Contemporary Engineering Sciences, Vol. 10, 2017, no. 6, 273 - 284
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ces.2017.717

FEEMD-DR Model for Forecasting Water Consumption

Nuramirah Akrom and Zuhaimy Ismail

Department of Mathematical Sciences
Faculty of Sciences
Universiti Teknologi Malaysia
81310, Skudai, Johor, Malaysia

Abstract

To improve forecasting accuracy and reliability, a combined model based on a favorable technique of "decomposed and ensemble" and a multivariate method of Dynamic Regression was proposed for forecasting water consumption data series. For the Dynamic Regression model, rainfall data was treated as the exogenous variable. The proposed combined model involved three main steps: first, water consumption and rainfall data series were decomposed into two Intrinsic Mode Functions (IMFs) using Fast Ensemble Empirical Mode Decomposition (FEEMD), then for each IMF of the separated data, the IMFs were predicted using Dynamic Regression, and lastly all the predicted IMFs were aggregated as the ensemble and forecasting process. By comparing the proposed combined FEEMD-DR model and the single model of Dynamic Regression, the results revealed that the FEEMD-DR model improved the forecasting accuracy for water consumption data series.

Keywords: Fast Ensemble Empirical Mode Decomposition (FEEMD), Dynamic Regression, Water consumption, Rainfall data

1 Introduction

Water scarcity is increasingly becoming a variable of interest due to the increasing severity of climate change [1]. Unpredictable rainfall patterns make water a resource that is insufficient in many countries. Therefore, the forecasting of water demand and supply has become one of the important endeavors for many countries.

For the purpose of water consumption modeling and forecasting, various methods have been proposed in previous researches such as by Makki et al. [1] which modeled and forecasted end-use water consumption using ADHEUC model. Romano and Kapelan [2] suggested forecasting a smart water distribution system using Artificial Neural Network (ANN) and evolutionary system. Brentan et al. [3] investigated short-term water demand forecast in Branca, Brazil, using a hybrid of support vector regression and Fourier time series for building a base prediction of water demand. Lomet et al. [4] applied statistical modeling for real domestic hot water consumption forecasting. Mamade et al. [5] implemented a spatial and temporal method for forecasting water consumption at District Metered Area (DMA) level with extensive measurements. Candelieri and Archetti [6] demonstrated time series clustering for the daily water pattern and applied support vector regression model that acted as a regression model in predicting each clustered daily water pattern.

The forecasting techniques reviewed above are complex and time-consuming [7]. In this study, we propose a combined "decomposed and ensemble" with multivariate method, namely FEEMD-DR model, for forecasting water consumption in Malaysia. The proposed FEEMD-DR model only needs three forecasting procedures: first was to apply Fast Ensemble Empirical Mode Decomposition (FEEMD) to decompose water consumption and rainfall time series data into two different Intrinsic Mode Functions (IMFs), then each IMF of the water consumption and rainfall data series was fitted into the Dynamic Regression model, and lastly the predicted IMFs were aggregated and made into the forecasting model. The monthly water consumption and rainfall data sets in Malaysia were collected and deployed to evaluate the effectiveness of the proposed forecasting technique.

The rest of the paper is organized as follows: Section 2 briefly introduces the methodology of the single model of Dynamic Regression and the combined model of FEEMD-DR, Section 3 introduces the water consumption and rainfall data series as well as the modeling and forecasting procedures for water consumption data series using DR and FEEMD-DR, then Section 4 presents the forecasting results using DR and FEEMD-DR, and Section 5 concludes the paper.

2 Methodology

2.1 Dynamic Regression

The general notation for the Dynamic Regression model can be described as follows:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + e_t \tag{1}$$

where y_t is the linear function of k predictor variables of $x_{1,t}, \cdots, x_{k,t}$, $\beta_0, \beta_1, \cdots, \beta_k$ is the parameter estimation, and e_t is the assuming uncorrelated error term. When $e_t = n_t$, the error series n_t follow the ARIMA model [8]. If the data series have seasonality, then the SARIMA model is used for error series.

