**RESEARCH ARTICLE**

# Forecasting of crude palm oil price using hybridizing wavelet and group method of data handling model

Huma Basheer , Azme Khamis[*]

*Department of Mathematics and Statistics, Faculty of Science, Technology and Human Development, Universiti Tun Hussein Onn Malaysia (UTHM), 86400 Parit Raja, Batu Pahat, Johor Bahru, Malaysia*

* Corresponding author: azme@uthm.edu.my

**Abstract**

Forecasting of Crude Palm Oil (CPO) is one of the most important and the largest vegetable oil traded in the world market. This study investigates the forecasting of Crude Palm Oil (CPO) price using a hybrid model of Group Method of Data Handling (GMDH) with wavelet decomposition. The original monthly time series data of CPO price were decomposed into the spectral band using discrete wavelet transform (DWT). After decomposition of the original time series data into subseries, the new series which is obtained acquired by adding all the effective components of the spectral band were given as an input time series data to GMDH model to forecast the CPO price of monthly time series data. The results performance of hybridized GMDH model is compared with the original GMDH model. The measurements results from the important statistical measures of errors such as mean absolute error (MAE) and the root mean square error (RMSE) showed that the hybrid GMDH model with wavelet decomposition gives more accurate result of predictions compared with the original GMDH model. The proposed combined wavelet-GMDH model verifies to be one of the best feasible results considering the forecasting issues of the time series data.

*Keywords*: Crude Palm Oil (CPO), GMDH, discrete wavelet decomposition, hybrid model, forecasting

## INTRODUCTION

Palm oil is one of the most important and the largest vegetable oil traded in the world market (Rasillo-Calle, Pelkmans, and Walter 2009). Palm oil is also considered "substitute goods" of any other vegetable oil because due to the price fluctuation, the food processor usually switch to the different vegetable oil (Thani 2013). Palm oil have different used, for example in cooking palm oil were used in different food items, useful chemicals, like bio-diesel, soaps production, products of personal care and bio-fuels (Sime Darby Plantation 2014).

The Palm oil has used in greater quantity as compared with any other vegetable oil (Mohammadi et al. 2015). Referring to the United States Department of Agriculture (USDA) report ((USDA) 2015) presents that the palm oil production is about 36,0% of the world vegetable oil production, while the other 8 vegetable oils the remaining 56,0% (Abdullah 2011). According to the data source provided by Malaysia Palm Oil Board (MPOB), Oil World Jan 2012-2013 report (MPOB 2014) for world palm oil distribution and supply, the largest country is Indonesia and the second largest country for palm oil producer is Malaysia, as shown in Fig 1.

The values of future markets rise from the understanding for the spot price pridiction are significant for managing the risks related with the trading a goods (MPOB 2014). Therefore, the precise and accurate prediction values are needed to assist the decision making process of the trader. Fig. 2 has been shown the graph for monthly oil prices data, which were taken from Malaysian Palm Oil Board (MPOB) into consideration from 1983 to 2014. The graph shows the highest price of major vegetable oils is
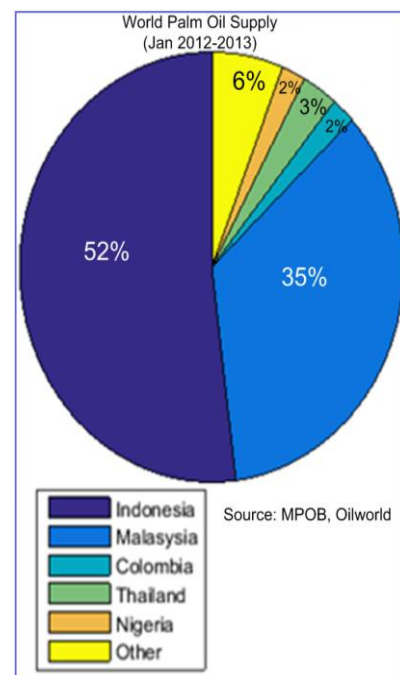


**Fig. 1** World palm oil supply by country

Coconut Oil (CNO), the medium price is Soybean Oil (SBO), and the lowest price is Crude Palm Oil (CPO). These vegetables oils are commerced worldwide among a huge number of producer and consumer. This study only target the CPO among all of these commodities.

