Original papers

Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia

Y.W. Soh, C.H. Koo, Y.F. Huang*, K.F. Fung

Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Jalan Bandar Sg. Long, Bandar Sg. Long, 43000 Kajang, Selangor, Malaysia

ARTICLE INFO

ABSTRACT

Drought forecasting is a vital for mitigating the impact of drought events on the economy, tourism, agriculture and water resource systems. This paper adopts the proposed Wavelet-ARIMA-ANN (WAANN) model and the latest Wavelet-Adaptive Neuro-Fuzzy Inference System (WANFIS) model to predict the Standardized Precipitation Evapotranspiration Index (SPEI) at the Langat River Basin for different time scales (1-month, 3-months and 6-months). Model input data pre-processing with wavelet decomposition for improving the performance of the models was carried out apriori. The historical SPEI from 1976 to 2007 were used in the WAANN and WANFIS models for predicting the SPEI for the test period from 2008 to 2015. The Adjusted Coefficient of Determination (R^2_{adj}), Root-Mean-Square-Error (RMSE), Mean Absolute Error (MAE), Willmott's Index of Agreement (d) and the Nash-Sutcliffe Coefficient of Efficiency (E) were used to assess the models. It was found that the prediction accuracy of the two models improved with time scale length. For the prediction of SPEI-1 (1-month), the errors associated with both models were considered relatively high. Based on the performance measures and graphical plots, the WAANN model is better for the prediction of SPEI-3 and SPEI-6. The WANFIS model had satisfactory prediction of the mid-term drought forecasting for all stations. The WAANN model developed in this study however, gives better accuracy for both, the short-term and mid-term drought forecasting.

1. Introduction

On the basis of the long-run average precipitation (normal precipitation) for a particular basin, the declining trend of precipitation indicates the initiation of droughts (Jalalkamali et al., 2015). Low relative humidity, temperatures, high wind velocity, rainfall characteristics including intensity, duration of precipitation and the distribution of rainfall during the crop growing seasons are important features of the droughts (Mishra and Singh, 2010). It has been reported in researches that the drought-induced conditions developed seasonally in the event of El Nino (also known as the warm phase of El Nino Southern Oscillation, ENSO). El Nino is induced by the reduction of trade winds and, in tandem with the increase of earth surface temperatures, appears to occur on the average, every 3–4 years (Paz et al., 2007).

Even though Malaysia receives an average of 2800 mm of precipitation annually, the country however is still subjected to prolonged dry spells; especially at the Langat River Basin, where the rapid urbanisation of the Kuala Lumpur City area has resulted in an increased demand for the freshwater supply (Pour et al., 2014). The ENSO

profoundly affects the condition of climate in Malaysia and in Southeast Asia. Yusof et al. (2012) applied the Kriging method to analyse the upward and downward trends during the occurrence of droughts in Peninsular Malaysia. Their results showed that the major regions of West Malaysia are subjected to an upward trend throughout the dry season, particularly in the eastern and western regions. It is extremely imperative for the water resources department to predict the drought intensity, severity and duration. With prior awareness of the onset of droughts, appropriate actions to mitigate the consequential damages can be considered. Precise and representative drought index series about the onset, extent and the end of the drought event allows the proper drought contingency plans to be established (Subash et al., 2011).

Drought monitoring and early warning are important phases to manage droughts (Bachmir et al., 2016). Among the approaches for drought forecasting, the use of artificial intelligence (AI) shows outstanding performance and accuracy (Masinde, 2014; Ozger et al., 2011; Belayneh et al., 2014). The flexibility and adaptability of AI is useful in predicting the occurrence of droughts that poses varying durations,
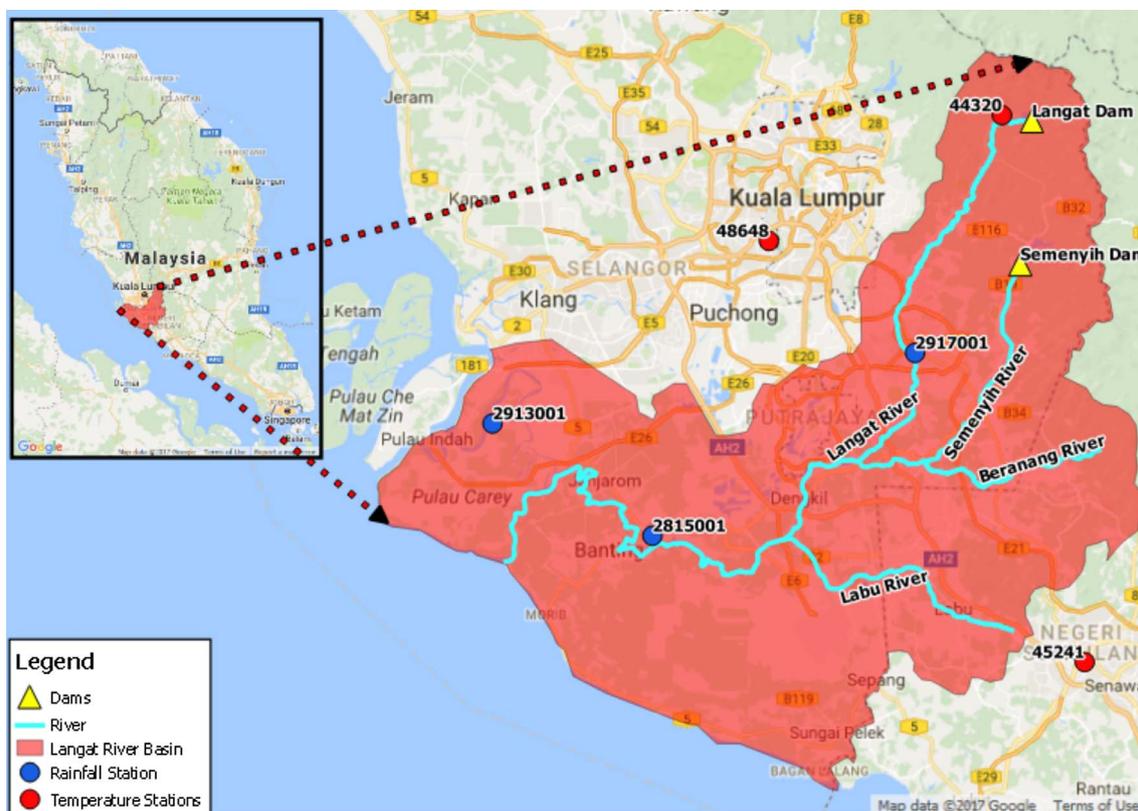

Fig. 1. Location of Langat River Basin.

frequencies and intensities; characteristics that are not easily determined effectively using empirical relations. Artificial Neural Network (ANN), Fuzzy Logic (FL) and Support Vector Regression (SVR) are the examples of AI models that can be used to forecast the time series modelling based on historical data (Djerbouai and Souag-Gamane, 2016). Another popular type of data-driven model is the stochastic method. The older Auto Regressive Integrated Moving Average Model (ARIMA) and Seasonal Auto Regressive Moving Integrated Average Model (SARIMA) are the most widely adopted stochastic approaches, as they are relatively simple yet providing excellent results (Mossad and Alazba, 2015; Bazrafshan et al., 2015).

In order to improve the prediction accuracy in the time series forecasting, different types of hybrid models were developed to minimize the error between the actual and predicted values. For example, both Choubin et al. (2014, 2016) explored the effectiveness of ANFIS model in forecasting the Standardized Precipitation Index (SPI) for different regions. Apart from that, ARIMA-ANN, Wavelet-ANN (WANN) and WANFIS are also effective forecasting tools among the hybrid models (Belayneh et al., 2014; Shabri, 2014). The wavelet transform had been proven to decompose the incoming signal into different frequencies (Djerbouai and Souag-Gamane, 2016; Belayneh and Adamowski, 2012, 2013; Özger et al., 2011, 2012) and thus, providing a more accurate output for the data driven models. By combining the strength of different models, the hybrid models can give a higher degree of accuracy.

