

Multiresolution Analysis Based on Wavelet Transform for Commodity Prices Time Series Forecasting

Supaporn Bunrit, Nittaya Kerdprasop, and Kittisak Kerdprasop

Abstract—This research proposes time series forecasting of commodity prices by using multiresolution analysis from wavelet transform. In this work, discrete wavelet transform based multiresolution decomposition of Daubechies family is used. Firstly, Daubechies wavelet transform is applied to the training set of time series data up to level four. The reconstruction values of the approximation part of wavelet from each level are then used for the forecasting process by ARIMA model. The validation set of data is used to analyze and select the best model from all 4 levels of multiresolution decomposition. Finally, best selected validating model is used for evaluating the remaining testing data set. The forecasting results by using multiresolution analysis are compared to the case where the original data are directly modeled and forecasted by ARIMA. Results based on the mean absolute percentage error evaluation from using multiresolution analysis are better for both of the two studied data including daily gold price and rubber price. By applying multiresolution analysis, the improvement is 10.83% for gold price and 42.68% for rubber price. The variances of errors from the proposed method on both data sets are also much less than directly use the original time series data for forecasting.

Index Terms—Multiresolution analysis, wavelet transform, commodity prices, time series forecasting.

I. INTRODUCTION

Commodities such as oil, gold, rice, copper, natural gas, sugar, and rubber are the basic goods used in commerce that is interchangeable with other commodities of the same type. They are most often used as inputs in the production of other goods or services. The quality of a given commodity may differ slightly, but it is essentially uniform across producers. As a result, product differentiation cannot be used for market competition. Price is incapable defined by producers; it depends on demand-supply of the market. In addition, many factors may affect the commodity price fluctuation, for example, US dollar price, inflation rate, speculators, and hoarding of goods.

By the nature of commodities that can be bought, sold or exchanged around the world, their prices will reflect the overall of world market. The recovery or the deflation of world economics can be manifested from commodities price. Price of commodities, thereby, is another one significant factor to prescribe the characteristics of world or country

Manuscript received January 26, 2018; revised March 1, 2018. This work was supported by grants from Suranaree University of Technology (SUT), Thailand.

The authors are with the School of Computer Engineering, SUT, 111 University Avenue, Muang, Nakhon Ratchasima 30000, Thailand (corresponding author: Supaporn Bunrit; tel.: +66944961244; e-mail: sbunrit@sut.ac.th, nittaya@sut.ac.th, kerdpras@sut.ac.th).

economics. Many stakeholders or investors are needed on keeping track or watching of commodities prices. Accurately estimating or forecasting of such prices is, as a result, needed from all people involved. Researching on commodities prices forecasting, as such, are continuously interesting and challenging. One biggest challenge comes from the fluctuation of commodities price that has been affected from many external factors. Time series data of commodities prices are always non-stationary. Therefore, novel techniques and methods in forecasting suitable for such kind of data are needed to explore and apply continuously in order to increase the forecasting accuracy.

The objective of this research is applying multiresolution analysis of wavelet transform for commodities prices time series forecasting. The proposed method is compared to the case that the multiresolution analysis has not been applied. ARIMA models are used in forecasting for such comparison. Two of commodities price time series data are studied which are gold and rubber prices in Thailand market.

II. LITERATURE REVIEWS

A. Wavelet Transform in Time Series Analysis

Wavelet transform is a powerful tool for a time-scale multiresolution analysis on time series and has been used to break down an original time series into different components, each of which may carry meaningful signals of the original time series. There are numerous applications of discrete wavelet transforms in various domains of time series data analysis and mining [1], including 1) wavelet-based time series similarity search, 2) wavelet-based time series classification, 3) wavelet-based clustering, 4) wavelet-based trend, surprise, and pattern detection, and 5) wavelet-based prediction or forecasting.

In the domain of similarity search, the capability of searching for similar or exact time series is required. The research in this area mostly focuses on feature extraction [2], [3], indexing, and similarity measurements [4], [5]. Indexing methods specially designed for multidimensional data have been proposed and explored for similarity search.