The spot prices predictions for future values are very important for different trading commodities (Abdullah 2011). Therefore, to assist and enhance the decision making process, we need the prediction of precise and accurate values of future markets. In this paper, we used hybrid wavelet-GMDH model to forecast the price of

CPO values precisely and accurately. The time series data were first decomposed into several simple time series using a discrete wavelet transform (DWT). Then, a GMDH model was used as a predictor method. The measurement results from the mean absolute error (MAE) and the root mean square error (RMSE) showed that the hybrid GMDH model with wavelet decomposition gives more accurate result for time series forecasting of CPO prices as compared with the original GMDH.
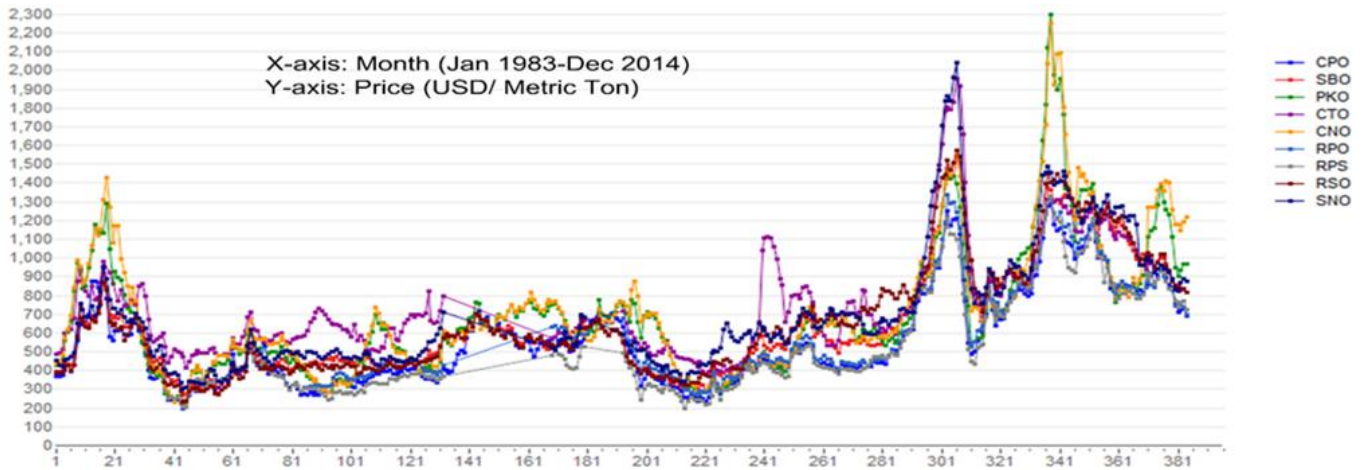


**Fig. 2** Price of major vegetables oils including crude palm oil (average per year1983-2014)

## GROUP METHOD OF DATA HANDLING (GMDH)

In modelling techniques, the algorithm of GMDH works and organized as a heuristic and computer oriented method which learn the relationship among the variables. In the aspect of time series forecasting, the algorithm of GMDH type neural network works as a lags, which determine the relationship among the variables. After learning the relationship among the variables, the GMDH type neural network automatically selects the way to follow in algorithm. To construct a high order polynomial, Ivakhnenko (1968) was used the first GMDH model (Anastasakis et al. 2001; Mottaghitalb 1996; Park and Pedrycz 2007). The Ivakhnenko polynomial is given by the following Eq. (1).

$$y = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} x_i x_j + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk} x_i x_j x_k + \ldots\ldots\ldots(1)$$

Here $M$ represents number of input (variables) and '$a$' represents the coefficients (weights) of the variables. Here '$y$' represent response variable and '$x_i$' and '$x_j$' represent the lagged time series data, which is to be regressed. Useually Eq. 1. can be written up to the square term as shown below.

$$y = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} x_i x_j \tag{2}$$

The algorithm of GMDH study all pairwise combination of $n$ lagged time series. Therefore, each combination enters    each neuron. It means that one result goes out as an output, when two input variables go in a neuron. For m = 2, the Ivakhnenko polynomial in equation 2 is given by the following equation (G.A.Ivakhnenko 1995).

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{3}$$

The GMDH type algorithm is consisting of layers and in these layers exist neurons. The neurons number in a single layer is determine by the number of input variables and is given by

$$T = {}^n C_2 = \frac{n!}{(n-2)!2!} \tag{4}$$

Where $T$ present the number of neurons, n represents number of input variable. The GMDH algorithm architecture is shown in Fig. 3. As there are four input variables in GMDH Algorithms Architecture, so the number of nodes in a starting layer to the algorithm is found to be six. In Equation 3, the coefficients are estimated in each neuron. The estimated coefficient and input variables in each neuron, are used for the predicted desired output.
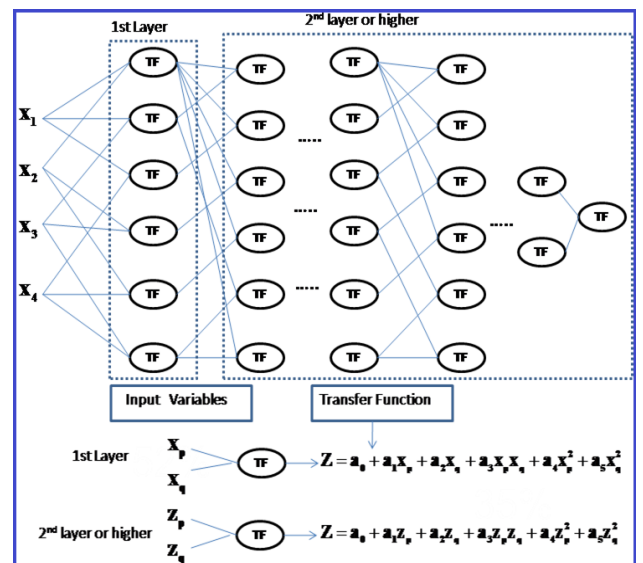