In 2013, the WANFIS model was first used in drought forecasting because the wavelet transform has the ability to improve the accuracy of the models. The performance of the WANFIS model was compared with the WANN, ANN and ANFIS. WANFIS model results were more precise than other models in the study and the wavelet transform was particularly good in improving meteorological drought forecasting. Shabri (2014) also proposed to use the WANFIS model to forecast the drought events in Malaysia. Subsequently, the WANFIS model was widely recognised as an effective tool to forecast the time series of

drought index. However, the performance of the WANFIS highly relied on the modelling of the ANFIS model. Seo et al. (2016) had pointed the drawbacks of using the ANFIS modelling that would lead to difficulty in the identification of network parameters and the number and type of membership functions. Moreover, the development of the ANFIS model is computationally expensive and complex.

The aim of this paper is to improve the drought forecasting procedure at the Langat River Basin, using hybrid modelling. An alternative model, the WAANN was proposed in this study since the wavelet transform is able to reduce the complexity of the time series and the ARIMA-ANN model was proven as an appropriate tool in the time series forecasting (Babu and Reddy, 2014; Khashei and Bijari, 2011). The latest hybrid model WANFIS in drought forecasting is used for comparison. The expected results are the SPEI series predicted with both models and the accompanying performance evaluations.

2. Methodology

2.1. Study area and data acquisition

The catchment area of the Langat River Basin in Selangor, Malaysia is about 2400 km², supplying approximately 65% of water usage in the state of Selangor. Two reservoirs in the Langat River Basin, namely the Semenyih dam (area of 56.6 km²) and the Hulu Langat dam (area of 41.0 km²) supply water to the state. The recorded daily precipitation and temperature for each station in the basin were retrieved from the Department of Irrigation and Drainage (DID) Malaysia and the Malaysian Meteorology Department (MMD). After determining the suitability of the data, the records from six stations were used. A set of daily meteorological data, including rainfall and temperature were obtained for these stations for the period of 1976–2015; with the exception of station 44320, where only the daily data from 1985 to 2015 is available. The location of the Langat River Basin is illustrated in Fig. 1 and the details of selected hydrometric stations in Langat River

Table 1
Details of the selected hydrometric stations at the Langat River Basin.

Station name	Station ID	Available data	Record's period
Pejabat JPS Sg. Manggis	2815001	Precipitation	1976–2015
P/K WLN P/S Telok Gong	2913001	Precipitation	1976–2015
RTM Kajang	2917001	Precipitation	1976–2015
Hospital Seremban	45241	Temperature	1976–2015
Ampangan Ulu Langat	44320	Temperature	1985–2015
Petaling Jaya	48648	Temperature	1976–2015

Basin are given in Table 1.

In the homogeneous test for climate time series, climatic factors are assumed to be the only factors contributing to the variations. However, changes in microenvironment, instrumentation and variations in the observation time may introduce inhomogeneity in a time series by hiding the true climatic patterns. It is essential to check the homogeneity of a long-term hydrological data record. Four absolute test methods: Standard Normal Homogeneity Test, Buishand Range Test, Pettitt Test and the Von Neumann Ratio Test were selected to verify the homogeneity of the time series.

2.2. Standardized precipitation evapotranspiration index (SPEI)

The recently developed drought index, SPEI is identified as a very appropriate tool for researching and monitoring drought conditions under warming since many researchers had conducted studies of SPEI drought analysis (Vicente-Serrano et al., 2010; Wang et al., 2014). The SPEI combines the multi-scale nature of the SPI while considering the Palmer Index sensitivity with evapotranspiration (based on temperature variation) using simple computation. The SPEI takes the resultant effects of precipitation, temperature or evapotranspiration into account in developing of the drought evolutionary stages (Beguería et al., 2014), in contrast to the Rainfall Decile Drought Index (RDDI) and the SPI that are solely assessed with rainfall data.

The computed value of SPI or SPEI indicating the drought severity based on the magnitude is such that the drought is categorised as mild if the value of SPI or SPEI ranges from 0 to -1, moderate if from -1 to -1.5, severe from -1.5 to -2 and extreme if less than -2. The category defined for the SPEI is identical with those of the SPI because they share a similarity in the computation that is based on the probability distribution (Tan et al., 2015). According to the development of SPEI by Vicente-Serrano et al. (2010), a total of 11 observations located in different parts of the world which include tropical, monsoon, Mediterranean, semi-arid, continental, cold, and oceanic climates were selected to develop the SPEI. Hence, the SPEI can be reliably used as the drought index in tropical countries, including Malaysia to represent the severity of the drought events.

2.2.1. Potential Evapotranspiration (PET)

The Potential Evapotranspiration (PET) represents the amount of moisture loss through evaporation and transpiration when there is sufficient availability of water. Many important papers have been published on this well researched subject in important journals over the past decades, and therefore suffice to say that the Thornthwaite method was adopted herein, for the estimation of the PET due to its simplicity and the data constraint.

2.2.2. Computation of SPEI

Based on the standard operating procedure mentioned in MetMalayisa, 2014, drought disasters in Malaysia is managed based on SPI index of time scales up to 6 months. Thus, the time scales of SPEI adopted in this study are 1, 3 and 6 months. The calculation of SPEI is based on the concept of deficit or surplus of water between the precipitation and the potential evapotranspiration based on the different time scales. Suffice to say here, the computation of the SPEI can be

performed based on the probability distribution $F(x)$ as expressed in Eq. (1):

$$SPEI = W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \tag{1}$$

where $W = [-2 \ln(P)]^{0.5}$ for $P \le .5$; $P = 1 - F(x)$; $C_0 = 2.515517$, $C_1 = 0.802853$, $C_2 = 0.010328$; $d_1 = 1.432788$, $d_2 = 0.189269$ and $d_3 = 0.001308$. If the value of P is greater than 0.5, then it substituted by 1-P and the sign of the final SPEI is reversed. The probability distribution $F(x)$ is calculated as presented in Eq. (2):

$$F(x) = \left[1 + \left(\frac{\alpha}{x-\gamma} \right)^\beta \right]^{-1} \tag{2}$$

where α represents the scale, β represents the shape, γ represents the origin parameters, for D values in the range ($\gamma > D < \alpha$). They can be determined using the L-moment method (Ahmad et al., 1988) with Eqs. (3)–(5).

$$\beta = \frac{2w_1 - w_0}{6w_1 - w_0 - 6w_2} \tag{3}$$

$$\alpha = \frac{(w_0 - 2w_1)\beta}{\Gamma\left(1 + \frac{1}{\beta}\right)\Gamma\left(1 - \frac{1}{\beta}\right)} \tag{4}$$

$$\gamma = w_0 - \alpha\Gamma\left(1 + \frac{1}{\beta}\right)\Gamma\left(1 - \frac{1}{\beta}\right) \tag{5}$$