2.2 Fast Ensemble Empirical Mode Decomposition (FEEMD)

In building the FEEMD algorithm, there were two conditions which the IMFs had to meet [9]:

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

Hence, the decomposition of the IMFs follows the following basic steps:

1) Set the ensemble number and the amplitude of the added white noise.
2) Add a white noise series to the targeted data with the selected amplitude and let the generated series be $x(t)$.
3) Use the cubic spline interpolation to link local maxima points and local minima points of $x(t)$ and obtain the upper and lower envelopes, respectively.
4) Calculate the mean curve $m_{11}(t)$ of the upper and lower envelopes and appoint the difference between $x(t)$ and $m_{11}(t)$ as $h_{11}(t)$:

$$h_{11}(t) = x(t) - m_{11}(t) \tag{2}$$

If $h_{11}(t)$ meets the two aforementioned conditions, it is regarded as the pioneer IMF or else, it is seen as an original series and thus repeated into the above processes until $h_{1k}(t)$ satisfies the requirements of becoming an IMF. Here, $c_1(t) = h_{1k}(t)$.

$$\begin{aligned} h_{12}(t) &= h_{11}(t) - m_{12}(t) \\ h_{13}(t) &= h_{12}(t) - m_{13}(t) \\ &\vdots \\ h_{1k}(t) &= h_{1(k-1)}(t) - m_{1k}(t) \end{aligned} \tag{3}$$

Notice that m_{ik} represents the mean curve formed at the k-th step of subtractions like Equation (3) for the purpose of obtaining the i-th IMF (c_i). Similarly, h_{ik} represents the component formed at the k-th step of subtractions like Equation (3) for the purpose of gaining the i-th IMF (c_i).

Standard deviation (SD) is used as a stopping criterion for the subtraction process. The sifting process is stopped when the SD value falls into range [0.2, 0.3] and outputs the IMF.

$$SD = \sum_{t=1}^{N} \frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{(h_{1(k-1)}(t))^2} \tag{4}$$

5) Regard the residual $r_1(t)$ after extracting $c_1(t)$ from $x(t)$ as a new series and repeat the above process to separate $c_2(t)$ from it until $r_n(t)$ becomes a monotonic function.

$$r_1(t) = x(t) - c_1(t),$$

$$r_2(t) = r_1(t) - c_2(t),$$

$$r_3(t) = r_2(t) - c_3(t),$$

...

Finally, $x(t)$ is expressed as:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{5}$$

where $c_i(t)(i = 1,2,\cdots,n)$ are N IMFs and $r_n(t)$ is a residual. For FEEMD, $r_n(t)$ is regarded as a monotonic function. We set the IMF for two iterations only, where the second IMF follows closely the trend of the actual data series. This is because the steps of building EEMD suggested by Wu and Huang [9] may be time consuming in decomposing IMFs. These forecasting steps have minimal procedure, yet improve forecast accuracy.

2.3 FEEMD-DR Model

The main purpose of this study was to forecast water consumption data using FEEMD-DR model. The steps in building the combined FEEMD-DR model are as follows:

Step 1: Decompose the original water consumption and rainfall data series using the FEEMD algorithm. In this study, the targeted number of IMFs for both water consumption and rainfall data series was two.

Step 2: For each IMF, predict the IMF series using the Dynamic Regression model; this step is known as IMF-DR.

Step 3: Aggregate all the predicted IMF-DR series and forecast the series.

2.4 Mean Average Percentage Error (MAPE)

To investigate the forecasting performance between the DR and FEEMD-DR models, the MAPE criterion was adopted. The detailed equation for MAPE is as follows:

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t} \right| \tag{6}$$

where N is the number of samples in the forecasted series, A_t is the actual data series, and F_t is the forecasted data series.