**Fig. 3** GMDH algorithms architecture

As using external criteria, *n* neurons are preferred and *T- n* neurons are excluded from the given network. The pridiction mean square error used in this study as an external criteria (Dag and Yozgatligil 2016). In the GMDH algorithms architecture shown in Fig. 3, from each layer four neuron are chosen and two neurons are excluded from the network. The outputs achieved from the chosen neurons turn into the inputs for the next layer. This progress continues until one neuron is chosen in the last layer, which is the pridicted value for the given time series data. The GMDH algorithms flow chart is shown in Fig. 4.
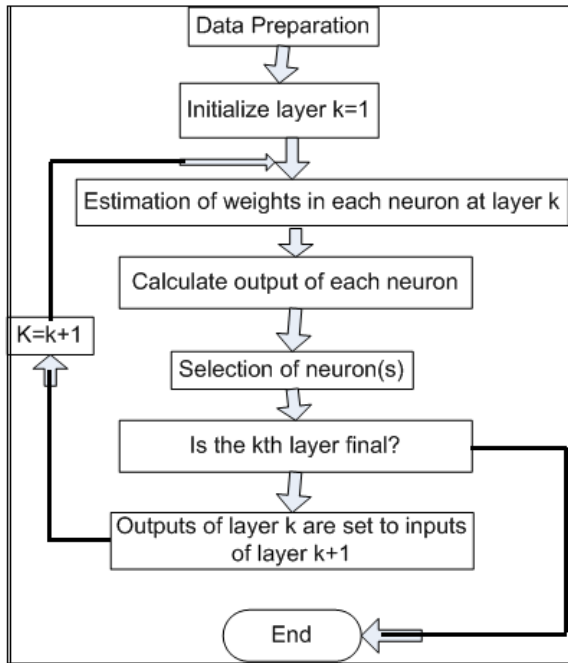


**Fig. 4** GMDH algorithms flow chart

## DISCRETE WAVELET TRANSFORM (DWT)

Wavelet transforms describes a mathematical tool, which shows strong representation of a time-frequency in the time domain for the analyzed signal (Danandeh Mehr, Kahya, and Olyaie 2013; Dash et al. 2007; Okkan and Ali Serbes 2013). Wavelet transformation produce effective disintegration of the actual time series data by taking useful information on different decoposition levels. The wavelet are selected based on their shape and their understanding to study the time series data in a special application. The wavelet families have different types, which can be used as a mother wavelet such as Daubechies, Haar, Symlet, Coiflet, Morlet and Mayer (Strange 1996). Wavelet transform can be divided either in a continuous wavelet transform (CWT) or discrete wavelet transform (DWT). The CWT is not often used for forecasting due to its computationally complex, where as DWT requires less computation time and is often used in forecasting applications (Shabri and Samsudin 2014). The variable which is decomposed by DWT is used Mallat discrete wavelet tranform algorithm (S. G. Mallat 1989). The Mallat algorithm used each data during decomposition process in such a way that no information is lost in any part of the original data.

The descrete wavelet transform decomposition process is a subsequent filtering process of the original input data into an approximate (As) and detail coefficients (Ds). The level of decomposition depends on the length of the data set (the number of values). The following formula is used for the level of decomposition number (Nourani, Alami, and Aminfar 2009).

$$L \; = \; \mathrm{int}[\log(N)] \qquad (5)$$

In this formula N is the number of values of time series data and L is decomposition level. Using the formula in equation (5), the decomposition levels for monthly crude palm oil (CPO) Price series data in this study is 3. The output of the discrete wavelet transform decomposition tree is shown in Fig. 5. The effective component of the output of DWT are chosen using correlation analysis.
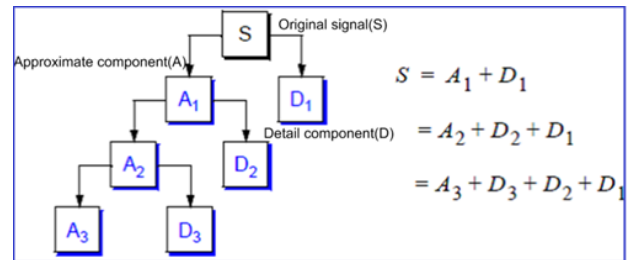


**Fig. 5** Discrete wavelet transform decomposition tree

.

