

Drought Forecasting using Wavelet-GMDH Model with Standardized Precipitation Index

Mohammed Salisu Alfa, Ani Bin Shabri, Muhammad Akram Shaari

Abstract: This paper proposes Wavelet-Group Methods of Data Handling (W-GMDH) model to explore its ability of drought forecasting. The W-GMDH model was developed by combining Discrete Wavelet Transform (DWT) and GMDH model using the Standardized Precipitation Index (SPI) drought data for forecasting to assess the effectiveness of the new (W-GMDH) model. These methods were used on four SPI data sets (SPI3, SPI6, SPI9 and SPI12). To achieve this, a 624 month of SPI data from January 1956 to December 2008 was used and divided into two parts (80% for training and 20% for testing). The results of the W-GMDH model were then compared with the conventional GMDH model using Root Mean Square Error (RMSE), Mean Average Error (MAE) and coefficient of correlation as the performance evaluation measures. Both results of the proposed W-GMDH model and the GMDH showed very clearly that the propose method can achieve the best forecasting performance in terms of accuracy for each of the SPI data series. The key role played by the DWT is to smooth the analysis of SPI data obtained after the wavelet decomposition which is also used to decompose the SPI data into different number of component series to minimize the forecasting error. In all the results computed, the proposed model has a minimum error indicating its superiority over the GMDH model. This indicates that W-GMDH model's performance has outweighed that of the conventional GMDH model in SPI drought forecasting. The research contributes to the discovering of viable forecasting of drought and demonstrates that the established model is good and appropriate for drought. In all the analysis W-GMDH model has the minimum error. The overall results showed that SPI12 has the minimum error among all the SPI data considered.

Keywords: Drought Forecasting; GMDH; Time Series; SPI; DWT.

I. INTRODUCTION

Drought forecasting is an important aspect for implementing suitable mitigation measures to decrease negative effects on socioeconomic activities of man and environment. Drought is associated with various climatic and hydrologic processes such as precipitation, temperature, streamflow [1]. The accuracy of drought forecasting is fundamental to several decision process which improves the efficiency of forecasting models. The Group Methods of Data Handling (GMDH) model, which is new in draught forecasting will be used in identifying and providing further investigation on the combination of GMDH with the wavelet decomposition.

Revised Manuscript Received on November 15, 2019

Mohammed Salisu Alfa, Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia (UTM) 81300 Johor, Malaysia

Ani Bin Shabri, Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia (UTM) 81300 Johor, Malaysia

Muhammad Akram Shaari, School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM) 81300 Johor, Malaysia

This is expected to improve the drought forecasting of the GMDH approach. To achieve this goal, the Auto-Regressive Integrated Moving Average (ARIMA) and Support Vector Machine (SVM) models are used as benchmark models in comparison with the standard GMDH. It is a well-known fact that many researchers [2]; [3]; among others) have used these models to improve the accuracy of forecasting problems. Similarly, of recent GMDH have been used by researchers such as [4]; [5]; [6]; [7]. In this current work, however, we shall consider combination of GMDH and discrete wavelet transform (DWT) for drought forecasting to improve the performance of the GMDH model. GMDH is a technique which develops nonlinear systems involving several input variables. The GMDH algorithm was initially presented by Ivakhnenko in 1968 to produce complex mathematical models to handle data samples whose motive was to build linear models [2]. The application of GMDH is used in different fields to model and forecast the behaviors of complex systems which are based on input-output series of data [7]. It is a well-known fact that many researches have been carry out using traditional models like ARIMA and Exponential smoothing and artificial intelligent techniques such as SVM and ANN with GMDH to improve the performance of GMDH approach. Therefore, the focus of this work is to consider the GMDH approach with the wavelet as alternative to the ones mentioned earlier for the improvement and better performance of drought forecasting. To undertake this work, a Standardized precipitation index (SPI) data will be used for the drought forecasting. The objective of this paper, therefore, is to develop a drought time series forecasting model using GMDH with wavelet to improve the forecasting accuracy and to compare the forecasting performance both GMDH and W-GMDH models.