For wavelet-based time series classification domain, the objective is to assign a class label to a time series. Examples of application domains are speech recognition [6], audio classification [7], texture classification [8], and medical signal classification [9].

The goal of wavelet-based clustering of time series is to group similar time series data together into the same clusters and put dissimilar time series into different clusters. It has been applied in diverse application domains, such as

clustering stocks in the stock market [10], and clustering gene expressions [11].

Wavelet-based trend, surprise, and pattern detection are deals with anomaly detection in time series. It helps users identify abnormal time series or parts of time series from the rest of the data. Detection of abnormal events includes network performance problem detection, intrusion detection, disease outbreak detection, bio-terrorist attack detection, and change detection in classification models that can be reviewed in [1].

Time series data prediction or forecasting is the task of forecasting data values in the future by given actual values of such a forecasting time point and the historical data values. The main idea behind time series prediction is to understand the movement or the characteristics of historical data. Works on wavelet transform in time series forecasting are the forecasting of stock prices, electricity and water consumption, river water level, Internet usage, disease outbreak and physiological symptoms [1].

B. Wavelet Transform in Commodities Price Time Series Forecasting

Although wavelet transform in time series forecasting use in a variety of domains in literature, there exists quite few that involves commodities prices. For commodities product, discrete wavelet transform (DWT) based method exists in forecasting of crude oil price [12], oil price [13], and natural gas price [14] that are the most interesting products in views of many researchers. There also has a work for forecasting metal prices that consists of aluminum, copper, lead, and zinc [15]. For our work, we investigate differently on two commodities products: gold price and rubber price.

III. THEORIES

A. Time Series

A time series is a sequence of data that demonstrate recorded of values of the same variable over time. Most commonly, it is a sequence taken at successive uniform time intervals. In order to permit valid statistical comprehension, these values must be repeatedly recorded, often over a year to many years period. Many application domains involve time series data, such as financial, meteorological, medical, social sciences, computer networks, and business [1]. Time series consist of four components: 1) seasonal variations that repeat over a specific period such as a day, week, month, season, etc., 2) trend variations that move up or down in a reasonably predictable pattern, 3) cyclical variations that correspond with business or economic 'boom-bust' cycles or follow their own peculiar cycles, and 4) random variations or fluctuations that do not fall under any of the above three classifications.

B. Wavelet Transform (WT)

WT is a transformation that decomposes the original data or signal $f(x)$ into many scales and many positions by a group of basic functions called "wavelets". A group of basis function, $\psi_{a,b}(x)$, comes from a process of dilation and translation of mother wavelet function, $\psi(x)$, expressed in (1)

[16].

$$\psi_{a,b}(x) = a^{-1/2} \psi\left(\frac{x-b}{a}\right) \quad a, b \in \mathbb{R}, a \neq 0 \quad (1)$$

where, a is a scale and b is a position.

For continuous WT of data or signal, $f(x)$ can be defined by (2).

$$W_{\psi(a,b)} f = \frac{1}{\sqrt{C_{\psi}}} \int_{-\infty}^{\infty} \psi_{a,b}(x) f(x) dx \quad (2)$$

where, C_{ψ} is a normalization constant.

But most calculation in practice, data or signals always come from discrete sampling. The discrete format of WT is then defined by equation 3.

$$W_{\psi(a,b)} f = \frac{1}{N} \frac{1}{\sqrt{|a|}} \sum_{k=1}^N \frac{1}{\sqrt{C_{\psi}}} \psi\left(\frac{x-b}{a}\right) f(x) \quad (3)$$

Mother wavelet function in DWT is selected to be the prototype for every basis functions in the transformation process. Therefore, all basis functions come from mother wavelet that are dilated (or compressed) and shift (or translated). Mother wavelet functions are proposed in many families; the most popular ones are Haar, Daubechies and Symlet.