### Hybridizing Wavelet and GMDH Model

The hybridizing wavelet-GMDH model is the enhancement of conventional GMDH model by combining of DWT method and GMDH model. In DWT, the original crude palm oil price series data was disintegate into 3 level, details components namely D1, D2 and D3 and approximate component (A3). Each of detail series ( D's) shows a specific role in the original time series and has different effects on the original crude palm oil price series . The effective components of DWT are selected based on the correlation analysis and used as a input GMDH model to improve the forcasting performance of the hybridizing wavelet-GMDH model (Badyalina 2014). The selection of the effective components of DWT based on the correlation analalysis. This also allows the GMDH to determine about the characteristics of the time series data and produce good estimation.

The main significance of using DWT is the smooth analysis of the crude palm oil price series, which is obtained after wavelet decomposition. Wavelet transform decompose complex data into a different number of series component signals in such a way that the error criterion is minimized (Basheer and Khamis 2016) . The wavelet is also used as a function of filter, which remove the noice from the original raw data and therefore it makes easy for GMDH model to analysis the pure and clean signals. As it is reported by (Cannas, Fanni, and See 2006), the result for which the network is trained with the preprocessed data achieved better result than the networks which is trained for raw signals on noisy data. This study used Daubechies family wavelet which had been performed on the DWT. The impotant of Daubechies family wavelet approximation is the accurately decomposition of of raw signals.

## RESEARCH METHODOLOGY

The Summary of research frame work and research methodology used in this study consists of the following steps and flowchart as shown in Fig. 6.

1. Secondary Data of oil palm yield were taken from Malaysia Palm Oil Board (MPOB) ranging from Jan, 1983 to Dec, 2015.
2. Determine the number of inputs.
3. Split the data into training and testing.
4. Performed a simulation experiment in the conventional GMDH.
5. Decompose the input variables into the various corresponding detail coefficients (Ds) approximation coefficients (As) with the discrete wavelet transform (DWT) using mat-lab simulation tool.
6. Mallat DWT algorithm was used to decompose the input variables into the DWT (S. G. Mallat 1989).

7. Using the co-efficient of determination ($R^2$) for component identification to find the most effective components of detail coefficients (Ds) and construct new input for hybrid GMDH with discrete wavelet decomposition as shown in Fig. 7.
8. Use RMSE and MAE to find the most useful model acquired through the original GMDH and combined wavelet-GMDH.
9. The original GMDH algorithm performance in all measures of simulative analysis compared to hybrid GMDH with wavelet decomposition.
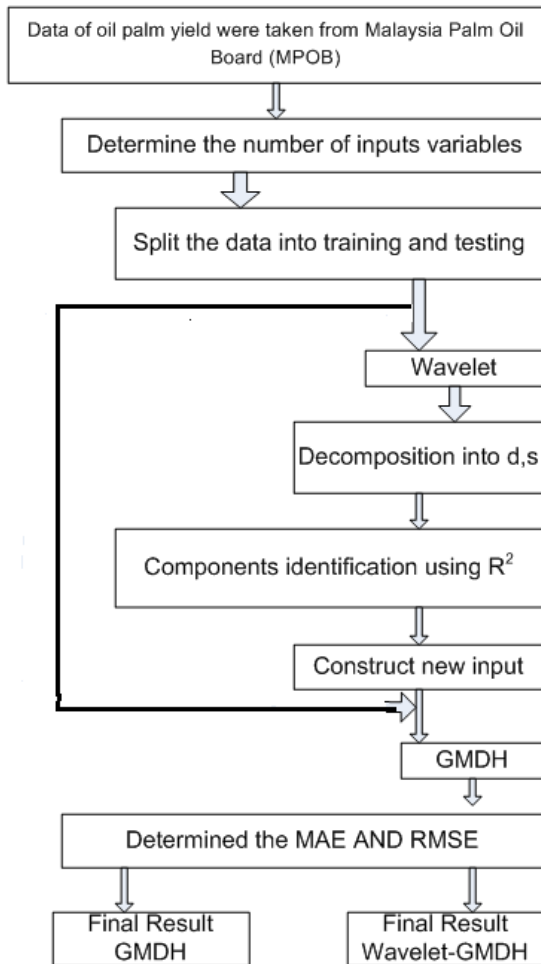
Where $y_t$ the forecasted value at data is point and $\hat{y}_t$ is the mean of the forecasted values at the time t, respectively and $N$ is represent the number of data points.
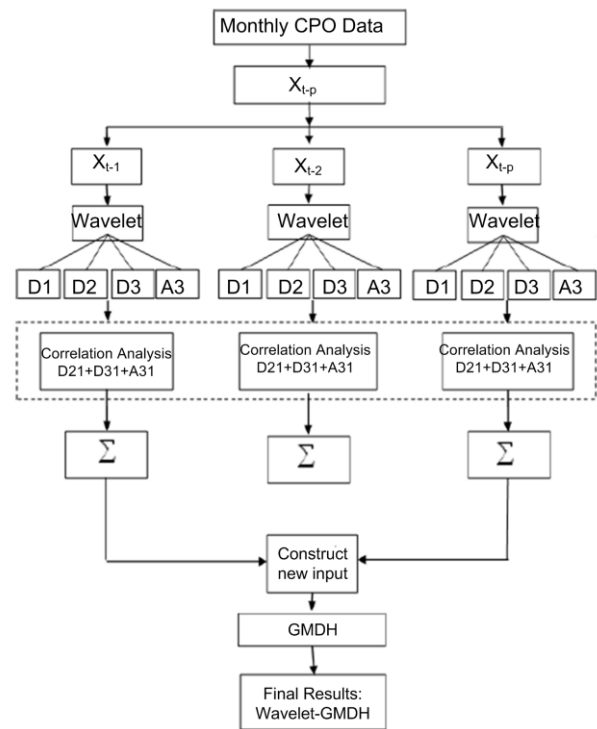


