

A hybrid GA-FCEEMD for forecasting natural gas demand

Nuramirah Akrom^a and Zuhaimy Ismail ^{*b}

^{a,b} Department of Mathematical Sciences, Faculty of Sciences,
Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia

*Corresponding author Email: *zuhaimyutm@gmail.com

April 2, 2018

Abstract

Forecasting natural gas demand is important for the distribution of the natural gas network. In this paper, a hybrid of Genetic Algorithm and Fast Cardinal Ensemble Empirical Mode Decomposition (GA-FCEEMD) was proposed in forecasting natural gas demand. This hybrid GA-FCEEMD model can effectively handle uncertainty, fluctuation, and volatility patterns of natural gas demand. Monthly natural gas demand data from January 2002 to November 2016 were employed to measure the forecasting performance of the proposed hybrid GA-FCEEMD model. The result revealed that in comparison to ARIMA model, ANN model, and GARCH model, GA-FCEEMD provided the highest forecasting accuracy in predicting natural gas demand.

Key Words:Hybridization; Genetic Algorithm (GA); Fast Cardinal Ensemble Empirical Mode Decomposition (FCEEMD); Natural Gas Demand; Forecasting

1 Introduction

Among the essential energy sources in the world, natural gas stands out as a clean and safe fuel (1). The applications of natural gas are

extensive, such as for transportation, power generation, and residential use(2). In the natural gas supply and demand system, the main objective is the sufficiency of the amount of natural gas from the source to the natural gas network, and it often requires the forecasting of future natural gas demand(3). Therefore, accurate and reliable forecasts of natural gas demand are important for economical and reliable operations of the distribution of natural gas network.

Different methods of forecasting have been developed to fill the gap in this field. For example, Bai & Li (1) proposed a structure-calibrated support vector regression (SC-SVR) approach to forecast the daily natural gas consumption; Zhang and Yang(2) suggested Bayesian Model Averaging to forecast Chinas natural gas consumption; Zeng and Li, (3) used a self-adapting intelligent grey model to predict natural gas demand in China; Potonik et al., (4) compared static and adaptive models for short-term residential natural gas forecasting in Croatia; and Panapakidis and Dagoumas (4) combined Wavelet Transform (WT), Genetic Algorithm (GA), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Feed-Forward Neural Network (FFNN) to forecast day-ahead natural gas demand.

Most of the aforementioned methods proposed by previous scholars involve a lot of steps and are time consuming, even if they do give good forecasting performance (5). To minimize the forecasting steps as well as the time used, we propose a new hybrid forecasting model of Genetic Algorithm and Fast Cardinal Ensemble Empirical Mode Decomposition (GA-FCEEMD). To forecast natural gas demand using this GA-FCEEMD model, only three forecasting steps are involved; first, the natural gas demands actual data are decomposed using FCEEMD algorithm into two Intrinsic Mode Functions (IMFs). Then, the second IMF is used as the intrinsic prediction trend of natural gas demand. Lastly, GA is integrated to obtain the minimum stopping criterion and free curve for FCEEMD and the Mean Absolute Percentage Error (MAPE) for the forecasting accuracy measurement.

This study is expected to help in forecasting natural gas demand for planning and distributing the supply to the natural gas network. We also propose minimal steps in forecasting that can be applied and implemented (6).

The rest of this paper is as follows: section 2 introduces the

forecast methodologies which are ARIMA, ANN, GARCH, and GA-FCEEMD, section 3 presents the natural gas demand data series and empirical results, and section 4 concludes the study.

2 Methodology

2.1 ARIMA Model

In ARIMA model, the future value of a variable is assumed to be a linear function of several past observations and random errors (7):

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + L + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - L - \theta_q \epsilon_{t-q} \quad (1)$$

where y_t is the actual value and ϵ_t is the random error at time period t . $\theta_i (i = 1, 2, \dots, p)$ and $\theta_j (j = 1, 2, \dots, q)$ are model parameters, where p is the order autoregressive part (AR) and q is the order of the moving average (MA) process.