II. METHODOLOGY

It is a well-known fact that the methodology aspect of any research deals with data collected and the method of analysis. Hence, it is a procedure to follow to achieve a successful results and findings. This section therefore presents GMDH model used in modeling the drought forecasting as a focus of this paper.

2.3 Group Methods of Data Handling (GMDH) model
Group Method of Data Handling model introduced by Ivakhnenko, (1971) used to solve complex non-linear multidimensional short data series. In building the GMDH algorithm, the model is given in the form of linear polynomial in which the structure of the model is determined by the amount of terms in its polynomial [4]. GMDH has been used successfully to deal with the uncertainty and non-linearity of the systems in

different disciplines in the area of ecology, economy, medical aspects of diagnosing the ailments, signal processing and control. This is focused on extending the scope of GMDH to involve discrete wavelet transform (DWT) for the purpose of forecasting drought. This will provide time-frequency localization in the future selection to choose the optimum features or input node from the model. since GMDH is a method to develop non-linear models involving several input variables, it has been used extensively for modeling and forecasting non-linear process which are complex [5]. The proposed model is focused on a multiplier structure that uses each pair of variables involving a second order polynomial given by:

$$y = V_0 + \sum_{i=1}^n c_i x_i + \sum_{i=1}^n \sum_{j=1}^n V_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n V_{ijk} x_i x_j x_k + \dots \tag{2}$$

This is called Kolmogorov-Gabor Polynomial. Where $X(x_1, x_2, \dots, x_n)$, is the input variable vector. Where n is the number of inputs $V(v_1, v_2, \dots, v_n)$. For most of the application the quadratic aspect is called Partial Descriptions (PD) for two variables only is used. It is in the form of:

$$\hat{y} = v_0 + v_1 x_i + v_2 x_j + v_3 x_i^2 + v_4 x_j^2 + v_5 x_i x_j \tag{3}$$

Which is used to predict the output. The input variables are $X(x_1, x_2, \dots, x_n)$ and output is set to $\{y\}$. The coefficients v_i for $i=0,1,\dots,5$ are determined by using the least square method.

$V(v_1, v_2, \dots, v_n)$, is the vector coefficient weight. GMDH is developed with polynomial in equation 1 as PD as contained in table1 where M is the total number of inputs

Table 1: Summary of the previous studies that used wavelet transform

Study conducted by Authors	Methods carried out
Zheng et al., (1999); Bashir and El-Hawary (2000); Chen et al., (2006); Dash and Nayak (2007)	Neural network and wavelet function
Soltani et al., (2000)	Autoregression (AR) model and the wavelet function
Chang-il et al., (2006)	The wavelet transformation and the regression model
Bao et al., (2007)	Hybridizing Wavelet and Least Squares Support Vector Machines for Crude Oil Price Forecasting
C.Stolojescu et al. Al. (2010)	A Wavelet Based Prediction Method for Time Series
Khashei et al., (2011)	A novel hybridization of artificial neural networks and ARIMA models for time series forecasting
Pandhiani et al., (2013)	Time Series Forecasting using Wavelet-Least Squares Support Vector Machines and

Wavelet Regression models for Monthly Stream Flow Data
--

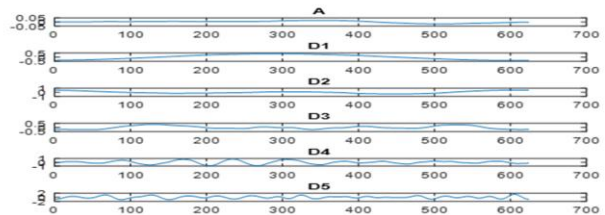
2.4 Discrete wavelet transform (DWT)