where Γ is the gamma function of β and w_ι ($\iota = 0,1,2...$) can be computed by probability weighted moments (PWMs) through the L-moment method (Hosking and Wallis, 1997);
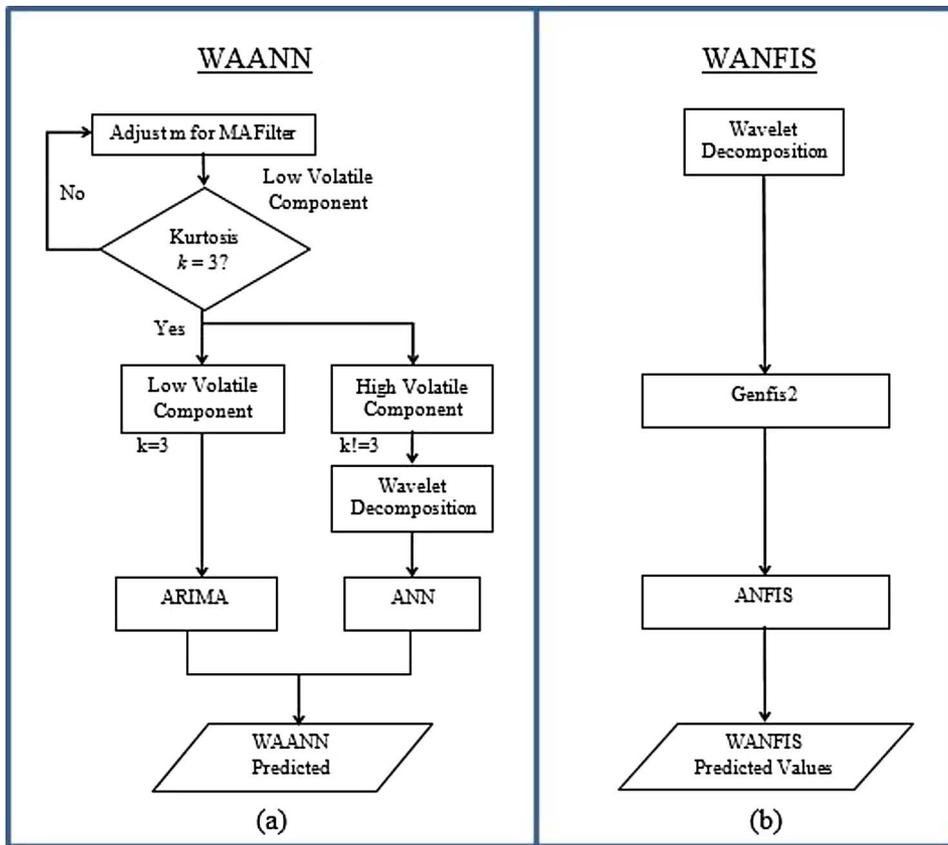
$$w_\iota = \frac{1}{n} \sum_{i=1}^{n} x_i \left(1 - \frac{i - 0.35}{n} \right)^\iota \tag{6}$$

where x_i is the ordered random sample ($x_1 < x_2... < x_n$) of D (D = Precipitation-Potential Evapotranspiration) and n represents the sample size.

2.3. ARIMA, ANN, ANFIS and discrete wavelet transform models

The ARIMA model is the stochastic model that is most widely used for both short-term and long-term drought forecasting. The ANNs have been widely used in hydrologic forecasting over the past decades. During the network training, the weights and biases in the hidden layer are adjusted to generate the target output (Kisi et al., 2016). The feed-forward neural network is one of the more popular approaches used for the time series forecasting. The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid of the neural network and fuzzy logic that applies the former for training procedure to fine tune the correlated parameter and function in a given sets of data (Shabri, 2014). The wavelet transform is defined as a time-dependent approach that decomposes the data series in the time-frequency space by providing a time-scale representation of processes and their connections. The original time series is decomposed into two components, high-frequency (high-pass) component and low-frequency (low- pass) component. Unlike the Fourier transform, the DWT localises a time series both in scaling and space and has some other favourable components where the wavelet function can be analysed more rapidly compared to the similar Fourier function (Kanika et al., 2012). The Haar wavelet was selected as the mother wavelet in this study and it was computed using Matlab®. The Haar wavelet has numerous advantages such as it is simple, fast and memory efficient (Deka et al., 2012). The historical SPEI series was used as the input for the one-dimensional wavelet transform to obtain the decomposed detail and approximation series.

2.4. Hybrid models

2.4.1. Wavelet-ARIMA-ANN (WAANN)

The development of WAANN requires the modelling of ARIMA-ANN, which is the key stage to obtain a good performance in the time series forecasting. Based on the modelling concept mentioned by Babu and Reddy (2014), the development of WAANN model is illustrated in Fig. 2(a) and the working procedure of the newly proposed WAANN in this paper can be summarised as follows.

Firstly, the length of Moving Average (MA) filter m is adjusted to generate a smoothened series y_{tr} with low-volatility as presented in Eq. (7):

$$y_{tr} = \frac{1}{m} \sum_{i=t-m+1}^{t} y_i \qquad (7)$$

Next, the high-volatility time series y_{res} can be obtained as the residual of the time series by using Eq. (8):

$$y_{res} = y_t - y_{tr} \qquad (8)$$

The obtained y_{tr} and y_{res} series were then used to develop the ARIMA and ANN, respectively.

For the modelling of ARIMA, the obtained y_{tr} series (low-volatility) from previous step was used as the input and modelled involving three stages known as model identification, estimation and diagnostic check. During the model identification, the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series were plotted to determine the significance lag. The selection of model was based on the lowest Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC) as shown in Eqs. (9) and (10):

$$AIC = -2logL + 2m \qquad (9)$$

$$SBC = -2logL + m\ ln(n) \qquad (10)$$

where m is the number of terms estimated in the model; L is the likelihood function of the ARIMA models and n is the number of observations.

For ANN, the generated y_{res} series was first put through wavelet decomposition before it was used as the input for the model's development. The feedforward multilayer perceptron was proposed as it is very effective and simple to compute for the hydrological forecasting. Firstly, the decomposed input data was normalised before the modelling of ANN as expressed in Eq. (11) (Mishra et al., 2007):

$$X_n = \frac{X_o - X_{min}}{X_{max} - X_{min}} \qquad (11)$$

where X_n is the normalised data; X_o is the observed data; X_{max} is the maximum value among the observed data and X_{min} is the minimum value among the observed data. The training algorithms, Levenberg Marquardt backpropagation (trainlm) and Bayesian regularisation backpropagation (trainbr) were selected to identify the most appropriate training algorithm for the SPEI time series. The training stage was conducted using 70% of the data samples; the next 15% was used in the testing stage, whilst the remaining 15% was used at the validation stage. The optimum number of lag and hidden neuron was chosen based on trial and error.

Finally, the predicted time series of WAANN model was generated using the results from ARIMA and ANN, by obtaining the sum of the series from both models.

2.4.2. Wavelet-Adaptive Neuro-Fuzzy Inference System (WANFIS)

The coupling of wavelet transformation and Adaptive Neuro-Fuzzy Inference System (ANFIS) in the field of hydrological forecasting is considered a relatively new approach in recent years. The combination of these two methods is known as WANFIS and is a better model compared to the stand-alone ANFIS. In this paper, a Sugeno-type FIS was adopted due to its effectiveness in computation and performs well with optimisation and adaptive techniques. The development of the WANFIS is shown in Fig. 2(b) and the working procedure can be

described as follows. Firstly, the decomposed historical SPEI series was normalised and used as the input variables for the development of fuzzy interference system (FIS) in the genfis2 (ANFIS with subtractive clustering). The radii in the genfis2 were adjusted to obtain more accurate results (Kisi et al., 2015). The ANFIS algorithm was then applied to the trained FIS to enhance the performance of the model. The number training epoch and initial step size in the ANFIS algorithm were determined through trial and error. The predicted outputs were then generated based on the *evalfis* algorithm in the Matlab® for the purpose of validation.

2.5. Model's performance evaluation

After modelling of the WAANN and the WANFIS models, the performance of both models in forecasting the monthly SPEI was statistically assessed with several key statistical parameters. The parameters suggested included the Root-Mean-Square-Error (RMSE), Mean Absolute Error (MAE), Willmott's Index of Agreement (d) and Nash-Sutcliffe Coefficient of Efficiency (E). The Adjusted Coefficient of Determination (R^2_{adj}) was adopted to evaluate the model performance in this study as well. The calculation of RMSE, MAE, d, E and R^2_{adj} are shown in Eqs. (12)–(17), respectively:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{n} \left(SPEI_{p_i} - SPEI_{o_i} \right)_t^2} \tag{12}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{n} |\left(SPEI_{p_i} - SPEI_{o_i} \right)_t| \tag{13}$$

$$d = 1 - \left(\frac{\sum_{i=1}^{n} (SPEI_{o_i} - SPEI_{p_i})^2}{\sum_{i=1}^{n} (|SPEI'_{p_i} - \overline{SPEI_o}| + |SPEI'_{o_i} - \overline{SPEI_o}|)^2} \right), 0 \leqslant d \leqslant 1 \tag{14}$$

$$E = 1 - \left[\frac{\sum_{i=1}^{n} (SPEI_{o_i} - SPEI_{p_i})^2}{\sum_{i=1}^{n} (SPEI_{o_i} - \overline{SPEI_o})^2} \right], 0 \leqslant E \leqslant 1 \tag{15}$$

$$R^2 = \left(\frac{\sum_{i=1}^{n} (SPEI_{o_i} - \overline{SPEI_{o_i}})(SPEI_{p_i} - \overline{SPEI_{p_i}})}{\sqrt{\sum_{i=1}^{n} (SPEI_{o_i} - \overline{SPEI_{o_i}})^2 (SPEI_{p_i} - \overline{SPEI_{p_i}})^2}} \right)^2 \tag{16}$$

$$R^2_{adj} = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \tag{17}$$

where N is the number of samples and p is the number of regression coefficients (including the intercept).