3 Data Series

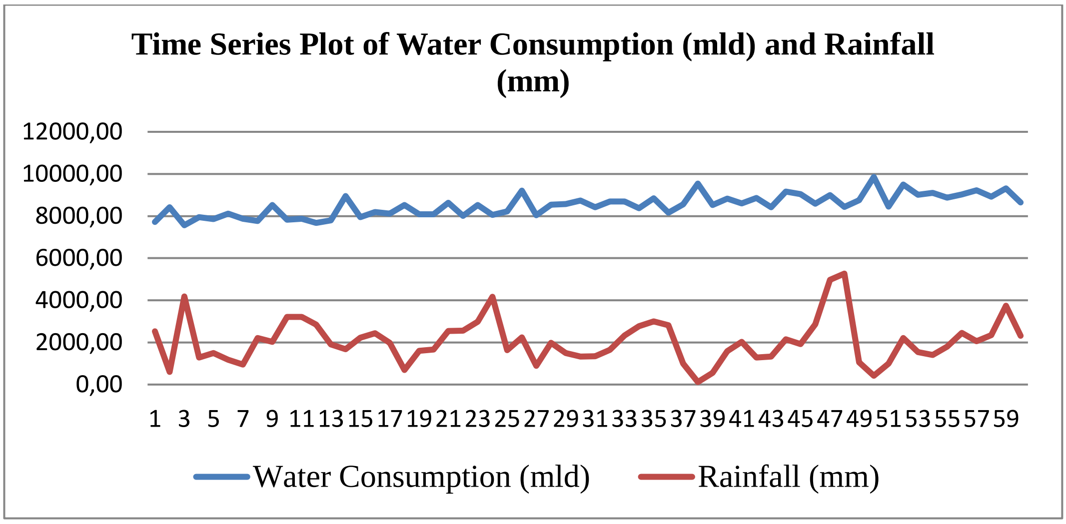


Figure 1: Time series plot of water consumption (mld) and rainfall (mm)

Figure 1 shows the time series plot of the water consumption and rainfall data. The water consumption data were measured in million liters per day (mld) and the rainfall data were measured in millimeter (mm). The data were plotted in Figure 1 which shows the monthly seasonality of approximately five years. The data set had 60 observations, corresponding to the period of January 2011 to December 2015. Based on Figure 1, the water consumption indicates an unstable pattern where there are ups and downs of water frequency per month. For the rainfall data, it clearly shows a random pattern.

3.1 Forecasting Computation

The forecasting steps popularized by Pankratz [8] were deployed in this study in order to model and forecast the water consumption data using the Dynamic Regression model. First, we plotted the ACF and PACF for the water consumption and rainfall data (see Figure 2). From Figure 2, the ACF and PACF plots for the water consumption and rainfall data showed a non-stationary series. Second, the water consumption and rainfall data were differenced, i.e. $d_1 = 1$, to eliminate non-stationary series from the water consumption and rainfall data (see Figure 3). Then, we differenced the data once again and this time, seasonal differencing was applied, i.e. $D_1 = 1$ and $s = 12$, since there existed monthly seasonality pattern for the data series (see Figure 4). After the second differencing, both data series became stationary (see Figure 5).

This step basically involved building the SARIMA model. Then, an SAS coding was built in order to estimate the parameter for the predictor variable and SARIMA model.