A wavelet is a mathematical function which is used in digital signal processing and image compression. Wavelet analysis is becoming a well-known tool because of its ability to show information within the signal in both the time and scale domains[6]. Wavelet is a mathematical procedure that involves the transformation of the original signal (most especially in the time domain) into a different domain in processing and in the analysis [7]. [8], in his study of hybrid wavelet and adaptive neuro-fuzzy inference system for drought forecasting stated that wavelet analysis is one of the most powerful tools to study time series. In another study, [9], described wavelet analysis as a multi-decomposition analysis that provide information for time and frequency domains and provide useful decompositions of the original time series for the wavelet-transformed data to improve the power of a forecasting model. Wavelet is a tool in time series forecasting whose importance is applied by many researchers. One of the basic objectives of wavelet transforms is to analyze the time series data. DWT is a mathematical instrument that gives strong representation involving a time-frequency in the time domain for a signal to be analyzed (Okkan, 2012 and Danandeh Mehr, Kahya, & O'zger, 2014). Wavelet transformations provide a very useful decomposition of the original time series by capturing a very useful information on the various decomposition levels. To obtain a number of decomposition level, the formula by (W. Wang & Ding, 2003 and Nourani, Alami and Aminfar, 2009) is applied. This is given as:

$$L = \text{int}[\log(N)] \tag{4}$$

Where L is the decomposition level and N is the number of the SPI data series. In the formula, the original SPI drought data series is decomposed into L components from A to D_{L-1} ($A_1, D_1, D_2, \dots, D_{L-1}$) which stands for different frequency components of the original data. In this paper, the number of decomposition levels, N, is equal to 5.

Instead of using the D's component separately, as input model, we employ added suitable D's component which is more useful and capable of increasing the forecast performances of the models.



2.5 Wavelet-GMDH model

We propose the combination of wavelet with the GMDH to obtain the W-GMDH model for the drought forecasting being proposed. This is to enhance the conventional GMDH by bringing the wavelet method and GMDH model



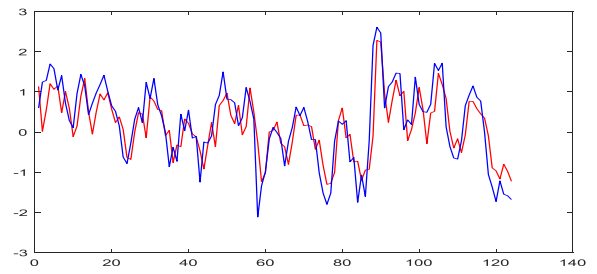
together to form one model. On the side of DWT, the SPI data series were decomposed into 5 levels (A, D₁, D₂, D₃, D₄, D₅) using the earlier given formula $L = \text{int}[\log(N)]$. The components of DWT are chosen and used as input GMDH model with the purpose of improving the forecasting performance of the combination of W-GMDH model [13]. The selection of these components of DWT is also to allow the GMDH to determine the features of the SPI data series which produces a good estimate. In the view of Zou & Yang, (2004), the combination of two models usually enhance the performance of the new model. The key importance of applying the DWT is to smooth the analysis of SPI data obtained after the wavelet decomposition. Wavelet transform decomposes the SPI data into different number of component series to minimize the error criterion [3]. The combination of two different models enhance the performance of the new model [14].