2.2 Artificial Neural Network Model

Artificial Neural Networks are a powerful, non-parametric tool used for pattern recognition, signal filtering, and interpolation. Each neural network connects a group of input variables $\{I_i\}, i = 1, \dots, k$ with a group of one or more output variables $\{O_j\}, j = 1, \dots, k$ and zero, one, or more so-called hidden layers, as illustrated in Figure 1. Neurons are connected between the layers through connections that are activated by reaching a threshold, because the evaluation of the function of transfer is based on the input parameters (8).

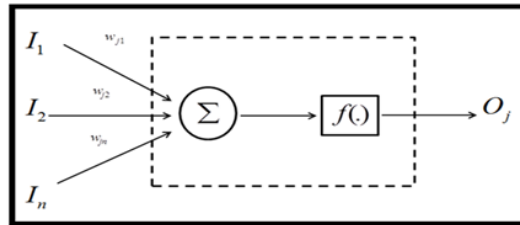


Figure 1. Illustration of a neural network

2.3 GARCH Model

The notation for GARCH (1,1) can be described as follows :

$$r_t = c + \theta_1 r_t - 1 + \epsilon_t \sigma_t \quad (2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (3)$$

where c is a constant, r_t is the actual data series in time t , θ is the parameter for autoregressive model r_{t-1} , ϵ_t is the error term, σ_t is the conditional variance, and α , β are the parameters.

2.4 Fast Cardinal Ensemble Empirical Mode Decomposition (FCEEMD) Model.

FCEEMD model was first popularized by (9). In building the FCEEMD model, there are two conditions which the IMFs have to meet:

- 1) In the whole data set, the number of zero points and the number of extreme points are equal to, or differ at most by, one
- 2) The mean values of the upper and lower envelopes at any point must be zero.

The decomposition process is shown as follows:

- 1) The ensemble number and the amplitude of the added white noise are set.
- 2) A white noise series is added to the targeted data with the set amplitude and the generated signal is denoted as $S(t)$.
- 3) The adjustable cubic trigonometric cardinal spline interpolation (CTCSI) is used to link the local maxima and local minima of $S(t)$ and the upper and lower envelopes are obtained, respectively.
- 4) The mean curve, $m_{11}(t)$, of the upper and lower envelopes are calculated and the difference between $S(t)$ and $m_{11}(t)$ is appointed as $h_{11}(t)$:

$$h_{11}(t) = s(t) - m_{11}(t) \quad (4)$$

If $h_{11}(t)$ meets the two aforementioned conditions, it can be regarded as the first IMF or else it will be seen as an original signal and must undergo the above steps again until $h_{1k}(t)$ satisfies the requirements to becoming an IMF.

Here, $c_1(t) = h_{1k}(t)$.

$$h_{12}(t) = h_{11}(t) - m_{12}(t)$$

$$h_{13}(t) = h_{12}(t) - m_{13}(t)$$

$$h_{1k}(t) = h_{k-1}(t) - m_{1k}(t) \quad (5)$$

Notice that m_{ik} represents the mean curve formed at the k -th step of the subtractions, like in Equation 5, for the purpose of gaining the i -th IMF (c_i). Similarly, h_{ik} represents the component formed at the k -th step of subtractions like in Equation 5 for purpose of gaining the i -th IMF (c_i). For FCEEMD, the targeted number of IMFs is two.

Standard deviation (SD) is often used as a stopping criterion for the subtraction process. That is, the sifting process will be stopped when the SD value falls into the range of $[0.2, 0.3]$.

The definition of basic functions and the description of the cubic trigonometric cardinal spline curve are given as follows (10):