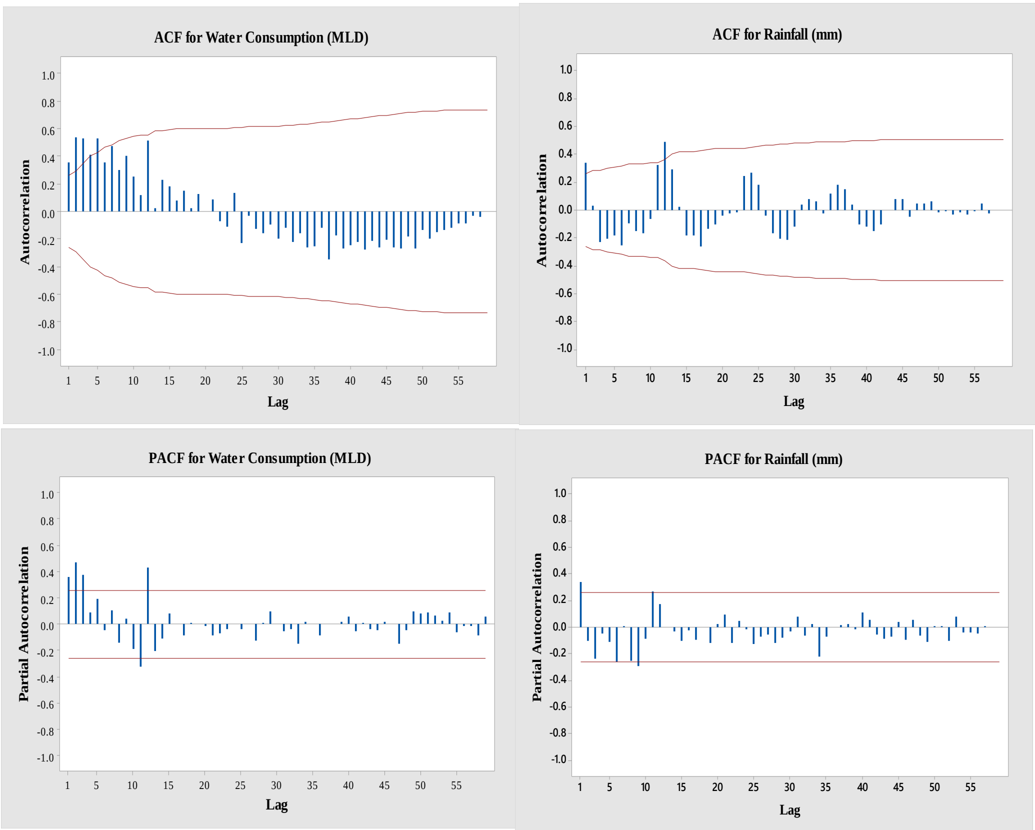
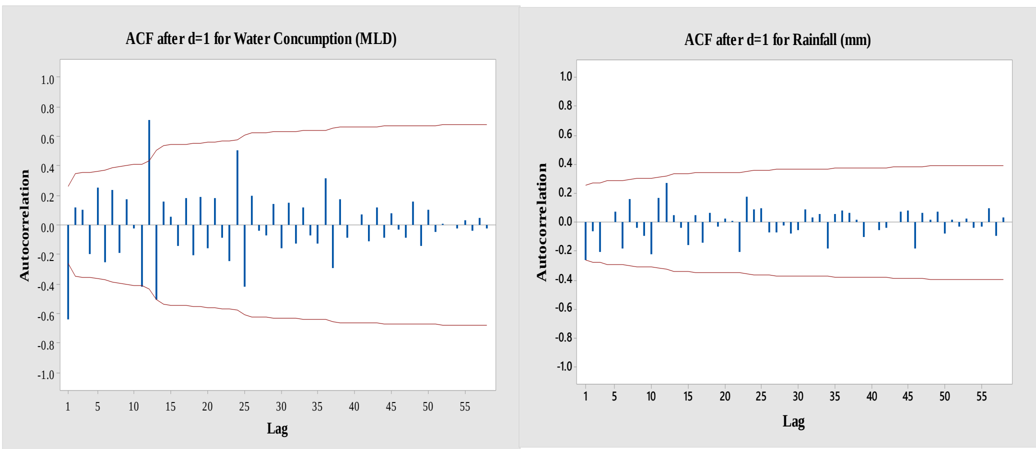


Figure 2: ACF and PACF plots for water consumption and rainfall data



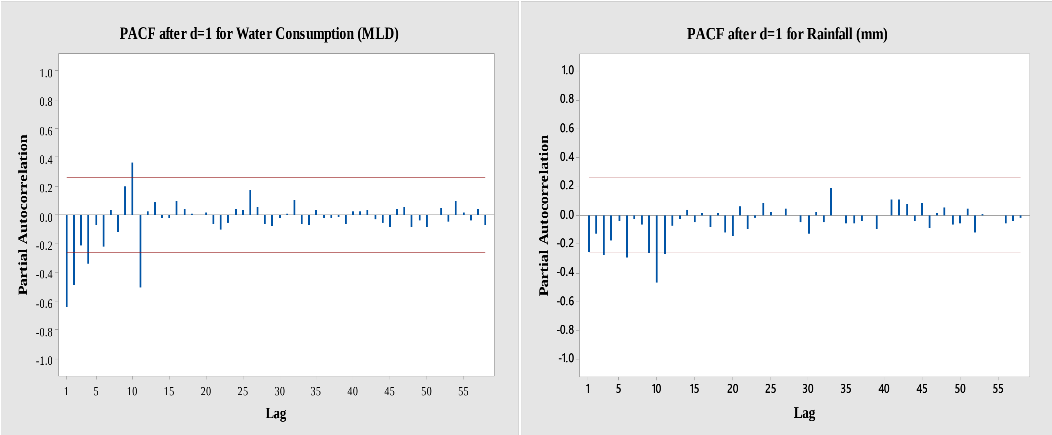


Figure 3: ACF and PACF plots for water consumption and rainfall data after $d = 1$