According to (3), DWT is a multiresolution analysis technique that its concept can be illustrated as in Fig. 1. It is a technique that analyzes the original data in different scales and positions (times). As shown in Fig. 1, the more stretch the wavelet function from the left, the longer the portion of the signal with which it is being compared, and thus the coarser the signal features being measured by the wavelet coefficients [17].

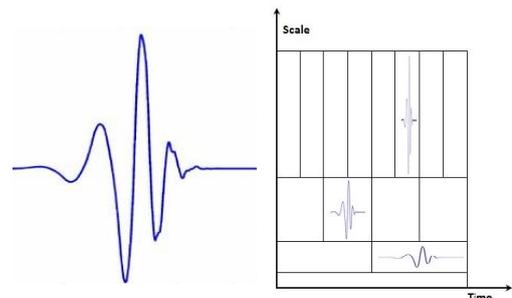


Fig. 1. Multiresolution analysis concept of DWT. Illustrate when mother wavelet function on the left is dilated and translated to fit at different scales and positions (times) represented on the right.

Fig. 2 shows the decomposition of DWT at each level. For level 1, the original data are transformed and decomposed into the coefficient of the approximation part (cA1) and the coefficient of the detail part (cD1). Number of values of the approximation part and the detail part is half of the original data. At level 2, only the cA1 part is further transformed and decomposed into the coefficient of the approximation part (cA2) and the coefficient of the detail part (cD2). If we want to further transform to the next level, only the approximation part of the previous level will be decomposed by the same scheme. As a result, if we want to reconstruct the original data from the transformation coefficient, we can reconstruct by equation 4, which is the case that we transform to level 3. In

practice, we can transform to any level, up to the original data and the applied applications.

$$\text{Original Data} = cA3 + cD3 + cD3 + cD1 \quad (4)$$

We can also reconstruct just the coefficient of the approximation part. That means we need only the main part of the original data by eliminating its noise or its detail part. We then just reconstruct by setting all cDi terms in (4) as zero according to its position in the reconstruction process.

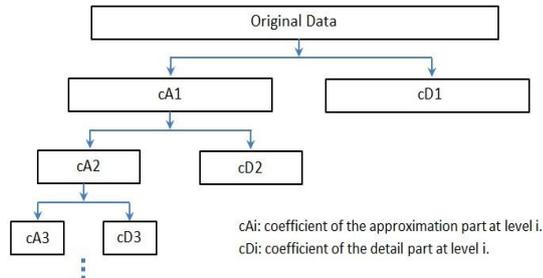


Fig. 2. DWT decomposition at each level. For level 1, the original data is transformed and decomposed into the coefficient of the approximation part ($cA1$) and the coefficient of the detail part ($cD1$). In level 2, only $cA1$ part is further decomposed in to $cA2$ and $cD2$.

C. ARIMA Model

ARIMA model is one of the widely used approach in time series forecasting [18]. The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. Lags of the stationarized series in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series. Random-walk and random-trend models, autoregressive models, and exponential smoothing models are, therefore, all special cases of ARIMA models.

A non-seasonal ARIMA model is classified as an "ARIMA(p, d, q)" model, where:

p is the number of autoregressive terms,

d is the number of non-seasonal differences needed for stationarity, and

q is the number of lagged forecast errors in the prediction equation.

The forecasting equation is constructed as follows. First, let y denote the d^{th} difference of Y , which means:

$$\text{If } d = 0: y_t = Y_t$$

$$\text{If } d = 1: y_t = Y_t - Y_{t-1}$$

$$\text{If } d = 2: y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

In terms of y , the general forecasting equation is:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (5)$$

Standard ARIMA models are often described by the triple (p, d, q) as described before. These define the structure of the model in terms of auto regressive (AR) order, the degree of differencing order, and moving average (MA) order that models to be used.