Definition 1. For any given real number λ and parameter variable $0 \leq t \leq 1$, the following functions are called basis functions of the cubic trigonometric cardinal spline with a shape controlling parameter:

$$\begin{cases} B_{0,3}(t) = \frac{2}{\pi}\lambda S^3 - \frac{2}{\pi}\lambda C^3 + \frac{2}{\pi}\lambda C^2 - \frac{2}{\pi}\lambda S \\ B_{1,3}(t) = (\frac{2}{\pi}\lambda - 1)S^3 - (1 - \frac{2}{\pi}\lambda)C^3 + (1 - \frac{2}{\pi}\lambda)S^2 + \frac{2}{\pi}\lambda C \\ B_{2,3}(t) = (1 - \frac{2}{\pi}\lambda)S^3 - (\frac{2}{\pi}\lambda - 1)C^3 + (1 - \frac{2}{\pi}\lambda)C^2 + \frac{2}{\pi}\lambda S \\ B_{3,3}(t) = -\frac{2}{\pi}\lambda S^3 + \frac{2}{\pi}\lambda C^3 + \frac{2}{\pi}\lambda S^2 - \frac{2}{\pi}\lambda C \end{cases} \quad (6)$$

where $S = \sin(\frac{\pi t}{2})$, $C = \cos(\frac{\pi t}{2})$. Basis functions also satisfy the following conditions:

$$\begin{aligned} B_{0,3}(0) &= 0, B_{0,3}(1) = 0, B'_{0,3}(0) = -\lambda, B'_{0,3}(2, 1) = 0 \\ B_{1,3}(0) &= 1, B_{1,3}(1) = 0, B'_{1,3}(0) = 0, B'_{1,3}(2, 1) = -\lambda \\ B_{2,3}(0) &= 0, B_{2,3}(1) = 1, B'_{2,3}(0) = \lambda, B'_{2,3}(2, 1) = 0 \\ B_{3,3}(0) &= 0, B_{3,3}(1) = 0, B'_{3,3}(0) = 0, B'_{3,3}(2, 1) = \lambda \\ B_{0,3}(t) &= B_{3,3}(1-t), B_{1,3}(t) = B_{2,3}(1-t) \end{aligned}$$

$$\sum_{j=0}^3 B_{j,3}(t) = 1$$

Definition 2. Let $q_r(t = 0, 1, \dots, n)$ be a set of given discrete points in R^2 or $R^3 (n \geq 3)$. For any four ordinal control points $q_{i-1}, q_i, q_{i+1}, q_{i+2}$, the parameter curve segments are defined as follows:

$$P_i(t) = \sum_{j=0}^3 q_{i-1+j} B_{j,3}(t)$$

$$= q_{i-1}B_{0,3}(t) + q_iB_{1,3}(t) + q_{i+1}B_{2,3}(t) + q_{i+2}B_{3,3}(t) \quad (7)$$

($0 \leq t \leq 1, i = 1, 2, \dots, n-2$ is the segment number). Here the curve segment, $P_i(t)$, refers to the segmented cubic trigonometric cardinal spline curve, which in turn forms the cubic trigonometric cardinal spline curve known as $P(t)$. Since the basis function expressions have a shape controlling parameter, λ , we can get different spline basis function expressions when the value of parameter λ changes.

2.5 GA-FCEEMD Model

To forecast using the GA-FCEEMD model, it involves three steps:

1) Apply the aforementioned FCEEMD algorithm, where the target number of IMFs is two.

2) Set IMF2 as the intrinsic prediction trend for the actual time series data.

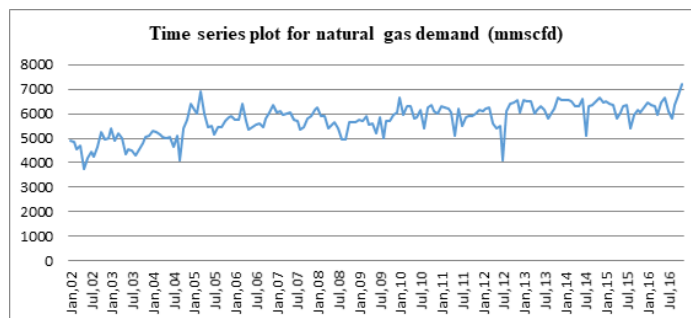
3) Obtain parameter and the stopping criterion for FCEEMD using GA, whose objective function is the MAPE. The calculation for MAPE is:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{A_t - F_t}{A_t} \right| \times 100 \quad (8)$$

where N is the number of observations in a series, A_t and F_t are actual and forecasted values in a time, .