**Fig. 7.** The framework structure for wavelet-GMDH



**Fig. 6** The study framework

### EVALUATION CRITERIA Software for Simulation Result

To calculate the performance of prediction of each model, we considered the important statistical measures of errors such as the MAE and RMSE, which are generally used for calculating the forecasting results of time series data. The important statistical measures of errors are given by the following formulas.

$$\mathbf{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t| \quad (5)$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \quad (6)$$

### SOFTWARE FOR SIMULATION RESULTS

The mat lab simulation tool was used for the wavelet decomposition time series data and the GMDH Shell (GMDH Shell DS 3.8.1) software was used for time series forecasting using Group Method of Data Handling (GMDH) algorithm. Learning algorithm is a method which controls the optimization process of modeling. In the current theory, which is used for the predictive modeling it is widely known that each model should present a trade-off between the accuracy and simplicity (Ratrout 2014). A.G.Ivakhnenko in 1968 discovered the knowledge discovery approach, which is based on the learning algorithm of GMDH. GMDH Shell also uses the learning algorithm of GMDH, which is based on the knowledge discovery approach of A.G.Ivakhnenko (1968). GMDH Shell is used as a tool for real-world data sets and targets different topics like knowledge discovery, predictive analytics, and forecasting software of time series, classification, regression, data-mining tools, and prediction (Shinohara, Dohi, and Osaki 1999).

### RESULTS AND DISCUSSION

The monthly CPO price (in USD) data have been used in this study and are taken from Malaysian Palm Oil Board (MPOB) from 1983 to 2015. MPOB is a government organization, which is responsible for the regulation, promotiom and development of the palm oil industry in Malaysia. The CPO data was obtained from January 1983 to May 2015. Total data for each of commodity price are 389. Given a set of 389 observation made at uniformly spaced time intervals, the location of CPO is rescaled to the time axis become the set of {1, 2…389}. For example the first location in 1983 is written as time 1, the second location in 1983 to 2 and so on. The time series data of crude palm oil divided into two parts namly one is training part, which observed the ability of methods on prediction (n =

384) and the other is testing part, which observed the forecasting (n = 5). The time series plot of CPO (Crude Palm Oil) after simulation results include predictions, model plots, and accuracy metrics is given in the following.

Fig. 8 has been shown the time series plot of monthly CPO prices for actual and predicted values using conventional GMDH model. The actual data, model fitted data and pridicted values are clearly shown in the Figure. The forecasting performance of statistical measures of errors such as MAE and RMSE calculated and were shown in Table 1.

Fig. 9 has been shown the discrete wavelet decomposition of original signal into details (Ds) and approximation (As) components using level 3 db. Using the co-efficient of determination ($R^2$) for component identification, to find the most effective components of detail coefficients (Ds) and constructed new input, which for hybrid GMDH with discrete wavelet decomposition as shown in Fig. 10.

Fig. 11 has been shown the time series plot of monthly CPO prices with predictions and forecasts obtained via Wavelet-GMDH. The actual data, model fitted data and pridicted values are clearly shown in the Figure, which showed very good fitting result as compared to conventional GMDH. The forecasting performance of statistical measures of errors such as MAE and RMSE calculated and were shown in Table 2.