IV. PROPOSED METHOD

In this research, we apply multiresolution analysis method

to the time series data based on Deubechies (Db) family of DWT. Many researches reveal that Db wavelet is suitable for non-stationary time series data. There can be many vanishing moment variables in Db. The most commonly used are Db2 to Db10. We used Db6 in our proposed work explained as follows:

A. Process of Model Creating and Validating

- 1) Dividing the original time series data set into three parts: the first 80% of time intervals are for training set, the following 10% are for validating set, and the remaining 10% are for testing set.
- 2) Applying multiresolution analysis by DWT of db6 to the training set up to level 4.
- 3) Reconstructing the approximation part of each level in step 2. Resulting from the reconstruction parts are in rAi , where $1 \leq i \leq 4$ is named for level number of DWT decomposition.
- 4) Modeling ARIMA for each of rAi in step3. Therefore, there are four ARIMA models. The forecasting of each model in this step is evaluated by the validating set. The best validated model, observed from the lowest mean absolute percentage error (MAPE), is selected for further forecasting purpose.

B. Process of Model Testing

In the process of model testing, the best validated model is evaluated on the testing set for comparing a multiresolution-based ARIMA model against the ARIMA forecasting on the original data. The process is as follows:

- 1) Applying i levels of DWT to the training set consecutively with the validating data set, where i came from rAi that was used in the validating step for model selection (step 4 of the model creating and validating process).
- 2) Reconstructing the approximation part of such i level and using that reconstruction values to model by ARIMA. Then, evaluating the forecasting performance of a model using the testing set.
- 3) Modeling ARIMA on the original training data set. This step is for comparing the performance of the proposed method, ARIMA is thus also modeled to the original data of the training set consecutively with the validating set. The evaluating the forecasting performance of a compared model is also done by the same testing set.
- 4) Measuring MAPE, percentage of MAPE improvement, and variance or errors to compare the forecasting performance from step 2 and step 3.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Sets

In this research, we experiment on two of commodity price data sets, which are daily gold price and rubber price in Thailand market. Gold price consists of 888 time intervals, as shown in Fig. 3. Rubber price consists of 258 time intervals, as shown in Fig. 4. All prices are quoted in Thai Baht.

For gold price, we divided the first 712 time intervals for training set (80% of data), the next consecutive 88 time

intervals for validating set (10% of data), and the remaining consecutive 88 time intervals for testing set (10% of data). The same portion is done for another data set. Therefore, for rubber price, 208 time intervals are for training set, the next 25 time intervals are for validating set, and the remaining 25 time intervals are for testing set.

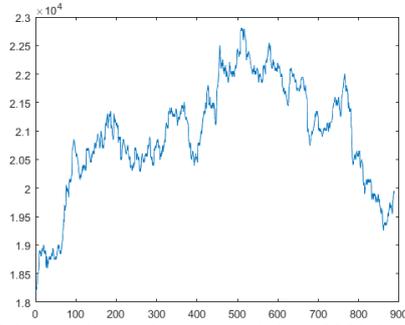


Fig. 3. Original gold price data set consisting of 888 time intervals.

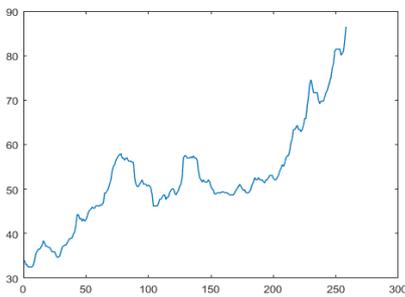


Fig. 4. Original rubber price data set consisting of 258 time intervals.

B. Experimental Results

Following our proposed method mentioned in Section IV, Fig. 5 shows plotting results of the reconstruction values in the process of model creating and validating after the training set of gold price data set is transformed by Db6 DWT up to level 4. Therefore, only the first 712 time intervals of data in Fig. 3 are used by such transformation. By our proposed method, only the coefficient of the approximation part of each level is used for reconstruction. Fig. 5(a) shows the reconstruction values from level 1 (rA1) and Fig. 5(b) is the reconstruction values from level 2 (rA2). The bottom row shows the reconstruction values from level 3 (rA3) in Fig. 5(c) and level 4 on (rA4) in Fig. 5(d), respectively.

