prediction models are based on the LSSVM prediction tool and share similar features in terms of the regression model. That is, the LSSVM prediction-based models rely on the Structural Risk Minimization (SRM) principle which minimizes an upper bound of generalization error rather than training error which is implemented in ANN (Bao et al., 2011). In addition, it is also well-documented that in practice, the BPNN usually requires a large amount of training data (up to 10 years of historical daily data) to ensure it performs well in prediction (Haidar & Wolff, 2011), and this is not the case for LSSVM (Xiang, & Jiang, 2009).

Table 17 shows the results obtained from the *t*-Test. From the table, it shows that the modifications that have been introduced are significant at 0.05% significance level. That is, the superiority of *e*ABC-LSSVM is significantly improved, not only better than ABC-LSSVM and its variants (*lv*ABC-LSSVM and *cm*ABC-LSSVM), but also with three other prediction models. The statistical results in graphical representation are visualized in Fig. 4 while the

comparison of the convergence rate between ABC-LSSVM, *e*ABC-LSSVM and GA-LSSVM are shown in Fig. 5. From Fig. 5, it clearly shows that *e*ABC-LSSVM is able to escape from the local minima area which finally resulted in higher prediction accuracy. On the other hand, the fitness curves of ABC-LSSVM decrease at the very beginning of the iterations and subsequently the curve becomes flat, which consequently causes the model to face premature convergence.

Table 13

Computer Configuration

	Configuration
CPU	Intel® Core™ i3-2330M processor
Memory	2GB
Hard Disk	6.00 GB of RAM
Operating System	64-bit Windows 7

Table 14

Properties of eABC/ABC

Properties	
SN	10
Limit	*SN*D*
Maximum iteration	100

Table 15

Comparison among Different Data Proportion for eABC-LSSVM

	A	B	C
Γ	795.7964	353.2191	74.4067
σ^2	273.2709	1.9226	1.3317
A	0.611	0.976	0.1871
MAPE (%)	7.9078	5.5392	6.0506
PA (%)	92.0922	94.4608	93.9494
sMAPE	8.2589	5.5583	6.3185
RMSPE	0.0942	0.0722	0.0737
Theil's U	0.0492	0.0358	0.0391

Table 16

Comparison among Different Prediction Models for CL Price Prediction

	γ	σ²	α	MAPE (%)	PA (%)	sMAPE	RMSPE	Theil's U
*e*ABC-LSSVM	353.2191	1.9226	0.976	5.5392	94.4608	5.5583	0.0722	0.0358
*lv*ABC-LSSVM	249.1869	1	0.4537	5.8177	94.1823	5.931	0.0745	0.0378
*cm*ABC-LSSVM	808.7678	4.1738	-	5.7081	94.2919	5.6888	0.0736	0.0361
ABC-LSSVM	304.5763	1	-	5.9263	94.0737	6.0508	0.0758	0.0386
GA-LSSVM	26.0286	17.7216	-	6.3512	93.6488	6.5168	0.075	0.0393
CV-LSSVM	16.461	17.7216	-	6.4917	93.5083	6.6787	0.0766	0.0403
BPNN	-	-	-	12.2009	87.7991	12.2333	0.1525	0.0808

Table 17

t-Test Results: eABC-LSSVM vs. Different Prediction Models

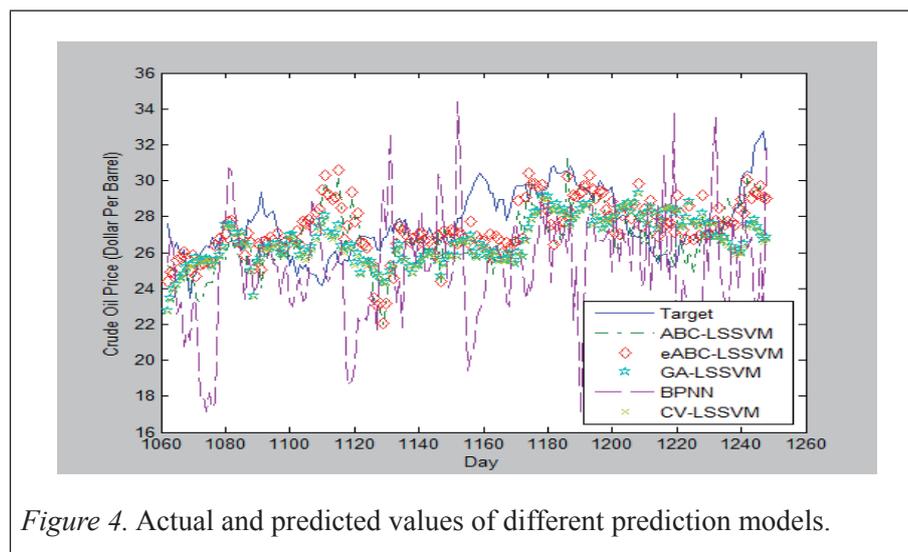| Prediction Models	Mean	Std. Deviation	Sig. (2-tailed)
*e*ABC-LSSVM - *lv*ABC-LSSVM	.3034	.3817	.000
*e*ABC-LSSVM – *cm*ABC-LSSVM	-.3534	.5634	.000
*e*ABC-LSSVM – ABC-LSSVM	.3560	.4598	.000
*e*ABC-LSSVM – GA-LSSVM	.6424	1.0609	.000
*e*ABC-LSSVM - BPNN	2.1831	3.4510	.000
*e*ABC-LSSVM – CV-LSSVM	.7292	1.0858	.000