A good model should have a lower RMSE and MAE which indicates low-accumulated errors. Besides, the values of R^2_{adj}, d and E should be close to the unity for a better fitting model. All these five parameters normally are closely related to each other where if the RMSE is low, the MAE tends to be low and the R^2_{adj}, d and E are found to be closer to unity.

3. Results and discussion

3.1. Homogeneity tests

The calibration of the homogeneity test was: firstly, the time series was divided into twelve months (from January until December) and then each month of the dataset was tested for homogeneity. The *p*-value for all the selected tests should be greater than the 5% significance

value to indicate that the time series is homogeneous in a strict sense. The four tests were tested with the null hypothesis where the data were assumed homogeneous. The results were classified into three categories including class 'useful', 'suspect' and 'doubtful'. The 'useful' categorization was when none or one out of four tests rejected the null hypothesis. When two of four tests rejected the null hypothesis, it was categorised as 'suspect'. On the other hand, the dataset was classified as 'doubtful' once three or four tests had rejected the null hypothesis.

Two rainfall stations, namely the Pejabat JPS Sg. Manggis station (2815001) and the P/KWLN P/S Telok Gong station (2913001) were classified as 'useful' due to none of four tests rejected the null hypothesis. The useful dataset of the RTM Kajang station (2917001) was 91.67% with only 8.33% was considered as 'suspect'. In comparison with the rainfall stations, the datasets of all temperature stations exhibited relatively poor homogeneity. Approximate 91.67% of data of the Hospital Seremban station (45241) was categorised as 'doubtful'. Similarly, the Ampangan Ulu Langat station (44320) had 50% of data considered as 'doubtful' and 16.67% in class 'suspect' and 33.33% as class 'useful'. On the other hand, the data of the Petaling Jaya station (48648) showed a better homogeneity with about 42% 'useful' and 33.33% 'doubtful'. Therefore, the Petaling Jaya station (48648) was the only station used in this study for the computation of historical Potential Evapotranspiration (PET) and SPEI.

3.2. Model development

After the computation of historical PET and SPEI, the SPEI time series was decomposed with the level of decomposition 3. The level of decomposition, L is determined based on the number of the dataset, N and determined by the equation $L = int[log(N)]$ (Tiwari and Chatterjee, 2010). The period of record was from 1976 to 2015 ($N = 480$), thus, the level of decomposition, $L = 3$.

The wavelet transform decomposed the time series into four components including A3, D1, D2 and D3 which were used as the inputs for ANN and ANFIS models. The development of WAANN and WANFIS required the modelling of ANN and ANFIS models. Both models used the historical SPEI from 1976 to 2007 as the training dataset. Modelling of the WANFIS models with 3-months lag showed a superior performance with high prediction accuracy. Increasing the lag beyond 3 months did not improve the performance of the WANFIS. While the significance lag of ANN varied with different SPEI series and it was determined via trial and error. Since each monthly SPEI had decomposed into four components through wavelet transformation, the total inputs became multiple of four with the lag to predict the one-step ahead value. Table 2 lists the network architecture of the best ANN and ANFIS models.

Table 2 indicates that the training algorithm for all the feedforward neural network models was Bayesian regularization backpropagation (trainbr) with the training epoch equal to 1000. This training function applied the Levenberg-Marquardt optimization to propagate the weight and bias data. The algorithm has the ability to train the network in the difficult and complex situation and provides early stopping to prevent overfitting. The hidden neurons of the ANN model were determined by trial and error as different time series fit better with a different number of hidden neurons. During the simulation, the optimum hidden neurons were identified with the best-measured goodness of fit (MAE and RMSE). Similar to the findings of Djerbouai and Souag-Gamane (2016), the wavelet decomposition reduced the SPEI series complexity and thus, a small number of the hidden neurons (between 2 and 4) was sufficient in this study to obtain the desired outputs in ANN. Increasing the number of hidden neurons beyond the optimum number may cause a deterioration of the network performance due to overfitting. Overfitting occurs when the training performance exhibits an excellent result but incapable to accurately predict the output in the validation stage as the network had memorised the training input data and unable to simulate the output of unknown data. Similar to the modelling of

Table 2
Network architecture of the best ANN and ANFIS models.

Stations	ID	Time Scale	ANN					ANFIS		
			Network architecture	Training algorithm	Hidden transfer function	Output transfer function		Radii	Epoch	Initial step size
Pejabat JPS Sg. Manggis	2815001	SPEI-1	28-3-1	Trainbr	Tansig	Linear		0.9	40	0.3
		SPEI-3	32-4-1	Trainbr	Tansig	Linear		1	40	0.3
		SPEI-6	20-4-1	Trainbr	Tansig	Linear		1	50	0.3
P/KWLN P/S Telok Gong	2913001	SPEI-1	28-2-1	Trainbr	Tansig	Linear		1	20	0.3
		SPEI-3	28-3-1	Trainbr	Tansig	Linear		1	20	0.3
		SPEI-6	32-3-1	Trainbr	Tansig	Linear		0.9	20	0.3
RTM Kajang	2917001	SPEI-1	28-2-1	Trainbr	Tansig	Linear		0.7	20	0.4
		SPEI-3	20-3-1	Trainbr	Tansig	Linear		0.8	20	0.3
		SPEI-6	28-3-1	Trainbr	Tansig	Linear		1	20	0.3

ANFIS, overfitting can be avoided by altering the number of the epoch to eliminate overtraining.

For the development of the hybrid model WAANN, the historical SPEI was separated into high volatility and low volatility components. The low volatility component was modelled with a linear ARIMA model which involved in three stages; known as model identification, estimation and diagnostic check.

During the model identification, the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series were plotted to determine the significance lag as shown in Fig. 3. Fig. 3 describes the significance lag of ACF at 1, 2, 3, 4, 5, 9, 10, 11, 12, 13 and 14; while the PACF had the significance lag at 1, 2, 3, 4, 5, 6 and 8. Hence, the ARIMA $(p,0,q)$ had a possible combination of $p = 1$ to 8 and $q = 1$ to 14. The best ARIMA model was distinguished based on the lowest Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC) as tabulated in Table 3. It was found that the ARIMA model fitted well for the RTM Kajang station (2917001) where all the predicted SPEI had low AIC and SBC compared to other stations. Among the different time scales studied, the best model was found at the RTM Kajang station (2917001) in forecasting the SPEI-6 with AIC = −2133.50 and SBC = −2096.42. In contrast, the lowest prediction accuracy of the model was found at the P/KWLN P/S Telok Gong station (2913001) in predicting the SPEI-1 with AIC = −420.63 and SBC = −366.51.

The overall performance of all the ARIMA models were considered good due to the separation of data by the moving average filter. With the kurtosis value close to 3 (Babu and Reddy, 2014), all the ARIMA models were able to forecast the SPEI time series with a high degree of accuracy despite the time scale (1, 3 and 6-months) of the series.