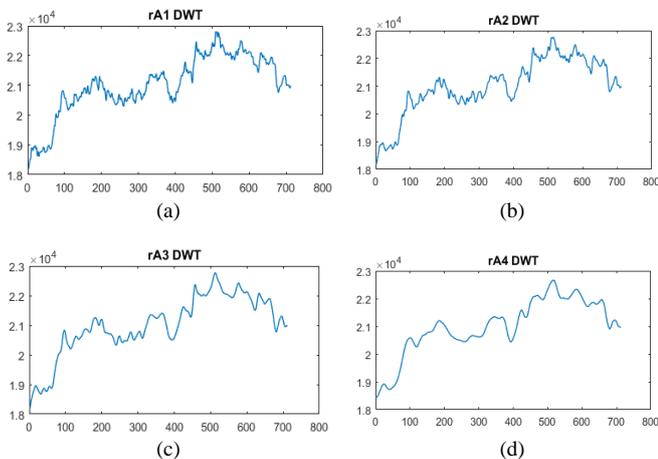


Fig. 5. Plot the reconstruction values from the approximation part of DWT at each level from level 1 to level 4 of the training set. Term rA1 in (a) is named for the reconstruction from the approximation part of level 1, named rA2 in (b) is for level 2, respectively.

Each rAi DWT data in Fig. 5 is then used to model by ARIMA and the forecasting results is evaluated by the validating set. Table I shows MAPE results when evaluating each model by the validating set. There are MAPE of all four models from multiresolution analysis by DWT and one model of the original data that are compared for both data sets. It reveals that the best validating model for gold price is rA3 of DWT with 1.7724 MAPE. For rubber price, the best validating model is rA4 of DWT with 5.1841 MAPE.

TABLE I: ERRORS (MAPE) FROM THE PROPOSED METHOD FOR FORECASTING THE VALIDATING SET BY ARIMA MODEL

Model built from Deubechies wavelet	Gold Price	Rubber Price
rA1 DWT	1.8757	6.8190
rA2 DWT	2.1163	8.1189
rA3 DWT	1.7724	7.1092
rA4 DWT	4.5257	5.1841
Original Data	1.9039	7.6979

TABLE II: THE RESULTS WHEN EVALUATED THE BEST MODEL (FROM TABLE I) ON TESTING SET COMPARED TO THE ORIGINAL DATA

Model	MAPE	Improvement on MAPE	Improvement on Variance of Errors
<i>Gold Price</i>			
rA3 DWT	3.6103	10.83%	26.78%
Original Data	4.0491	NA	NA
<i>Rubber Price</i>			
rA4 DWT	3.7819	42.86%	62.16%
Original Data	6.6189	NA	NA

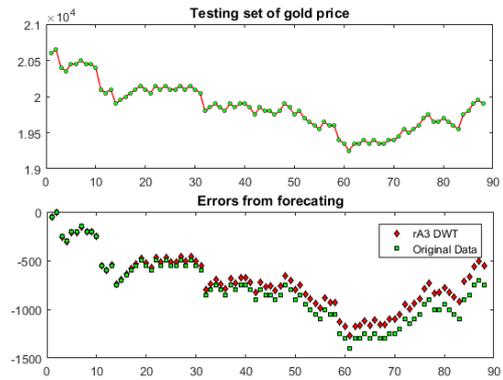


Fig. 6. Gold price forecasting over the 88 time intervals in the testing data set (above) and the forecasting errors comparing to the original data (below).

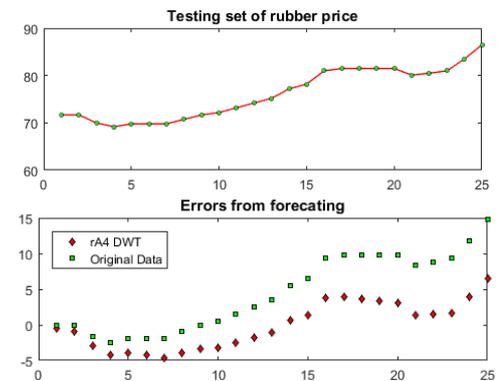


Fig. 7. Rubber price forecasting over the 25 time intervals in the testing data set (above) and the forecasting errors comparing to the original data (below).