3 Natural Gas Demand

The natural gas demand time series data were collected from Malaysia's Ministry of Finance (MOF) office. The data consisted of 179 observations and ranged from January 2002 to November 2016. Figure 2 shows the time series plot for the natural gas demand. The data were measured using million standard cubic foot per day (mmscfd)⁴. From the plot, we noticed that the trend of the natural gas demand increased, and in between there were quite a number of fluctuations and erratic ups and down, and also a volatility pattern in the data series.

Figure 2. Natural Gas Demand (mmscfd)⁴

3.1 Forecasting Modeling and Results

In this section, the modeling and forecasting of the natural gas demand are compared between ARIMA, ANN, GARCH, and GA-FCEEMD methods. The natural gas demand data have been divided into two, which are the training sets (167 observations) and testing sets (12 observations).

For ARIMA model, the actual natural gas demand data were differenced to obtain a stationary or trend-free series. Parameters p and q for the potential model were determined using Auto-correlation Function (ACF) and Partial Autocorrelation Function (PACF). After running a trial-and-error process, ARIMA(1,1,0) was chosen as the fitting model for the natural gas demand.

For ANN model, firstly the natural gas demand data were normalized to (0,1) values. Then, we designed a minimal fitting ANN model to the actual data which is composed of 2 input layers, a 3-hidden-layer network with 13 neurons per layer, a hyperbolic tangent-sigmoid transfer function hidden layer, and a linear transfer function output layer.

As for GARCH model, the modeling process and the gathering of parameter values for α_0 , α_1 and β_1 were performed directly using Excel after we maximized the Likelihood Estimation. The final values for $\alpha_0=0.0006$, $\alpha_0=0.3454$ and $\beta_1=0$. For GA-FCEEMD model, after we integrated the parameters for the stopping criterion and the value of λ with GA, the final stopping criterion for the natural gas demand data that gave the minimal MAPE was 0.9 and $\lambda=0.6$. Figure 3 shows IMF1 and IMF2 for the natural gas demand after

having implemented the GA-FCEEMD method.

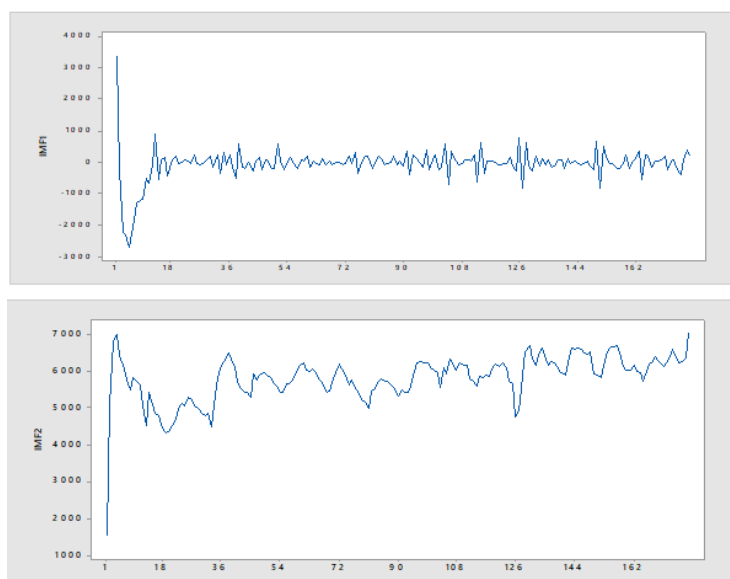


Figure 3. IMFs for Natural Gas Demand Data

From Figure 3, the 1st to 10th observations of IMF1 and IMF2 showed unstable spikes in the frequency. These ten initial observations did affect the intrinsic prediction trend of the natural gas demand.

After the modeling process, we forecasted the natural gas demand. Table 1 depicts the MAPE for the training and testing of the natural gas demand data.