After model identification, the parameters of autoregressive (AR) and moving average (MA) were identified in the estimation stage. For the diagnostic check, it requires fulfilling the assumption of Box-Jenkins' theory where the residuals should be independent, homoscedastic and normally distributed in order to validate the performance of the ARIMA model. The first step conducting the diagnostic check of the models is to analyse the residuals from the ACF and PACF. Fig. 4 illustrates a good plot of Residual Autocorrelation Function (RACF) and Residual Partial Autocorrelation Function (RPACF) for ARIMA (8, 0, 14) for the SPEI-6 prediction at the Pejabat JPS Sg. Manggis station (2815001); as an example. The ARIMA model is considered well-fitted with the time series as the RACF and RPACF lie within the confidence boundaries, which clearly indicates the residuals are white noise and independent.

Next, the Breusch-Pagan test was used to test for the homoscedasticity of residuals. As shown in Table 4, among the stations studied, the P/KWLN P/S Telok Gong station (2913001) is the only station that failed the test with the p-value less than 5% of significance value. The results showed that the rest of the SPEI residuals were homoscedastic and constant which fulfilled the assumption of Box-Jenkins'
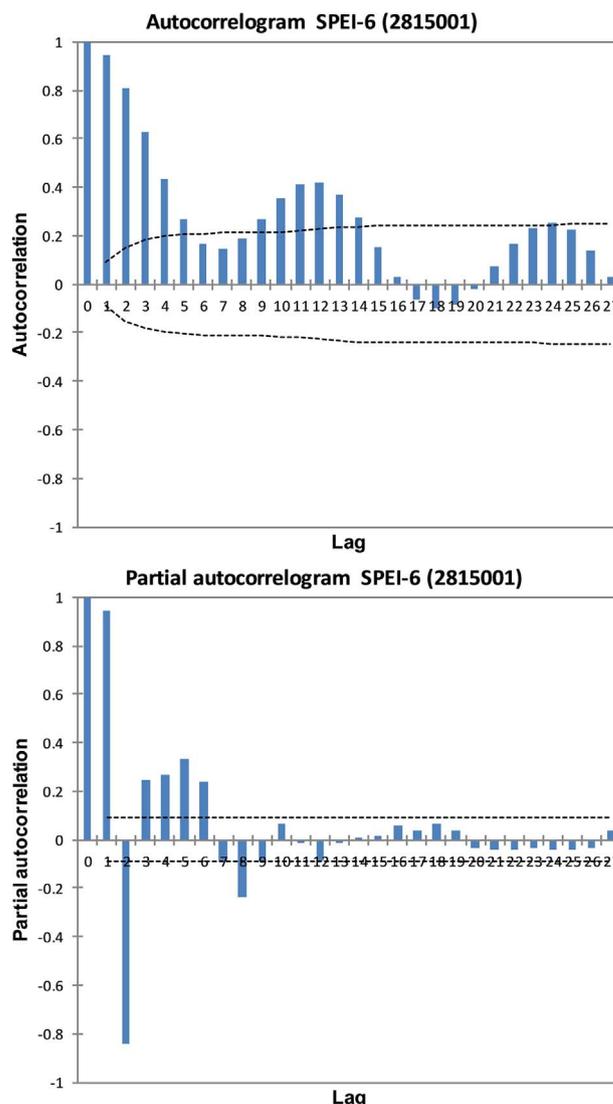

Fig. 3. ACF and PACF for SPEI-6 of Station Pejabat JPS Sg. Manggis (2815001).

theory.

The last step in the diagnostic check requires testing for the normality of standardised residual based on the Shapiro-Wilk test, the Anderson-Darling test and the Jarque-Bera test with 5% significance value as listed in Table 5. All the SPEI series for the Pejabat JPS Sg. Manggis station (2815001) and the RTM Kajang station (2917001) passed all the three tests, which indicated the assumption for the linear

Table 3
Best ARIMA models with AIC and SBC.

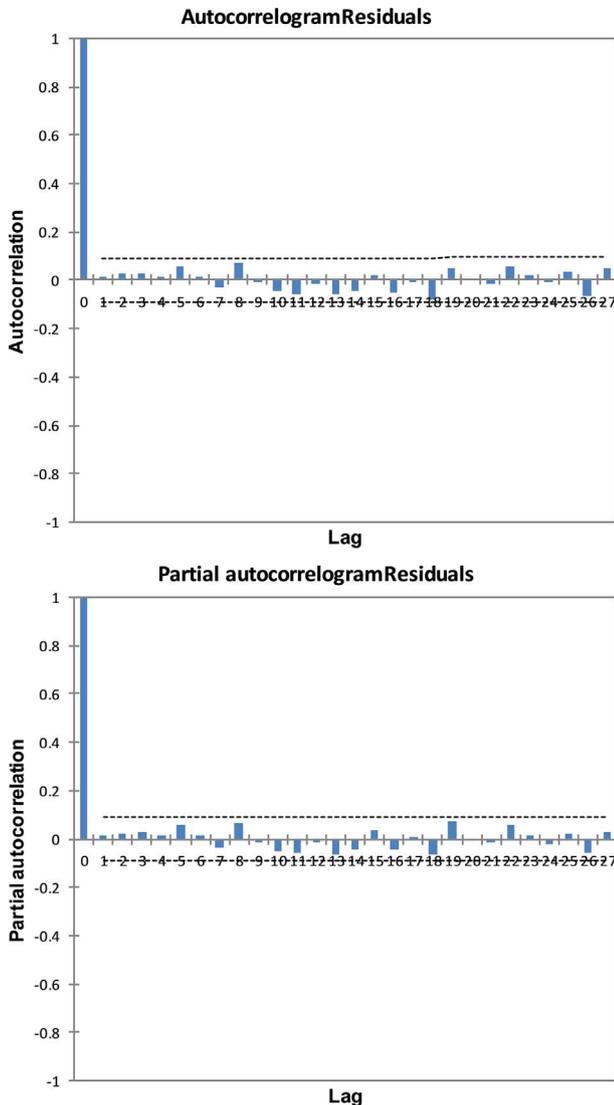| Stations	ID	Time Scale	Model	m	Kurtosis	AIC	SBC
Pejabat JPS Sg. Manggis	2815001	SPEI-1	ARIMA (0, 0, 13)	7	2.917	−526.01	−467.78
		SPEI-3	ARIMA (15, 0, 14)	7	2.784	−1189.59	−1064.88
		SPEI-6	ARIMA (8, 0, 14)	5	2.631	−1264.13	−1168.57
P/KWLN P/S Telok Gong	2913001	SPEI-1	ARIMA (6, 0, 6)	6	2.961	−420.63	−366.51
		SPEI-3	ARIMA (15, 0, 6)	8	3.022	−1186.62	−1095.22
		SPEI-6	ARIMA (8, 0, 7)	8	3.014	−1506.48	−1440.14
RTM Kajang	2917001	SPEI-1	ARIMA (13, 0, 12)	22	2.991	−1323.01	−1215.65
		SPEI-3	ARIMA (11, 0, 10)	21	2.999	−1750.77	−1659.98
		SPEI-6	ARIMA (3, 0, 5)	21	3.009	−2133.50	−2096.42

Fig. 4. RACF and RPACF for ARIMA (8, 0, 14) for SPEI-6 of Station Pejabat JPS Sg. Manggis (2815001).

regression of ARIMA model was fully fulfilled. Meanwhile, the results depicted that the SPEI-6 for the P/KWLN P/S Telok Gong station (2913001) failed the Anderson-Darling test. However, with the passing of the other two tests, it is still assumed that the linear regression of ARIMA was fully fulfilled.

3.3. Models performance evaluation

The performance of both WAANN and WANFIS models were

Table 4
Breusch-Pagan Test for Residuals Homoscedasticity.

Stations	ID	Time scale	p-value	Remark
Pejabat JPS Sg. Manggis	2815001	SPEI-1	0.577	Homoscedastic
		SPEI-3	0.892	Homoscedastic
		SPEI-6	0.987	Homoscedastic
P/KWLN P/S Telok Gong	2913001	SPEI-1	0.301	Homoscedastic
		SPEI-3	0.634	Homoscedastic
		SPEI-6	0.041	Heteroscedastic
RTM Kajang	2917001	SPEI-1	0.932	Homoscedastic
		SPEI-3	0.875	Homoscedastic
		SPEI-6	0.722	Homoscedastic

Table 5
Normality Test for Standardised Residuals.