Table 3 has been shown the comparison of the GMDH algorithms with Hybrid wavelet-GMDH model on crude palm oil prices in terms of Prediction and forecasting values of MAE and RMSE respectively. The wavelet-GMDH algorithm has been shown the best results about the forecasting performance and its prediction accuracy as compared to the conventional GMDH model.
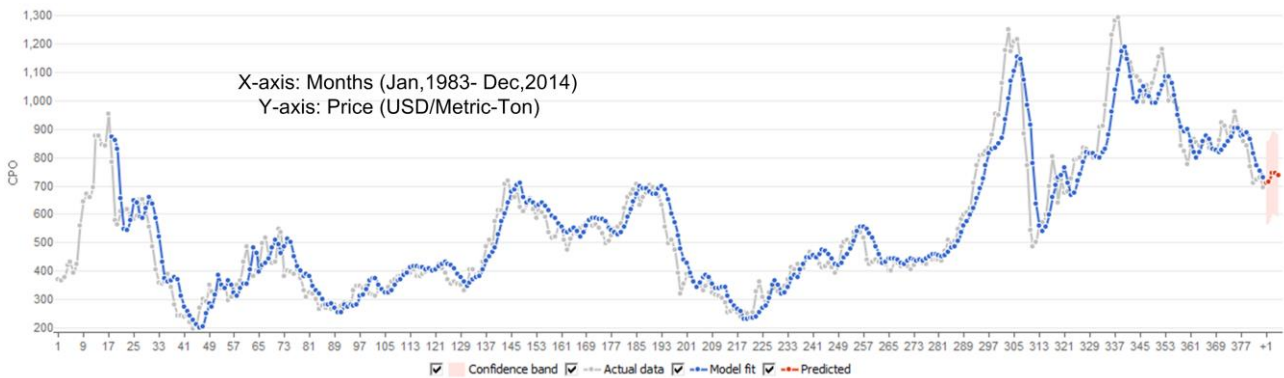


**Fig. 8**. The time series plot of monthly CPO prices with predictions and forecasts obtained via GMDH
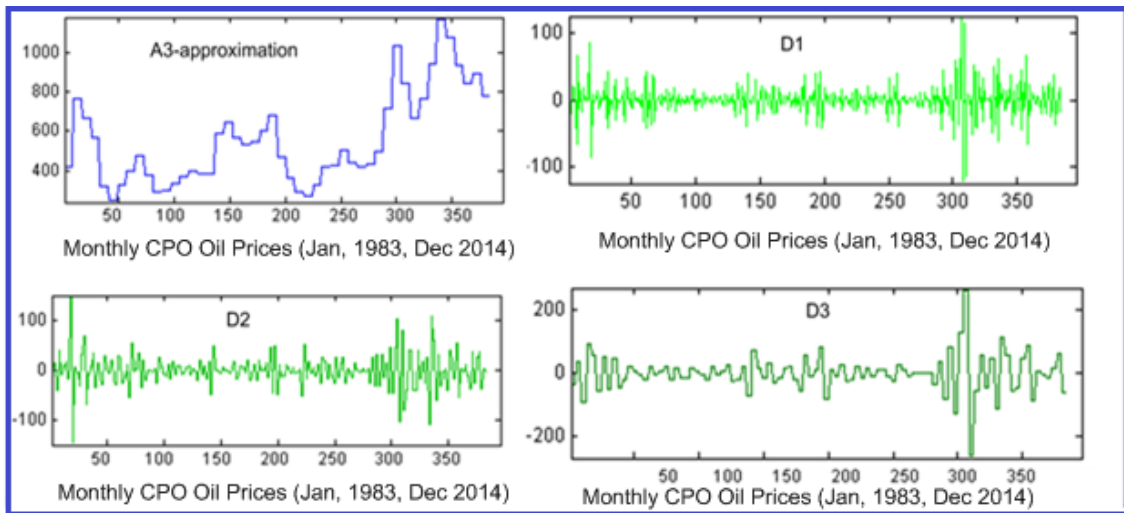