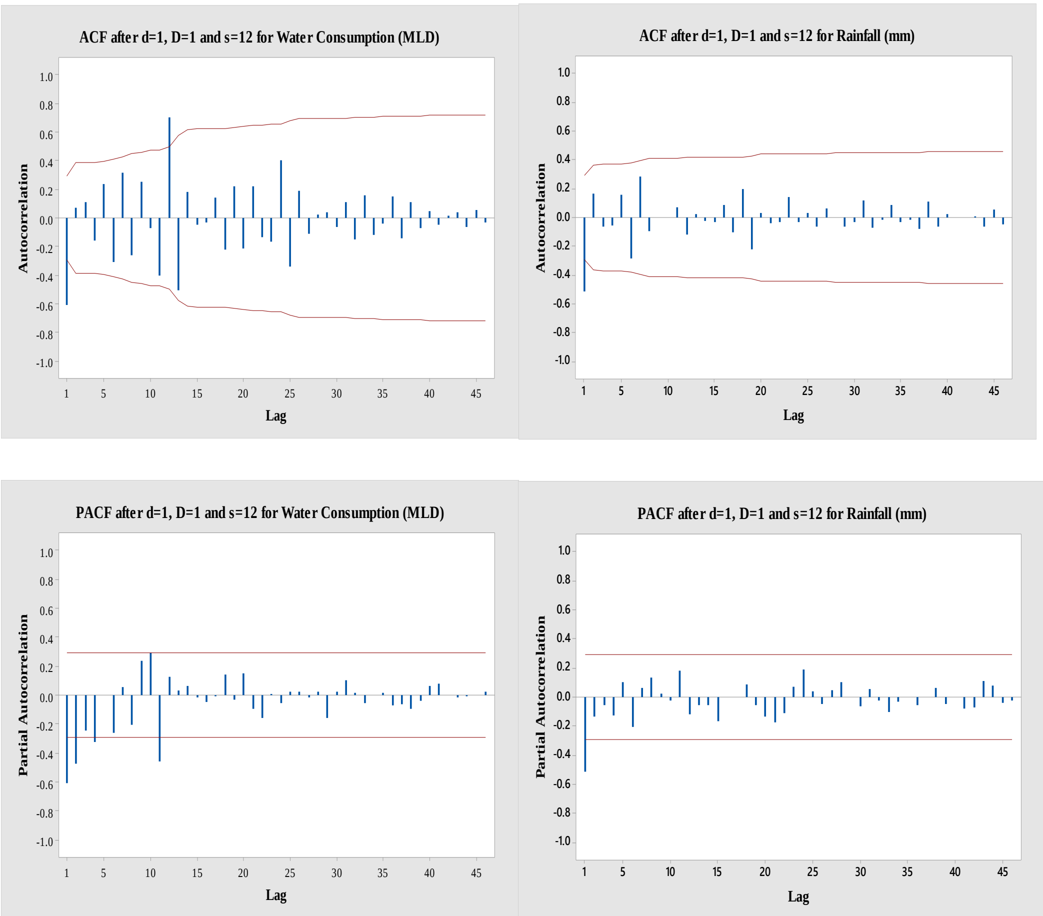


Figure 4: ACF and PACF plots for water consumption and rainfall after $D_1 = 1$ and $s = 12$