Stations	ID	Time scale	Shapiro-Wilk	Anderson-Darling	Jarque-Bera
Pejabat JPS Sg. Manggis	2815001	SPEI-1	0.102	0.080	0.074
		SPEI-3	0.568	0.618	0.565
		SPEI-6	0.738	0.741	0.456
P/KWLN P/S Telok Gong	2913001	SPEI-1	0.265	0.118	0.193
		SPEI-3	0.476	0.739	0.637
		SPEI-6	0.095	0.036	0.163
RTM Kajang	2917001	SPEI-1	0.269	0.918	0.572
		SPEI-3	0.340	0.464	0.082
		SPEI-6	0.079	0.090	0.076

compared with the statistical measures including the Adjusted Coefficient of Determination (R^2_{adj}), Root-Mean-Square-Error (RMSE), Mean Absolute Error (MAE), Willmott's Index of Agreement (d) and the Nash-Sutcliffe Coefficient of Efficiency (E) as shown in Table 6. The performance validation was carried out over the 96 months' test period (2008–2015) for all the SPEI series (SPEI-1, -3 and -6).

As illustrated in Table 6, the prediction from WAANN model for the SPEI-6 of the RTM Kajang station (2917001) has the best results, with the highest R^2_{adj} (0.9603), d (0.9896) and E (0.9600) and the lowest MAE (0.1364) and RMSE (0.1810). In contrast, the worst prediction was found in the WANFIS model for the SPEI-1 of the Pejabat JPS Sg. Manggis station (2815001) (R^2_{adj} = 0.4994, MAE = 0.5090, RMSE = 0.7582, d = 0.8296 and E = 0.4916). Clearly, as the time scale increases from SPEI-1 to SPEI-6, the performance of both WAANN and WANFIS improved significantly as a higher SPEI would reduce the white noise by increasing the filter length (Mishra and Desai, 2005). Consequently, it is interesting to note that both WAANN and WANFIS models can predict the higher time scale SPEI more accurately due to less fluctuation.

In terms of forecasting ability for SPEI-1, the WAANN model performed better than the WANFIS model for the Pejabat JPS Sg. Manggis station (2815001) where higher R^2_{adj}, d and E with lower MAE and

Table 6
Statistical measures of the WAANN and WANFIS performance for one month lead time.

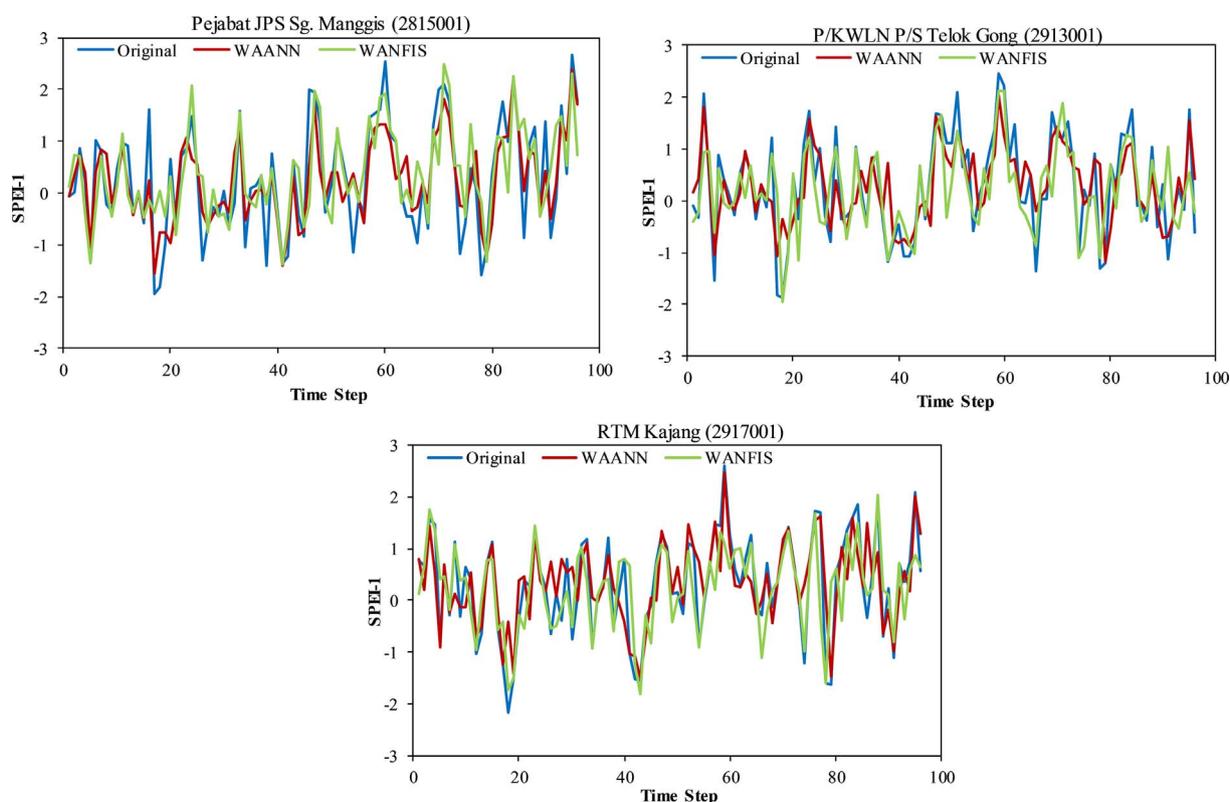| Stations	ID	Time Scale	Models	R^2 (adj)	MAE	RMSE	d	E
Pejabat JPS Sg. Manggis	2815001	SPEI-1	WAANN	0.7083	0.4313	0.5825	0.8955	0.6999
			WANFIS	0.4994	0.5090	0.7582	0.8296	0.4916
		SPEI-3	WAANN	0.9226	0.2183	0.2827	0.9801	0.9230
			WANFIS	0.8242	0.2797	0.4311	0.9525	0.8210
		SPEI-6	WAANN	0.9408	0.1730	0.2493	0.9847	0.9404
			WANFIS	0.8891	0.2328	0.3435	0.9711	0.8868
P/KWLN P/S Telok Gong	2913001	SPEI-1	WAANN	0.5428	0.4664	0.6458	0.8380	0.5470
			WANFIS	0.5874	0.4425	0.6216	0.8629	0.5802
		SPEI-3	WAANN	0.8948	0.2033	0.2994	0.9726	0.8932
			WANFIS	0.7979	0.2802	0.4123	0.9424	0.7973
		SPEI-6	WAANN	0.9545	0.1517	0.2020	0.9884	0.9550
			WANFIS	0.9133	0.2124	0.2792	0.9771	0.9140
RTM Kajang	2917001	SPEI-1	WAANN	0.6007	0.3895	0.6011	0.8702	0.6032
			WANFIS	0.6071	0.4012	0.6017	0.8722	0.6024
		SPEI-3	WAANN	0.8663	0.2400	0.3487	0.9640	0.8667
			WANFIS	0.7901	0.2636	0.4381	0.9412	0.7896
		SPEI-6	WAANN	0.9603	0.1364	0.1810	0.9896	0.9600
			WANFIS	0.8619	0.2458	0.3445	0.9587	0.8550

Fig. 5. Observed and Predicted SPEI-1.

RMSE were achieved. Besides, the WAANN obtained a lower MAE and RMSE and higher E in predicting the SPEI-1 for the RTM Kajang station (2917001). However, the WANFIS model outperformed the WAANN model in predicting the SPEI-1 for the P/KWLN P/S Telok Gong station (2913001). Hence, the performance of the models in forecasting the highly sensitive SPEI-1 depends on the dynamic properties and complexity of the time series which vary from station to station. As the error obtained from the ARIMA model was considerably small, thus, the prediction error obtained from the ANN became predominant for the WAANN model's performance. Overall, both WAANN and WANFIS models for SPEI-1 exhibited the largest value of errors (MAE and RMSE) compared to the other SPEIs due to the models being less effective in detecting the white noise under a short lead time. Hence, the SPEI-1

was a less representative time scale to use in the drought forecasting as it was extremely sensitive and unstable.