**Fig. 9**. Discrete wavelet decomposition of original signal into details (Ds) and approximation (As) Components using level 3 db
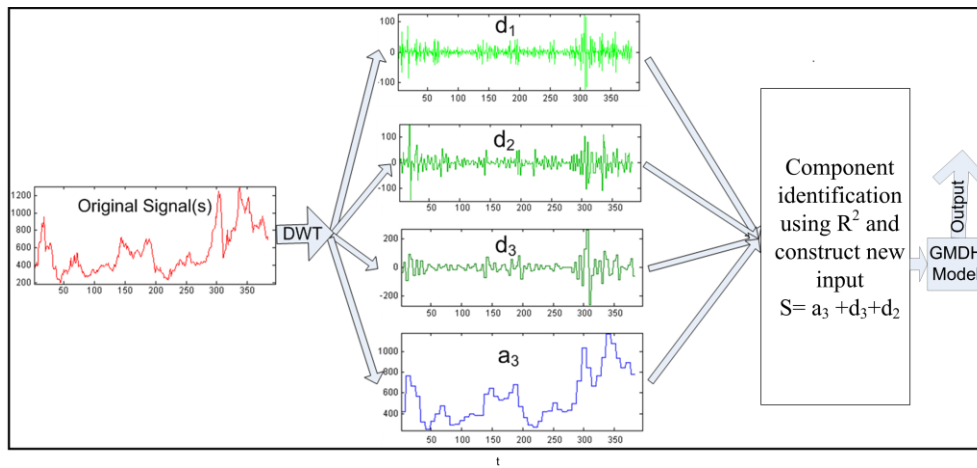
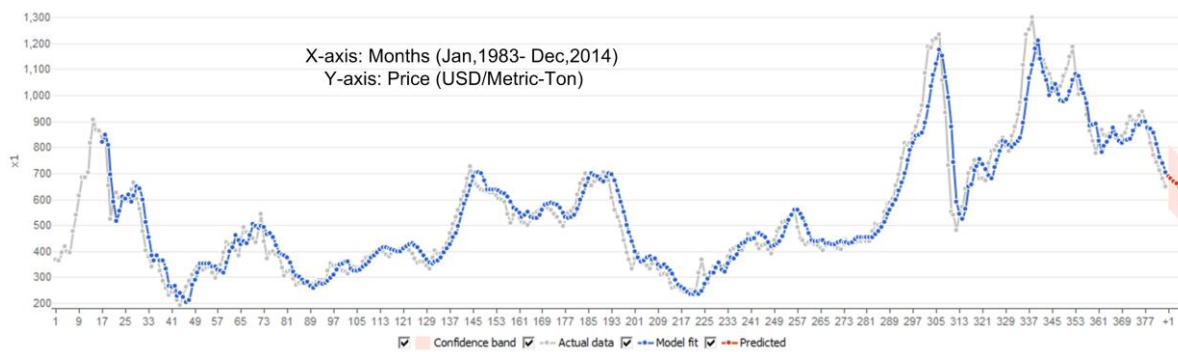**Fig. 10** Hybrid wavelet- GMDH model. Inside Discrete wavelet decomposition of original signal using level 3 db.



**Fig. 11** The time series plot of monthly CPO prices with predictions and forecasts obtained via Wavelet-GMDH

**Table 1** Forecasting performance indices of Conventional GMDH

|  | Actual Testing Data (A) | Forecasting Data by Conventional GMDH Model (F) | Error (A-F) | Square[Error(A-F)] |  |
|---|---|---|---|---|---|
| 1 | 653.75 | 711 | -57.25 | 3277.5625 |  |
| 2 | 655.4 | 715 | -59.6 | 3552.16 |  |
| 3 | 630 | 745 | -115 | 13225 |  |
| 4 | 626.85 | 744 | -117.15 | 13724.1225 |  |
| 5 | 631.25 | 735 | -103.75 | 10764.0625 |  |
| **MAE** |  |  | **90.55** |  |  |
| **RMSE** |  |  | **94.38528222** |  |  |

**Table 2** Forecasting performance indices of Hybrid Wavelet-GMDH Model

|  | Actual(A) Testing Data | Forecasting Data by Hybrid Wavelet-GMDH Model | Error (A-F) | Square[Error(A-F)] |  |
|---|---|---|---|---|---|
| 1 | 653.75 | 688 | -34.25 | 1173.063 |  |
| 2 | 655.4 | 681 | -25.6 | 655.36 |  |
| 3 | 630 | 667 | -37 | 1369 |  |
| 4 | 626.85 | 660 | -33.15 | 1098.923 |  |
| 5 | 631.25 | 644 | -12.75 | 162.5625 |  |
| **MAE** |  |  | **28.55** |  |  |
| **RMSE** |  |  | **29.86271** |  |  |

**Table 3** Prediction and Forecasting performance indices of GMDH and Hybrid wavelet-GMDH

| Model | Prediction (MAE) | Prediction (RMSE) | Forecasting (MAE) | Forecasting (RMSE) |
|---|---|---|---|---|
| GMDH | 50.60 | 72.9805 | 90.55 | 94.38 |
| Wavelet-GMDH | 43.56 | 61.7259 | 28.55 | 29.86 |

## CONCLUSION

In conclusion, we have developed a hybrid wavelet-GMDH model by combining the wavelet transform and GMDH model for modeling and forecasting of the CPO price. For this hybridization, first we have used GMDH model to the CPO price without any data preprocessing. Second, we have used discrete wavelet decomposition (DWT) to original crude palm oil data series to decompose into several simple time series. Third, the effective components of wavelet decomposition were used as an input to the GMDH model to forecast the CPO price. The results obtained from the hybridized GMDH were compared with the original GMDH model. The accuracy measurement of hybrid wavelet-GMDH model to forecast the price of CPO values produced better result of forecasting as compared with the original GMDH.