III. RESULTS AND FINDINGS

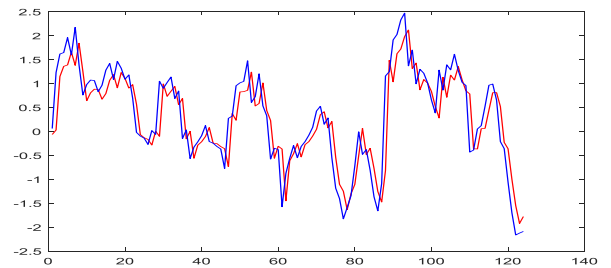
In this study, the SPI data set consists of 624 for the period of 52 years from January 1956 to December 2008 were used. The data from January 1956 to December 1998 which consists of 500 datasets representing 80% were used for training and the remaining 124 datasets representing 20% from 1999 to 2008 were used for testing. The results of the precipitation correspond to the 3, 6, 9 and 12-months data sets were used and their corresponding Standardized Precipitation Indexes (SPIs) computed. The time series for these SPIs values are calculated. The aim to consider the overall precipitation for the periods of 3, 6, 9 and 12 months was because of the classification of drought to be a short-term, medium-term and long-term for SPI3, SPI6 and SPI9 and for SPI12 respectively. The working of GMDH is building the successive layers with complex networks which are created using second-order polynomial function. The first layer is obtained by computing regression of input variables and then select the best ones. This is followed by the second layer and computing the regression of the best values of the first layer with the input variables in which only the best is chosen using the algorithm. The process continues until the selection criterion is achieved. To design the GMDH model, the following variables must be determined, the number of input nodes and layers. The selection of the number of inputs must correspond to that of the variables. The determination of the optimal number of input nodes is crucial and complicated. To produce GMDH and W-GMDH, six levels of inputs (2, 4, 6, 8, 10 and 12) and five layers for the experimental analysis to reduce the tedious computational burden.

GMDH Spi Graphs

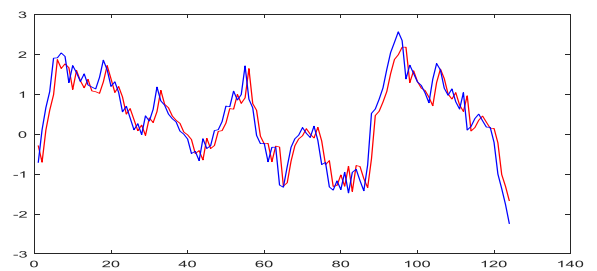
Spi3



Spi6



Spi9



Spi12

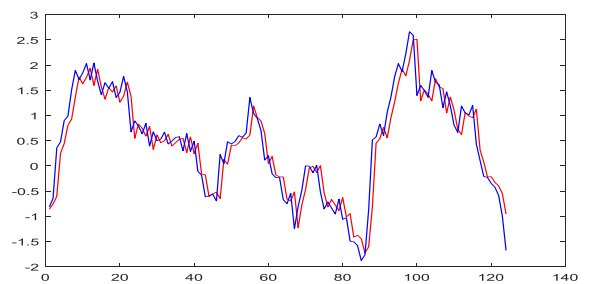


Table 2: showing the selected best GMDH models from all the SPI Data inputs

Data	Training			Testing		
	RMS	MAE	R	RMSE	MAE	R
Spi3	0.642	0.5107	0.7633	0.6232	0.4900	0.7768
	8	7	8	6	7	5
Spi6	0.471	0.3732	0.8865	0.4436	0.3277	0.9009
	4	6	5	5	7	7
Spi9	0.381	0.3055	0.9282	0.3650	0.2806	0.9322
		6	9	1	2	3
Spi1	0.366	0.2858	0.9327	0.3542	0.2701	0.9382
	2	3		3		6

Table 2: showing the selected best SPI data inputs for the GMDH model with respect to training and testing data. The selected SPI data gave the minimum RMSE, MAE and highest R in terms of the evaluation measures. In comparison, however,

Spi12 data is the best since it has the least error in the training and testing phases.

Spi12	0.21148	0.15875	0.97863	0.19413	0.14925	0.98232
-------	----------------	----------------	----------------	----------------	----------------	----------------

Table 3: showing the selected best W-GMDH models from all the SPI Data inputs

Data	Training			Testing		
	RMSE	MAE	R	RMSE	MAE	R
Spi3	0.38112	0.29732	0.92503	0.36933	0.29574	0.92689
Spi6	0.27659	0.21291	0.96373	0.24634	0.20004	0.97049
Spi9	0.26290	0.19113	0.96704	0.19121	0.14716	0.98250

Table 3 shows the best SPI data inputs selected for W-GMDH model among all the SPI data inputs (inputs 2, 4, 6, 8, 10 and 12) for the training and testing. The chosen SPI is the minimum error value using RMSE, MAE and highest R performance evaluation measures. Furthermore, Spi12 presented the best performance in both the training and testing phases.