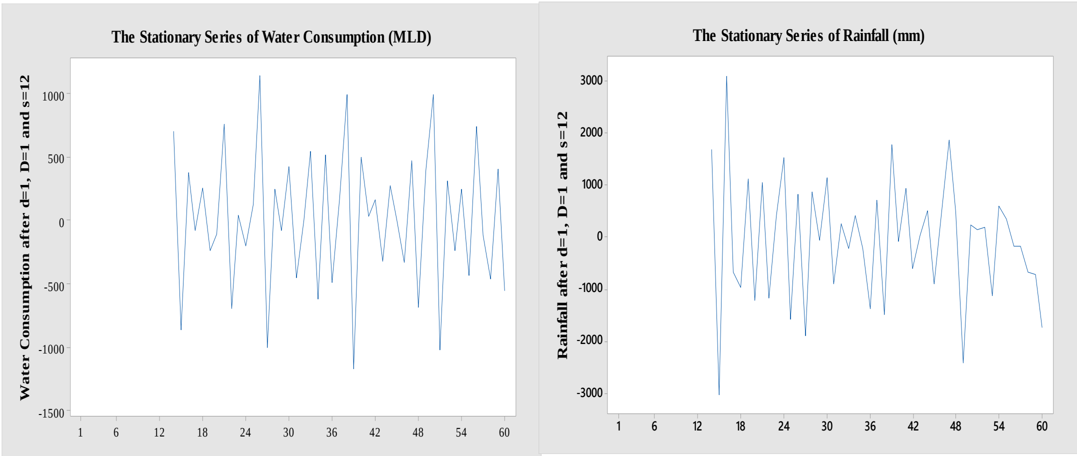


Figure 5: Stationary series of water consumption and rainfall data

Table 1 shows the parameter estimation for the predictor variable and SARIMA model.

Table 1: Parameter estimation for predictor variable and SARIMA model

Seasonal ARIMA Model		Transfer Function Model	
Parameter	Estimate	Parameter	Estimate
θ_1	-0.05135	ω_0	-0.01872
ϕ_1	1.39515	ω_1	-0.02650
ϕ_{12}	0.85901	δ_1	0.19603
		δ_{12}	0.47333

After estimating the parameters, the forecasting process occurred in the next step for the Dynamic Regression model. The DR model for the water consumption data is:

$$y_t = \frac{(-0.01872-0.02650B)}{(0.19603B+0.47333B^{12})} x_t + \frac{(1.39515B+0.85901B^{12})}{(-0.05135B)} n_t \quad (7)$$

3.2 FEEMD-DR Models Architecture

An algorithm using Matlab 2010 was built for the FEEMD-DR architecture. Figure 6 shows IMF1 and IMF2 for the water consumption and rainfall data series. From Figure 6, both IMF1s were in high frequency while IMF2s showed the actual smoothing trend of the data.