Figure 4. Actual and predicted values of different prediction models.
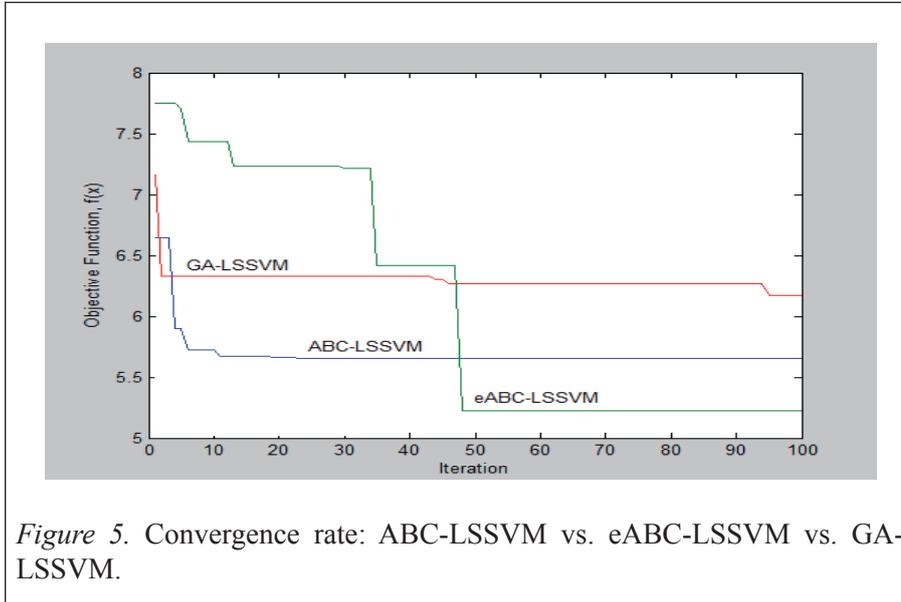
Figure 5. Convergence rate: ABC-LSSVM vs. eABC-LSSVM vs. GA-LSSVM.

CONCLUSION

This study reports the empirical results which examine the feasibility of *e*ABC-LSSVM in predicting crude oil prices. The performance of the proposed prediction model was guided using five statistical metrics, namely MAPE, PA, sMAPE, RMSPE and Theil's U. Later, the comparison with several comparable competitors was made. Two modifications that have been introduced makes *e*ABC superior in the searching strategy as compared to THE original ABC which is highly dependent on single equation in the searching process. In addition, the proposed technique also has proven its capability in avoiding premature convergence which finally led to a good generalization performance. Consequently, the combination of both modifications leads to the superiority of *e*ABC-LSSVM which may become a promising prediction model for the context of study.

REFERENCES

Abdullah, S. N., & Zeng, X. (2010). *Machine learning approach for crude oil price prediction with Artificial Neural Networks-Quantitative (ANN-Q) model.* Proceedings of the International Joint Conference on Neural Networks (IJCNN).

Afshin, M., Sadeghian, A., & Raahemifar, K. (2007). *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.

Ahmed Farid, J., & Salahudin, F. (2010). *Risk frameworks and applications* (2nd ed.). Karachi: Alchemy Technologies.

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

Armstrong, J. S. (2001). *A handbook for researchers and practitioners*. New York: Springer.

Bahamish, H. A. A., & Abdullah, R. (2010). *Prediction of C-peptide structure using artificial bee colony algorithm.* Proceedings of the International Symposium in Information Technology (ITSim).

Bahamish, H. A. A., Abdullah, R., & Abdul Salam, R. (2009). *Protein tertiary structure prediction using artificial bee colony algorithm.* Proceedings of the Third Asia International Conference on Modelling and Simulation.

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.

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.

Chen, Q., Wu, Y., & Chen, X. (2008). *Research on customers demand forecasting for E-business web site based on LS-SVM.* Proceedings of the International Symposium on Electronic Commerce and Security.