Table 1: Forecasting Performances

Method	Training	Testing
	165 observations (MAPE %)	12 observations (MAPE %)
ARIMA	5.4025	5.4619
ANN	5.4823	5.1662
GARCH	5.6125	4.9435
GA-FCEEMD	5.3822	2.6642

From Table 1, GA-FCEEMD gave the best forecasting performances as compared to the other methods. For the data training, the MAPE for GA-FCEEMD achieved the best forecasting accuracy

despite the average accuracy. This is because the first ten observations for IMF1 and IMF2 were not very stable. Meanwhile for the data testing, the proposed hybrid GA-FCEEMDs forecast performance continuously improved since the IMFs were already stable. The time series plot of forecast testing among the four models is shown in Figure 4.

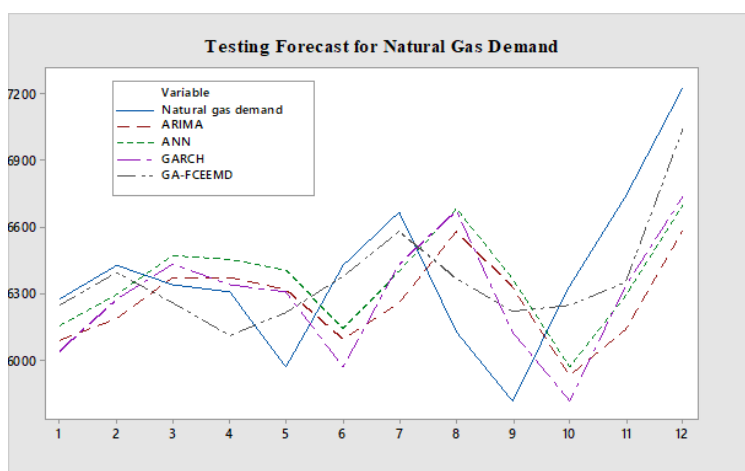


Figure 4. Forecast Testing of Natural Gas Demand using ARIMA, ANN, GARCH, and GA-FCEEMD Models

4 Conclusion

A new hybrid GA-FCEEMD forecasting model was proposed in this study. To investigate the forecasting performance of the new hybrid model, a natural gas demand data set was used as the experimental data for this study. The forecasting performance of the proposed GA-FCEEMD model was compared to that of ARIMA, ANN, and GARCH models. The results revealed that the forecasting of natural gas demand using GA-FCEEMD gave the highest accuracy, especially in data testing.

Acknowledgement: This work was supported by UTM under the Tier 1 grant, vote no. 11H89, and MyBrain15 scholarship through MyPhD.

References

- [1] Bai Y, Li C. Daily natural gas consumption forecasting based on a structure-calibrated support vector regression approach. *Energy and Buildings*. 2016;127:571-9.
- [2] Zhang W, Yang J. Forecasting natural gas consumption in China by Bayesian Model Averaging. *Energy Reports*. 2015;1:216-20.
- [3] Zeng B, Li C. Forecasting the natural gas demand in China using a self-adapting intelligent grey model. *Energy*. 2016;112:810-25.
- [4] Potonik P, Soldo B, imunovi G, ari T, Jeromen A, Govekar E. Comparison of static and adaptive models for short-term residential natural gas forecasting in Croatia. *Applied Energy*. 2014;129:94-103.
- [5] Liu H, Tian H, Liang X, Li Y. New wind speed forecasting approaches using fast ensemble empirical model decomposition , genetic algorithm , Mind Evolutionary Algorithm and Arti ficial Neural Networks. 2015;83:1066-75.
- [6] Niu M, Wang Y, Sun S, Li Y. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM $_{2.5}$ concentration forecasting. 2016;134:168-80.
- [7] Box GE, Jenkins G. *Time Series Analysis: Box/Time Series Analysis*. Hoboken: John Wiley and Sons, Inc.; 1994.
- [8] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representation by error propagation, parallel distributed processing. *Explorations in the Microstructure of Cognition*. 1986;1:318-62.
- [9] Zhao D, Huang Z, Li H. An Improved EEMD Method Based on the Adjustable Cubic Trigonometric Cardinal Spline Interpolation. 2017.
- [10] Ping ATAYD, Zhang Y. Adjustable Cubic Triangle Cardinal Interpolation Spline Curve. *Mathematical Numerical Singularities*. 2015;37:34-41.