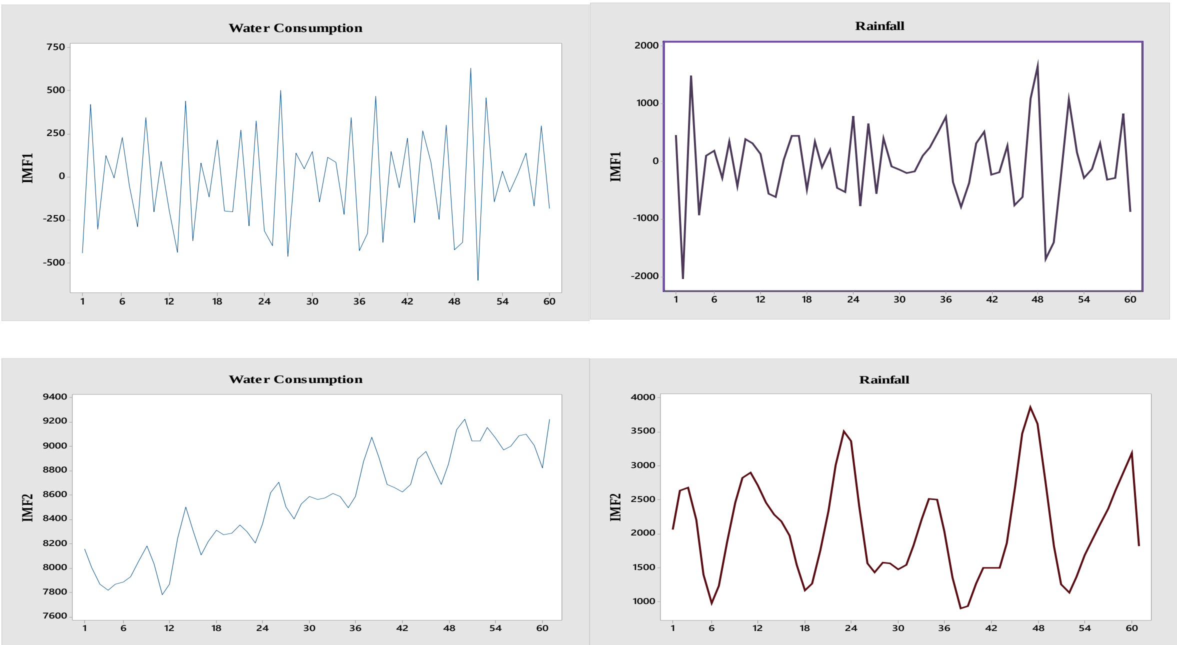


Figure 6: IMF1 and IMF2 for water consumption and rainfall data series

Next, the DR model was fitted for each IMF of water consumption. The IMF for rainfall was used as the exogenous variable for the IMF of water consumption. Table 2 presents the IMF-DR model for each IMF. From the table, the parameter estimation for each IMF was minimized since after FEEMD sifting process, the IMF was easy to fit into the DR model.

Table 2: IMF-DR models

IMF	IMF-DR Models
IMF1	$\widehat{IMF1}_t = \dfrac{(-0.0026 - 0.00498B)}{(1 + 1.8536B + 1B^2)} x_t + (1 - B)n_t$
IMF2	$\widehat{IMF2}_t = \dfrac{(-0.1052 - 0.04107B)}{(1 + 0.07626B + 0.97442B^2)} x_t + (1 + 0.84232B)n_t$

After fitting the IMF-DR model to each IMF, all the IMF-DR predicted equations were aggregated as follows:

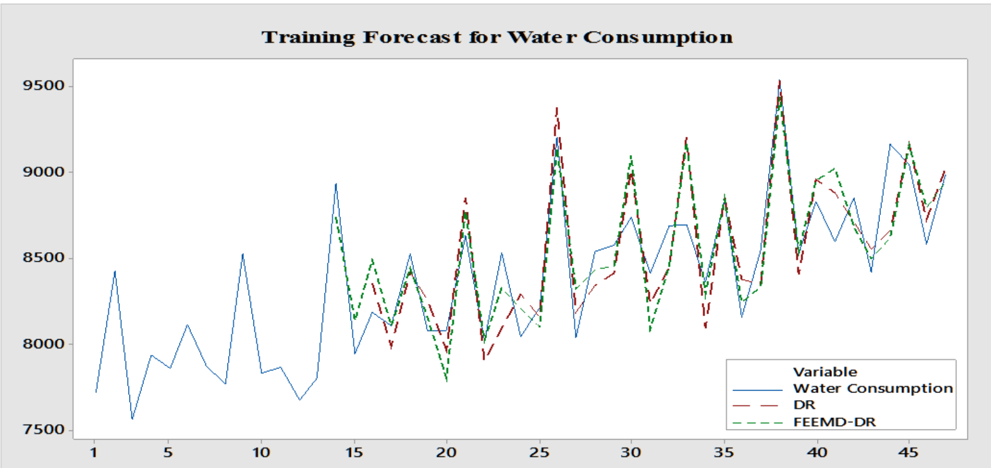
$$\widehat{IMF} = \widehat{IMF1} + \widehat{IMF2} \tag{8}$$

4 Forecasting Results

As for the forecasting results, the data were divided into training (48 observations) and testing (12 observations) forecasts. Table 3 shows the forecasting results for the training and testing water consumption data sets using the single model of DR and the combined model of FEEMD-DR.

Models	MAPE (%)	
	Training	Testing
DR	2.1815	2.9474
FEEMD-DR	2.0461	2.3639

Figure 7 demonstrates the training and testing forecasts for the water consumption using DR and FEEMD-DR models.