El-Abd, M. (2012). Performance assessment of foraging algorithms vs. evolutionary algorithms. *Information Sciences, 182*, 243-263.

Frechtling, D. C. (2001). *Forecasting tourism demand: Methods and strategies*. Woburn: Elsevier.

Fu, H., Liu, S., & Sun, F. (2010). *Ship motion prediction based on AGA-LSSVM.* Proceedings of the International Conference on Mechatronics and Automation (ICMA).

Gao, W.-f., & Liu, S.-Y. (2012). A modified artificial bee colony. *Computers & Operations Research, 39*, 687-697.

Hadavandi, E., Ghanbari, A., & Abbasian-Naghneh, S. (2010). *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).

Haidar, I, & Wolff, C. R.(2011). Forecasting crude oil price (revisited). Proceeding of the 30th USAEE Conference (Accepted).

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.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques* (3rd ed.). Waltham: Morgan Kaufmman Publishers.

Haupt, R. L., & Haupt, S. E. (1998). *Practical genetic algorithms*. New York: Wiley Pub.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*, 679-688.

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

Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization.* Erciyes University, Engineering Faculty, Computer Engineering Department.

Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *Journal Global Optimization, 39*, 459-471.

Karaboga, D., Gorkemli, B., Ozturk, C., & Karaboga, N. (2012). A comprehensive survey: Artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review.* Springer.

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.

Lendasse, A., Ji, Y., Reyhani, N., & Verleysen, M. (2005). *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.

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, 8*, 2074-2080.

Lomax, R. G. (2007). *An introduction to statistical concepts*. New Jersey: Lawrence Erlbaum Associates.

Maimon, O., & Rokach, L. (2010). *Data mining and knowledge discovery handbook* (2nd ed.). New York: Springer.

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 from http://economics.sbs.ohio-state.edu/jhm/programs/STABRND.M

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),* January.

Mustaffa, Z., & Yusof, Y. (2011). Optimizing LSSVM using ABC for non-volatile financial prediction. *Australian Journal of Basic and Applied Sciences, 5*(11), 549-556.

Ou, P., & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science, 3*(12), 28-42.

Pelkmans, K., Suykens, J. A. K., Gestel, T. V., Brabanter, J. D., Lukas, L., Hamer, B., et al. (2002). *LS-SVMlab: A matlab/C toolbox for least squares support vector machines.*

Ping, L., & Jian, Y. (2009). *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).

Shah, H., & Ghazali, R. (2011). *Prediction of earthquake magnitude by an improved ABC-MLP.* Proceedings of the Developments in E-systems Engineering (DeSE).

Sun, S.-X., Huang, Y., & Zi, S.-Y. (2009). *Research on weapon system cost forecasting model based on chaos optimization LSSVM.* Proceedings of the 16th International Conference on Management Science & Engineering Moscow, Russia.

Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least squares support vector machines*: World Scientific Publishing.

Tan, G., Yan, J., Gao, C., & Yang, S. (2012). Prediction of water quality time series data based on least squares support vector machines. *Procedia Engineering, 31*(2012), *1194-1199.*

Tan, S. T. (2010). *Applied mathematics for the managerial, life, and social sciences* (6th ed.). Belmont, USA: Richard Stratton.

Tseng, N.-J., & Lee, Y.-H. (2006). Obtaining forecasting capability by examining equity valuation models: A case of Taiwan's sea cargo industry. *The Journal of Human Resource and Adult Learning, May,* 117-122.

Van Gestel, T., Suykens, J. A. K., Baestaens, D. E., Lambrechts, A., Lanckriet, G., Vandaele, B., et al. (2001). Financial 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.

Vlad, S. (2010). Investigation of chaotic behavior in euro-leu exchange rate. *Journal of Applied Computer Science and Mathematics, 8*(4), 67-71.

Wang, J., Kou, L., Hou, X., & Zhou, Z. (2010). *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).

Wang, Y., & Li, Y. (2010). *Applying LS-SVM to predict primary energy consumption.* Proceedings of the International Conference on E-Product E-Service and E-Entertainment (ICEEE).

Wei, S., & Jie, Z. (2008). *Forecasting day ahead spot electricity prices based on GASVM.* Proceedings of the International Conference on Internet Computing in Science and Engineering (ICICSE).

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.

Xiang, Y., & Jiang, L. (2009). *Water quality prediction using LS-SVM and particle swarm optimization.* Proceedings of the Second International Workshop on Knowledge Discovery and Data Mining (WKDD).

Yang, X.-S. (2010). *Nature-inspired metaheuristic algorithms* (2nd ed.). United Kingdom: Luniver Press.

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.

Zaki, N. M., Deris, S., & Chin, K. K. (2002). Extending the decomposition algorithm for support vector machines training. *Journal of ICT, 1*(1), 17-29.

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.