Table 4: Showing The Comparison Best GMDH And W-GMDH Models From All The SPI Data Input

Data	Model	Training			Testing		
		RMSE	MAE	R	RMSE	MAE	R
Spi3	GMDH		0.51077	0.76338	0.62326	0.49007	0.77685
	W-GMDH	0.38112	0.29732	0.92503	0.36933	0.29574	0.92689
Spi6	GMDH	0.47142	0.37326	0.88655	0.44365	0.32777	0.90097
	W-GMDH	0.27659	0.21291	0.96373	0.24634	0.20004	0.9705
Spi9	GMDH	0.38109	0.30557	0.92829	0.36501	0.28062	0.93223
	W-GMDH	0.2629	0.19113	0.96705	0.19121	0.14716	0.98251
Spi12	GMDH	0.36631	0.2858	0.93275	0.35423	0.27012	0.93825
	W-GMDH	0.21148	0.15875	0.97863	0.19412	0.14925	0.98232

Page Layout

Table 4 describes the different results of the analysis with respect to the training and testing carried out for all the SPI data series used in the study. The results are the summary of the best selected inputs from the six inputs considered (inputs 2, 4, 6, 8, 10 and 12). For each model, the inputs with the least error (RMSE and MAE) and the highest coefficient of correlation (R) were chosen for comparison. In each of the SPI (Spi3, Spi6, Spi9 and Spi12) data, the W-GMDH model recorded the least in terms of RMSE and MAE indicating that it is better than the GMDH model in drought forecasting. The results indicate that Spi12 data has the smallest error in both GMDH and W-GMDH when compared with the other three data series. However, W-GMDH model which is the proposed model maintains the minimum error of RMSE and MAE and highest R for Spi12 data series.

IV. MEASURES OF PERFORMANCE EVALUATION CRITERIA

Generally, the performance evaluation criteria are widely used in evaluating the results of time series forecasting as described in various literature. These error measures are used as provided by Hyndman, (2014) and Shcherbakov & Brebels, (2013). In the view of Makridakis et al., (1998), ‘standards’ statistical measures of the forecast accuracy are made up of MSE and MAE. The criteria for judging the best model are how relatively small these values are in both the training and testing of the data. This is necessary to quantify the amount by which an estimator differs from the original

(true) value. That is why the measure with smaller values is usually selected as the best. These measures include, the Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Average Error (MAE), Mean Average Percentage Error (MAPE) and Correlation Coefficient (R). The performance of drought forecasting model is assessed by computing these measures of accuracy. However, for this study, three prediction evaluation criteria methods (RMSE, MAE and R) are used in our drought forecasting. Another popular application of these measures of performance was used by Dawson, Abrahart, & See, (2007) and defined as follows: -

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|} \tag{5}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{6}$$

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\left[\sum_{i=1}^n (y_i - \bar{y}_i)^2 \right]^{1/2} \left[\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)^2 \right]^{1/2}} \tag{7}$$

Where \hat{y}_i are the predicted values at time i , and y_i are the actual values at time i , and n is the number of Predictions.