## ACKNOWLEDGEMENT

## REFERENCES

Abdullah, R. 2011. World Palm Oil supply, demand, price and prospects: Focus on Malaysian and Indonesian Palm oil industries. *Oil Palm Industry Economic Journal,* 11(2): 13–25.

Anastasakis, L., L Anastasakis N., Mort, and N Mort. 2001. The development of self-organization techniques in modelling: A review of the group method of data handling (GMDH). *Group* (813). Retrieved from http://eprints.whiterose.ac.uk/83130/

Badyalina, B. 2014. Streamflow estimation at ungauged site using wavelet group method of data handling in Peninsular Malaysia. *International Journal of Mathematical Analysis,* 8(11): 513–24.

Basheer, H., and Khamis, A. 2016. A Hybrid group method of data handling (GMDH) with the wavelet decomposition for time series forecasting. *ARPN Journal of Engineering and Applied Sciences,* 11(18): 10792–10800.

Cannas, B., Alessandra F., and Linda S. 2006. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Physics and Chemistry of the Earth* ,31: 1164–1171.

Dag, O. and Ceylan, Y. 2016. GMDH: An R Package for short term forecasting via GMDH-Type neural network algorithms. *The R Journal,* 8(2016): 379–86.

Danandeh Mehr, Ali, Ercan Kahya, and Ehsan Olyaie. 2013. Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *Journal of Hydrology,* 505: 240–49.

Dash, P. K., Maya Nayak, M. R. Senapati, and I.W.C. Lee. 2007. Mining for similarities in time series data using wavelet-based feature vectors and neural networks. *Engineering Applications of Artificial Intelligence,* 20(2): 185–201.

Ivakhnenko, G.A., and Ivakhnenko, A.G. 1995. The review of problems solvable by algorithms of group method of data handling (GMDH). *Pattern Recognition and Image Analysis,* 5(4): 527–35.

Mohammadi, S., Fatimah, M. A., Bilash, K. B., and Abdulla, I. 2015. System dynamics analysis of the determinants of the malaysian palm oil price. *American Journal of Applied Sciences,* 12(5): 355–62.

Mottaghitalb, M. 1996. Genetic algorithms and group method of data handling-type neural networks applications in poultry science. In Roeva, O. (Ed.) *Real-World Applications of Genetic Algorithms* (pp 219-234). Croatia: In Tech.

Malaysian Palm Oil Board. 2014. Review *Overview of the Malaysian Oil Palm Industry 2014*. Malaysia : MPOB.

Nourani, V., Alami, M. T., and Aminfar, M. H. 2009. A combined neural-wavelet model for prediction of ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence,* 22: 466–72.

Okkan, U., and Serbes, Z. A. 2013. The combined use of wavelet transform and black box models in reservoir inflow modeling. *Journal of Hydrology and Hydromechanics,* 61(2): 112–19.

Park, H. -S., and Pedrycz, W. 2007. Evolutionary design of hybrid self-organizing fuzzy polynomial neural networks with the aid of information granulation. *Expert Systems with Application,* 33: 830–46.

Rasillo-Calle, F., Pelkmans, L., and Walter, A. 2009. A Global overview of vegetable oils, with reference to biodiesel. *IEA Bioenergy, Task 40* (June): 1–89.

Ratrout, N. T. 2014. Short-Term traffic flow prediction using group method data handling (GMDH)-Based Abductive Networks. *Arabian Journal for Science and Engineering* 39(2): 631–46.

Mallat, S. G. 1989. A Theory for multiresolution signal decomposition : The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 11(7): 674–93.

Shabri, Ani, and Ruhaidah Samsudin. 2014. Daily Crude oil price forecasting using hybridizing wavelet and artificial neural network model. *Mathematical Problems in Engineering,* 2014: 1–10.

Shinohara, Yasuhide, Tadashi Dohi, and Shunji Osaki. 1999. Predictive Evaluation for software testing progress via GMDH networks. *Electronics and Communications in Japan,* 82(5): 22–28.

Sime Darby Plantation. 2014. *Palm Oil Facts & Figures*.

Strang, G. and Nguyen, T. 1996. *Wavelet Toolbox*. Wellesley-Cambridge Press.

Mungmee, P., Buaklom, P., Pungprasert, P., Sinlapaprechar, P., Suksa-ard, C., and Raweewan, M. 2013. A framework study of factors effecting crude palm oil price behavior in Thailand Morrakot Raweewan. *Proceedings of the 4th International Conference on Engineering, Project, and Production Management (EPPM 2013)*, 23-25 October 2013. Bangkok, Thailand, 647–53.

United States Department of Agriculture (USDA). 2015. Report *ISSN: 1949-1522 Released June 30, 2015, by the National Agricultural Statistics Service (NASS), Agricultural Statistics Board, United States Department of Agriculture (USDA).*