The best validating model for each data set in Table I is then used to model and forecast in the process of model testing. Table II show the results when evaluating the best validating model by the testing set of data. For gold price, the

testing set of data consists of 88 time intervals used for forecasting; results are shown in Fig. 6. Model named rA3 DWT from the proposed method which is the best validating model from Table I is evaluated on this testing set. We can see that the model rA3 DWT is better than a model from the original data with 10.83% of MAPE improvement. Fig. 6 also shows errors from forecasting of each time step from both models. When consider such errors distribution, variance of error from the proposed rA3 DWT model is better than the model from the original data with 26.78% improvement.

For the model testing of rubber price data set, the proposed model named rA4 DWT from Table I which is the best validating model is evaluated on the testing set; the results are shown in Fig 7. MAPE results are also compared to a model from the original data set. Percent of MAPE improvement by the proposed rA4 DWT model is 42.86% improvement compared to a model from the original data. Variance of forecasting errors from the rA4 DWT model is better than the model from the original data with 62.16% improvement.

In our experiment, we implement DWT by Matlab and implement ARIMA in R Package. The orders of p , d , and q in ARIMA come from *auto.arima* function in such package. The package uses a combination of unit root tests, minimization of the AIC, and MLE to obtain orders of p , d , and q .

C. Discussion

We have found from the experiments that if the original data are firstly transformed by the multiresolution analysis method based on Deubechies DWT before creating the forecasting model, the forecasting performance are much better than a direct use of the untransformed data, especially for the rubber price data set. Our proposed method achieves good performance because DWT can eliminate some noise or the undesired part from the original data. It, however, depends on the nature of each data set that what is the undesired part of such data. We, thus, design our method to select the suitable level of DWT for each data set by firstly creating and validating the model. This is based on the assumption that the nature of data in the testing set should be the same as the training and the validating sets.

The experimental results on the two of commodity price data sets reveal that our assumption is precise. For gold price data set, the suitable level of DWT is 3. Thus, the main part of data after filtering out the undesired part is the reconstruction values from the approximation part at the third level (rA3). Using only this part to create the forecasting model is much better than modeling directly on the original data. For rubber price data, the best transformation by DWT is at level 4. The performance of rubber price forecasting is very satisfactory from the 42.86% of MAPE improvement and 62.16% improvement on reducing variance of errors.

VI. CONCLUSION

We proposed multiresolution analysis method to the time series data forecasting based on Deubechies family of DWT. The proposed method consists of two main processes which are process of model creating and validating and process of model testing. The best validating model from the first process is selected and used for creating the model in the next

process and forecasting the testing set of data. Experiments are conducted on two of commodity price data sets which are daily gold price and rubber price of Thailand market. MAPE results from using multiresolution analysis are better for both data sets. The improvement is 10.83% for gold price and 42.86% for rubber price. The variances of errors from the proposed method of both data sets are also much less than the direct use of the original time series data for forecasting.