V. CONCLUSION

From the various experimental analysis carried out, we can develop and propose a wavelet-GMDH by combining the DWT and GMDH for drought forecasting. To achieve this, GMDH is used with SPI data without preprocessing the data. Secondly, DWT is used to decompose SPI data into many sample series. Thirdly, the components of the wavelet decomposition were applied as input to GMDH model for the purpose of drought forecasting and the results were then compared. The performance accuracy measure of the W-GMDH for drought errors gave a better forecasting results compared with the GMDH model as presented in the tables. The performance evaluation of the proposed W-GMDH model based on RMSE, MAE and R was then compared with the GMDH model, the results showed great improvement and very potential method in SPI drought forecasting. The major key role played by the DWT is to smooth the analysis of SPI data obtained after the wavelet decomposition which is also used to decompose the SPI data into different number of

component series to minimize the forecasting error. The results revealed that W-GMDH model surpasses the GMDH since it produces the least error for all the inputs of SPI data series considered. With respect to the different SPI data sets SPI12 was also judged to be the best among other SPI data series for drought forecasting. Since W-GMDH has performed better than the GMDH in all the SPI data used according to various results obtained, it is then recommended as a tool which can be used for drought forecasting.

REFERENCES

1. A. F. Van Loon, M. H. J. Van Huijgevoort, and H. A. J. Van Lanen, "Evaluation of drought propagation in an ensemble mean of large-scale hydrological models," pp. 4057–4078, 2012.
2. A. G. Ivakhnenko, "Polynomial Theory of Complex Systems Polynomial Theory of Complex Systems," no. 4, pp. 364–378, 1971.
3. H. Basheer and A. Bin Khamis, "A HYBRID GROUP METHOD OF DATA HANDLING (GMDH) WITH THE WAVELET DECOMPOSITION FOR TIME SERIES FORECASTING: A REVIEW," vol. 11, no. 18, pp. 10792–10800, 2016.
4. A. G. Ivakhnenko and G. A. Ivakhnenko, "Problems of Further Development of the Group Method of Data Handling Algorithms . Part I," vol. 10, no. 2, pp. 187–194, 2000.
5. Y. Kim, H. Suk, and J. D. Plummer, "Environmental Modelling & Software A wavelet-based autoregressive fuzzy model for forecasting algal blooms," Environ. Model. Softw., vol. 62, pp. 1–10, 2014.
6. V. Nourani, M. T. Alami, and F. D. Vousoughi, "Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling," J. Hydrol., vol. 524, pp. 255–269, 2015.
7. B. Dong, Y. Mao, I. D. Dinov, Z. Tu, and Y. Shi, "Wavelet-Based Representation of Biological Shapes Wavelet-Based Representation of Biological," no. June 2014, 2009.
8. A. Shabri, "A hybrid wavelet analysis and adaptive neuro-fuzzy inference system for drought forecasting," Appl. Math. Sci., vol. 8, no. 139, pp. 6909–6918, 2014.
9. A. Shabri and R. Samsudin, "Fishery Landing Forecasting Using Wavelet-Based Autoregressive Integrated Moving Average Models," Math. Probl. Eng., vol. 2015, 2015.
10. U. Okkan, "Using wavelet transform to improve generalization capability of feed forward neural networks in monthly runoff prediction," vol. 7, no. 17, pp. 1690–1703, 2012.
11. A. Danandeh Mehr, E. Kahya, and M. O'zger, "A gene-wavelet model for long lead time drought forecasting," J. Hydrol., vol. 517,

- pp. 691–699, 2014.
12. W. Wang and J. Ding, "Wavelet Network Model and Its Application to the Prediction of Hydrology," vol. 1, no. 1, 2003.
13. B. Badyalina, "Streamflow Estimation at Ungauged Site Using Wavelet Group Method of Data Handling in Peninsular Malaysia," vol. 8, no. 11, pp. 513–524, 2014.
14. H. Zou and Y. Yang, "Combining time series models for forecasting," vol. 20, pp. 69–84, 2004.
15. L. R. J. Hyndman, "Forecasting: Principles & Practice," no. September, 2014.
16. M. V. Shcherbakov and A. Brebels, "A Survey of Forecast Error Measures," vol. 24, no. 4, pp. 171–176, 2013.
17. C. W. Dawson, R. J. Abrahart, and L. M. See, "HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts," Environ. Model. Softw., vol. 22, no. 7, pp. 1034–1052, 2007.