The results showed that with the performance measures for SPEI-3 and SPEI-6, using the WAANN model resulted in better prediction accuracy. It is observed that the WAANN greatly outperformed the WANFIS for all the stations to predict the future SPEI-3 and -6. This is probably because of the SPEI-3 and -6 were less fluctuating and contained lesser white noise compared to the SPEI-1. Therefore, the performance of the ANN model improved significantly and subsequently enhanced the performance of WAANN model. In contrast, the weakness of WANFIS model was unable to predict the output reliably when the input signals had fallen outside of the values of the training dataset. The overall result is the WAANN model outperformed the WANFIS model in
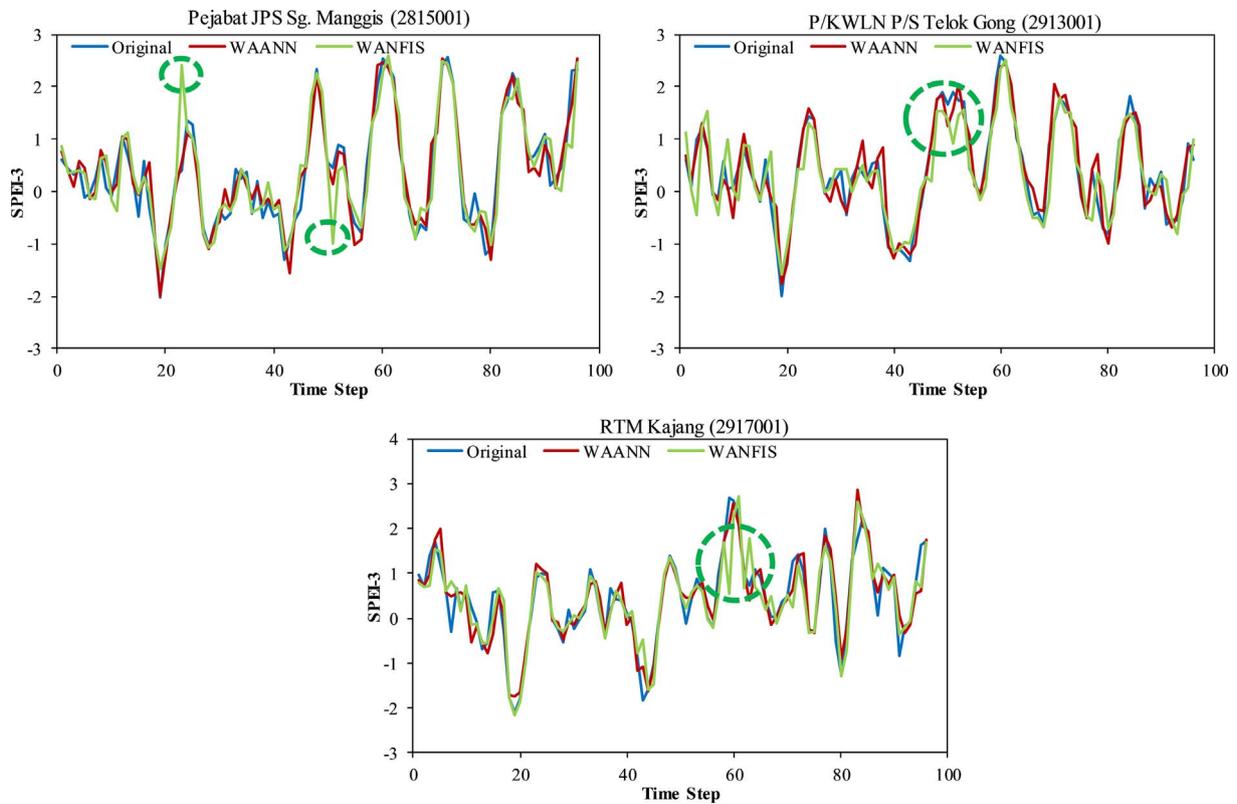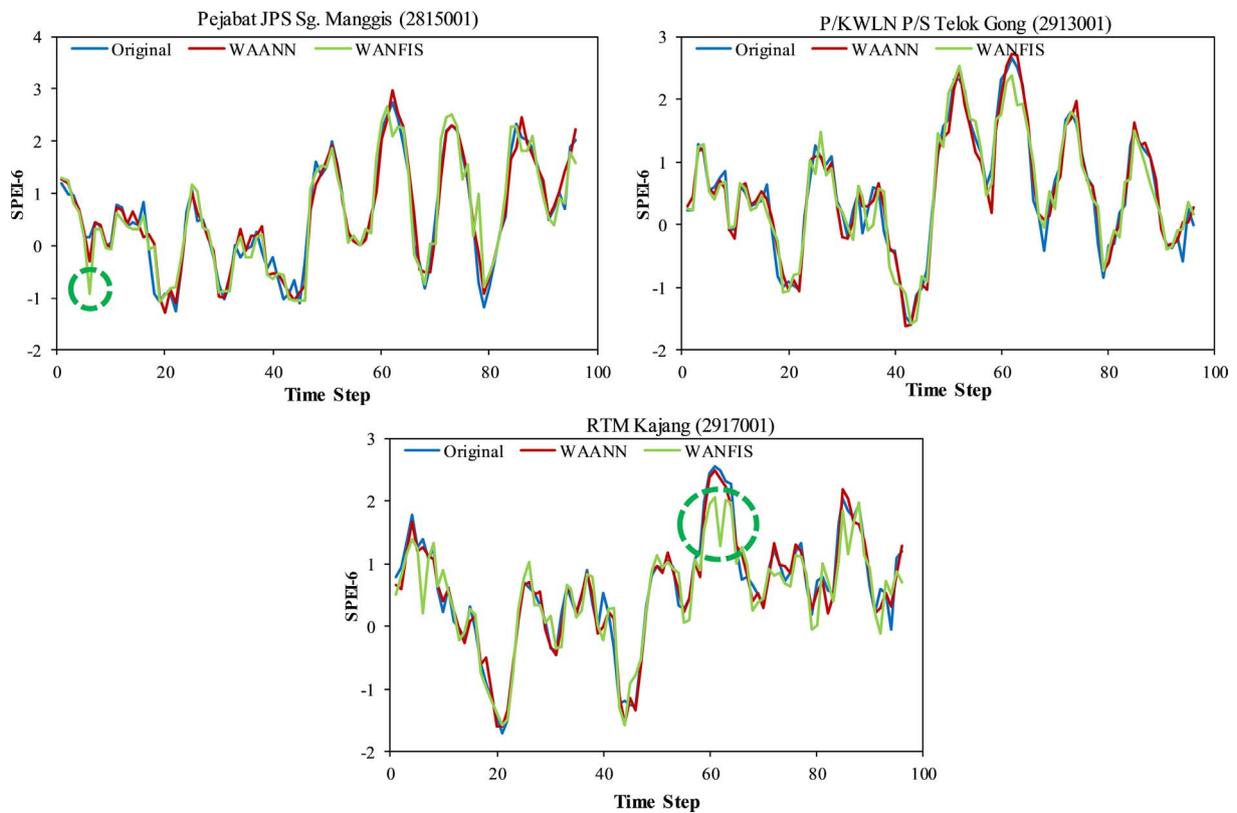
Fig. 6. Observed and Predicted SPEI-3.

Fig. 7. Observed and Predicted SPEI-6.

a reliable fashion, in forecasting the SPEI-3 and SPEI-6 for all the stations.