REFERENCES

- [1] P. Chaovalit, A. Gangopadhyay, G. Karabatis, and Z. Chen, "Discrete wavelet transform based time series analysis and mining," *ACM Computing Surveys*, vol. 43, no. 2, pp. 1-37, January 2011.
- [2] H. Yoon, K. Yang, and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series," *IEEE Trans on Know Data Eng*, vol. 17, no. 9, pp. 1186-1198, 2005.
- [3] U. Acharya, S. Sree, A. Alvin, R. Yanti, and J. Suri, "Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals," *International Journal of Neural Systems*, vol. 22, no. 2, April 2012.
- [4] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," in *Proc. 4th Int Conf on Foundations of Data Organization and Algorithms*, pp.69-84, 1993.
- [5] T. Li, Q. Li, S. Zhu, and M. Ogihara, "A survey on wavelet applications in data mining," *ACM SIGKDD Explorations Newsletter*, vol. 4, no. 2, pp. 49-68, 2002.
- [6] K. Kishore and P. Satish, "Emotion recognition in speech using MFCC and wavelet features," in *Proc. 3rd IEEE Int Advance Computing Conference*, 2013, pp. 842-847.
- [7] C. Lin, S. Chen, and T. Truong, "Audio classification and categorization based on wavelets and support vector Machine," *IEEE Trans on Speech and Audi Processing*, vol. 13, no. 5, pp. 644-651, 2005.
- [8] S. Gai, G. Yang, and S. Zhang, "Multiscale texture classification using reduced quaternion wavelet transform," *Int J Electronics and Communications*, vol. 67, no. 3, pp. 233-241, March 2013..
- [9] T. Gandhia, B. Panigrahib, and S. Ananda, "A comparative study of wavelet families for EEG signal classification," *Neurocomputing*, vol. 74, no.17, pp. 3051-3057, October 2011.
- [10] M. Madaleno and C. Pinho, "International stock market indices comovements: a new look," *Int J Finance & Economics*, vol. 17, pp. 89-102, 2012.
- [11] B. Kim, T. McMurry, W. Zhao, R. Wu, and A. Berg, "Wavelet-based functional clustering for patterns of high-dimensional dynamic gene expression," *J Comp Biology*, vol. 17, no. 8, pp. 1067-1080, 2010.
- [12] R. Jammazi and C. Aloui, "Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling," *Energy Economics*, vol. 34, no. 3, pp. 828-841, May 2012.
- [13] E. G. de Souza e Silva, L. Legey, and E. A. de Souza e Silva, "Forecasting oil price trends using wavelets and hidden Markov models," *Energy Economics*, vol. 32, no. 6, pp. 1507-1519, 2010.
- [14] J. Jin and J. Kim. (2015). Forecasting natural gas prices using wavelets, time series, and artificial neural networks. *PLOS ONE*. [Online]. 10(11). Available: <https://doi.org/10.1371/journal.pone.0142064>
- [15] T. Kriechbaumer, A. Angus, D. Parsonsa, and M. R. Casadoa, "An improved wavelet-ARIMA approach for forecasting metal price," *Resources Policy*, vol. 39, pp. 32-41, March 2014.
- [16] I. Daubechies, *Ten Lectures on Wavelets*, SIAM Press, 1992.
- [17] R. C. Singh, "Wavelet transforms in time series analysis," *J of Istanbul Aydın Üniversitesi Dergisi.*, vol. 3, no. 9, pp. 61-94, 2013.
- [18] A. Dimitros and S. G. Hall, "ARIMA models and the Box-Jenkins methodology," in *Applied Econometrics*, 2nd ed, Palgrave MacMillan, 2011, pp. 265-286.



S. Bunrit is a lecturer with computer engineering school, SUT. She received her bachelor degree in Science (Mathematics) from Kasetsart University, Thailand, in 1997, master degree in Science (Computer Science) from Chulalongkorn University, Thailand, in 2001. Her research of interest includes artificial neural network, deep learning, machine learning, digital image processing, computer vision, and time series analysis.



N. Kerdprasop is an associate professor with computer engineering school, SUT. She received her bachelor degree in Radiation Techniques from Mahidol University, Thailand, in 1985, master degree in Computer Science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in Computer Science from Nova Southeastern University, U.S.A, in 1999. Her research of interest includes data mining, artificial intelligence, and

intelligent databases.



K. Kerdprasop is an associate professor and chair of the School of Computer Engineering, SUT. He received his bachelor degree in Mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in Computer Science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in Computer Science from Nova Southeastern University, U.S.A., in 1999. His current research includes data mining, artificial intelligence,

computational statistics.