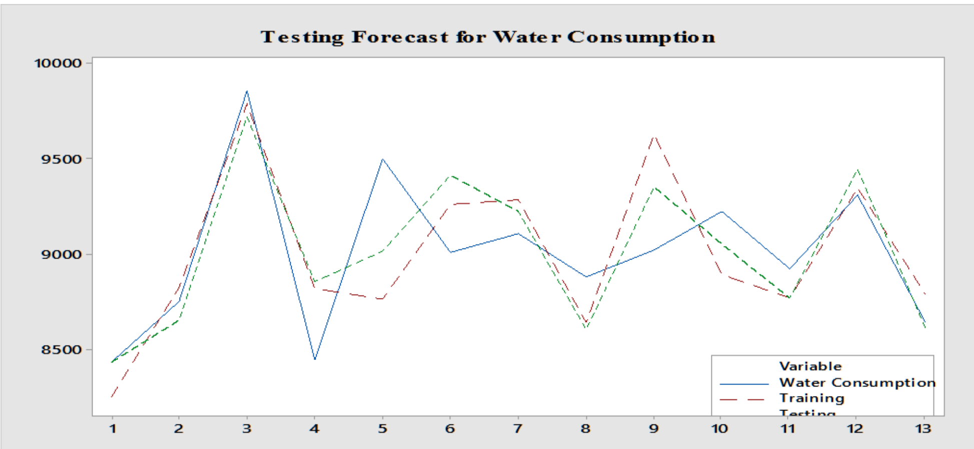


Figure 7: Training and testing forecasts for water consumption using DR and FEEMD-DR models

From Figure 7, the combined FEEMD-DR model gave higher forecasting accuracy than the DR model. This was due to the sifting process of the FEEMD algorithm which differentiated the first IMF from second IMF. For the first IMF, the frequency was higher as compared to the second IMF. The different frequency of each IMF gave a predicted DR model that was closer to the water consumption’s actual data series.

5 Conclusions

In this study, a new combined FEEMD-DR model had been proposed for water consumption prediction. The forecasting performance between single Dynamic Regression model and combined FEEMD-DR model were compared. The forecasting results for the FEEMD-DR model were slightly better than the single DR model.

Acknowledgements. This study was fully supported by MyPhD Scholarship and UTM Tier 1 Grant Vote No. 11H89.

References

[1] A.A. Makki, R.A. Stewart, C.D. Beal, and K. Panuwatwanich, Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption, *Resources, Conservation and Recycling*, **95** (2015), 15-37. https://doi.org/10.1016/j.resconrec.2014.11.009

[2] M. Romano, and Z. Kapelan, Adaptive water demand forecasting for near real-time management of smart water distribution systems, *Environmental Modelling and Software*, **60** (2014), 265-276. https://doi.org/10.1016/j.envsoft.2014.06.016

[3] B. M. Brentan, E. Luvizotto Jr, M. Herrera, J. Izquierdo, and R. Pérez-García, Hybrid regression model for near real-time urban water demand forecasting, *Journal of Computational and Applied Mathematics*, **309** (2017), 532-541. https://doi.org/10.1016/j.cam.2016.02.009

[4] A. Lomet, F. Suard, and D. Chèze, Statistical modeling for real domestic hot water consumption forecasting, *Energy Procedia*, **70** (2015), 379-387. https://doi.org/10.1016/j.egypro.2015.02.138

[5] A. Mamade, D. Loureiro, D. Covas, S.T. Coelho and C. Amado, Spatial and temporal forecasting of water consumption at the DMA level using extensive measurements, *Procedia Engineering*, **70** (2014), 1063-1073. https://doi.org/10.1016/j.proeng.2014.02.118

[6] A. Candelieri and F. Archetti, Identifying typical urban water demand patterns for a reliable short-term forecasting-The icewater project approach, *Procedia Engineering*, **89** (2014), 1004-1012. https://doi.org/10.1016/j.proeng.2014.11.218

[7] H. Liu, H. Tian, X. Liang, and Y. Li, New wind speed forecasting approaches using fast ensemble empirical mode decomposition, genetic algorithm, mind evolutionary algorithm and artificial neural networks, *Renewable Energy*, **83** (2015), 1066-1075. https://doi.org/10.1016/j.renene.2015.06.004

[8] A. Pankratz, *Forecasting with Dynamic Regression Models*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1991. https://doi.org/10.1002/9781118150528

[9] Z. Wu and N. E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Adv. Adapt. Data Anal.*, **1** (2009), no. 1, 1-41. https://doi.org/10.1142/s1793536909000047

Received: January 28, 2017; Published: February 21, 2017