Figs. 5–7 illustrate the observed and predicted SPEI-1, SPEI-3 and SPEI-6 graphical plots respectively, for all the stations. The extreme outliers of the predicted SPEI for WAANN and WANFIS were identified except for the highly fluctuating SPEI-1. Most of the outliers were found in the time series for SPEI-3 and SPEI-6 especially for the Pejabat JPS Sg. Manggis station (2815001) and the P/KWLN P/S Telok Gong station (2913001). By observing the plots, the predicted values of WANFIS yielded more extreme outliers compared to the WAANN. Hence, the performance of the WANFIS was relatively lower than the WAANN in SPEI-3 and SPEI-6 forecasting. The graphical results shows the limitation of the WANFIS model when the input data were outside of the training input range, whence the prediction accuracy deteriorated.

4. Conclusions

Two hybrid models, namely the WAANN and WANFIS were selected to forecast the future SPEI at the Langat River Basin, Malaysia. The input data for development of both models was pre-processed with wavelet transform to enhance the performance of the models. Based on the key statistical parameters, the performance of both WAANN and WANFIS models improved due to the reduction of white noise by increasing the filter length, when the time scale increased. For the prediction of SPEI-1, the errors obtained by both WAANN and WANFIS models were considered relatively high compared to SPEI-3 and SPEI-6 because both models were less effective in detecting the white noise for highly fluctuating SPEI-1. It was also found that the prediction accuracy of the WAANN was superior to the WANFIS for the SPEI-3 and SPEI-6 prediction for all the stations. In predicting the SPEI-3 and SPEI-6, both WAANN and WANFIS models performed very well since the white noise and outliers were greatly reduced. Nevertheless, the ability of the WAANN model in predicting the SPEI-3 and SPEI-6 was superior to the WANFIS model at all the stations. Apart from that, the WANFIS prediction values yielded more extreme outliers compared to the WAANN prediction values also showed the limitation of the WANFIS model where the prediction accuracy declined when the input data were outside the training input range. Finally, the WAANN model is the better drought forecasting model and it is not limited by the training input range and provided accurate predictions for both short-term and mid-term drought conditions.

Acknowledgments

References

Ahmad, M.I., Sinclair, C.D., Werritty, A., 1988. Log-logistic flood frequency analysis. J. Hydrol. 98, 205–224.

Babu, C.N., Reddy, B.E., 2014. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Appl. Soft Comput. 23, 27–38.

Bachmir, S., Svensson, C., Hannaford, J., Barker, L.J., Stahl, K., 2016. A quantitative analysis to objectively appraise drought indicators and model drought impacts. Hydrol. Earth Syst. Sci. 20, 2589–2609.

Bazrafshan, O., Salajegheh, A., Bazrafshan, J., Mahdavi, M., Marj, A.F., 2015. Hydrological drought forecasting using ARIMA models (Case Study : Karkheh Basin). Ecopersia 3, 1099–1117.

Beguería, S., Vicente-Serrano, S.M., Reig, F., Latorre, B., 2014. Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int. J. Climatol. 34 (10), 3001–3023.

Belayneh, A., Adamowski, J., 2012. Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. Appl. Comput. Intell. Soft Comput. 2012, 1–13.

Belayneh, A., Adamowski, J., 2013. Drought forecasting using new machine learning methods. J. Water Land Dev. 18 (I–VI), 3–12.

Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. J. Hydrol. 508, 418–429.

Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., Ahmad, S., Attarod, P., 2014. Drought forecasting in a semi-arid watershed using climate signals: a Neuro-fuzzy modeling approach. J. Mount. Sci. 11 (6), 1593–1605.

Choubin, B., Malekian, A., Golshan, M., 2016. Application of several data-driven techniques to predict a standardized precipitation index. Atmosfera 29 (2), 121–128.

Deka, P.C., Haque, L., Banhatti, A.G., 2012. Discrete wavelet-Ann approach in time series flow forecasting-a case study of Brahmaputra river. IJEE 5 (4), 673–685.

Djerbouai, S., Souag-Gamane, D., 2016. Drought forecasting using neural networks, wavelet neural networks, and stochastic models: case of the Algerois Basin in North Algeria. Water Resour. Manage. 30 (7), 2445–2464.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis – An Approach Based on L-Moments. Cambridge University Press, Cambridge.

Jalalkamali, A., Moradi, M., Moradi, N., 2015. Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. Int. J. Environ. Sci. Technol. 12 (4), 1201–1210.

Kanika, E., Dhillon, N., Sharama, E.K., 2012. Comparative analysis of discrete wavelet transform and fast wavelet transform on image compression. Int. J. Eng. Res. Technol. 1 (5), 1–7.

Khashei, M., Bijari, M., 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Appl. Soft Comput. 11 (2), 2664–2675.

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Niazi, F., 2015. Long-term monthly evapotranspiration modeling by several data-driven methods without climatic data. Comput. Electron. Agric. 115, 66–77.

Kisi, O., Genc, O., Dinc, S., Zounemat-Kermani, M., 2016. Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree. Comput. Electron. Agric. 122, 112–117.

Masinde, M., 2014. Artificial neural networks models for predicting effective drought index: factoring effects of rainfall variability. Mitig. Adapt Strateg. Glob. Change. 19 (8), 1139–1162.

Malaysia Meteorological Department, 2014. Country Reports – Malaysia. MetMalayisa, Selangor.

Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. Stoch. Environ. Res. Risk A 19 (5), 326–339.

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391 (1), 202–216.

Mishra, A.K., Desai, V.R., Singh, V.P., 2007. Drought forecasting using a hybrid stochastic and neural network model. J. Hydrol. Eng. 12 (6), 626–638.

Mossad, A., Alazba, A.A., 2015. Drought forecasting using stochastic models in a hyper-arid climate. Atmosphere 6 (4), 410–430.

Özger, M., Mishra, A.K., Singh, V.P., 2011. Estimating Palmer Drought Severity Index using a wavelet fuzzy logic model based on meteorological variables. Int. J. Climatol. 31 (13), 2021–2032.

Özger, M., Mishra, A.K., Singh, V.P., 2012. Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in Texas. J. Hydrometeorol. 13 (1), 284–297.

Paz, J.O., Fraisse, C.W., Hatch, L.U., y Garcia, A.G., Guerra, L.C., Uryasev, O., Bellow, J.G., Jones, J.W., Hoogenboom, G., 2007. Development of an ENSO-based irrigation decision support tool for peanut production in the southeastern US. Comput. Electron. Agric. 55 (1), 28–35.

Pour, S.H., Harun, S.B., Shahid, S., 2014. Genetic programming for the downscaling of extreme rainfall events on the East Coast of Peninsular Malaysia. Atmosphere 5 (4), 914–936.

Seo, Y., Kim, S., Kisi, O., Singh, V.P., Parasuraman, K., 2016. River stage forecasting using wavelet packet decomposition and machine learning models. Water Resour. Manage. 30 (11), 4011–4035.

Shabri, A., 2014. A hybrid wavelet analysis and adaptive Neuro-Fuzzy inference system for drought forecasting. Appl. Math. Sci. 8 (139), 6909–6918.

Subash, N., Mohan, H.R., Banukumar, K., 2011. Comparing water-vegetative indices for rice (Oryza sativa L.)–wheat (Triticum aestivum L.) drought assessment. Comput. Electron. Agric. 77 (2), 175–187.

Tan, C.P., Yang, J.P., Li, M., 2015. Temporal-spatial variation of drought indicated by SPI and SPEI in Ningxia Hui autonomous region, China. Atmosphere 6 (10), 1399–1421.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar Drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Climate 23, 1696–1718.

Wang, Q.F., Shi, P.J., Lei, T.J., Geng, G.P., Liu, J.H., Mo, X.Y., Li, X.H., Zhou, H.K., Liu, D.C., 2014. Temporal-spatial characteristics of severe drought events and their impact on agriculture on a global scale. Quatern. Int. 349, 10–21.

Yusof, F., Foo, H.M., Suhaila, S., Kong, C.Y., 2012. Trend analysis for drought event in peninsular Malaysia. Jurnal Teknologi (Sci. Eng.) 57 (1), 211–218.