International Journal of u- and e- Service, Science and Technology
Vol.9, No. 3 (2016), pp.53-66
http://dx.doi.org/10.14257/ijunesst.2016.9.3.07

Drought Identification, Monitoring and Forcasting for Selangor River Basin

Daniel Hong and Kee An Hong

UCSI University, Hong and Associates
danielhong10101994@hotmail.com, keeanhong@yahoo.com

Abstract

The severe 2014 drought recorded in the Selangor river basin has affected the everyday life of three million people inhabited in the northern area of Selangor and the neighboring federal capital of Malaysia, the city of Kuala Lumpur, where 70% of the source of water supply comes from Selangor dam in the upper reach of Selangor basin. Of particular importance is the water rationing imposed by the water authority in April 2014 lasting for one month and the shortage of food supply in the dry period as a result of reduction in food supply from the Selangor area. As such, drought monitoring, identification and forecasting play an important role in the planning and management of natural resources and water resource systems in the country. The purpose of this paper is to use established scientific methods and available hydrological data to identify, monitor, and forecast droughts for the planning, management and formulating drought strategies to reduce and mitigate the adverse effect of drought impacts. Standardized precipitation index (SPI) has been used as a conventional tool to identify and monitor drought occurrences. To achieve the aims, we use average long term monthly rainfall data for eight stations covering both the dry and wet seasons from Selangor river basin to derive the SPI values for durations of 3 to 9 months. These drought indicators, which are time series derived from rainfall data together with the multi-layer artificial neural networks model were used for drought forecasting for the basin. Forecasting were made for SPI with a one month ahead lead time as forecasting accuracy is reduced for longer lead times. This has been shown by studies carried out elsewhere. Our finding indicates that more accurate predictions are achieved using SPI of longer durations, i.e. 6 and 9 months. This is consistent with findings of studies by others.

Keywords: *Drought, Neural Network*

1. Introduction

Essentially, drought occurrence is due to the shortage of water, that is, in a period of time when water availability is less than a specified amount at a particular place, drought occurs. Drought is considered as a natural disaster that can have great impacts on regular human activities. An example is the 2014 drought occurred in Selangor basin, the drought is felt both economically and socially. Consequences are the imposition of water rationing due to the lack of water for human consumption and the losses in industrial sector and reduction in food productions.

Drought is a natural phenomenon which can be short, lasting only a few months, or it can be persistent for years before the climatic conditions return to normal. Drought is difficult to identify and quantify as its occurrence is always hard to detect. Therefore, an effective drought monitoring and forecasting system is an important tool for formulating mitigation strategies. Basically, the needs are the evaluation of current drought condition and forecasting of future drought occurrence in an area like Selangor. Traditionally, drought indices are normally used to evaluate and forecast drought occurrence. Drought indices like the standardized precipitation index (SPI), the Palmer index and the crop

ISSN: 2005-4246 IJUNESST

moisture index are commonly used for this purpose. For this study, the SPI method has been used to derive drought index as this method has the following advantages: (1) SPI is standardized and ensures independence from geographical position as the index in question is calculated with respect to the average precipitation in the same place. (2) the said characteristic makes the SPI useful as a primary drought index because it is simple , spatially invariant in its interpretation and probabilistic nature allow it to be used in risk and decision making analysis. For drought forecasting, we use the simple multi-layer perceptron (MLP) artificial neural network model with SPI as input time series for the model.

2. Materials and Methods

2.1. The Study Area

Records of rainfall stations of the Selangor river basin were used for this drought analysis. Figure 1 shows a location map of Selangor basin up to the Public Works Department (PWD) water intake point. Selangor river basin up to the intake point has an area of 1450 km² and the maximum length and width of the basin are 48 km and 39 km respectively. About 30% of the basin is steep mountainous country above 600 m, 38% is in hilly country and the remainder undulating low terrain. A large portion (two-thirds) of the basin is under jungle and the remainder under rubber, oil palm, paddy, maize and vegetable cultivation. In the eastern half fine to coarse granite and other allied rocks are found and sandstone is found in the western half of the basin. Wet seasons occur in April and May in the south west monsoon season and October to December in the north east monsoon season. Dry periods generally dominate in January to March and June to September. Rainfall stations with long term records are shown in Figure 1. Some short term rainfall stations which are usually located within 10 km of the respective stations shown in Figure 1 are also available. Figure 2 is the mean annual rainfall map of Selangor. It can be seen from the map that the mean annual rainfall generally varies from 2400 mm to 3000 mm. Although the basin basically receives higher rainfall than other parts of the world, drought has frequently been recorded as drought occurs when the rainfall amount for a certain period falls below the normal level recorded in the past, for the particular time scale.

2.2. Rainfall Data

Rainfall data and the periods of record available are shown in Table 1.

To obtain more accurate results, relatively continuous data with adequate length are required for SPI calculations. Considering the continuity, the availability of concurrent records and data available from nearby stations to infill the missing records for key stations chosen, it is considered appropriate to use data for the period 1948-1993. Detailed checks show that data of eight stations listed in Table 1 were suitable for use in this study.

2.3. Infilling of Missing Data

In order to preserve continuity of the monthly rainfall data for SPI calculations and data consistency checks, estimates of missing data were made for years up to 1993. As usual, it is noted that not all the stations in the study area are with complete data and stations where only a few months of data are missing were in filled using the records of nearby stations to increase the record length. No attempts were made to infill for stations with missing data more than a year. In this way it is possible to estimate the missing data for most of the stations as there are available nearby stations near the key rainfall stations listed in Table 1.

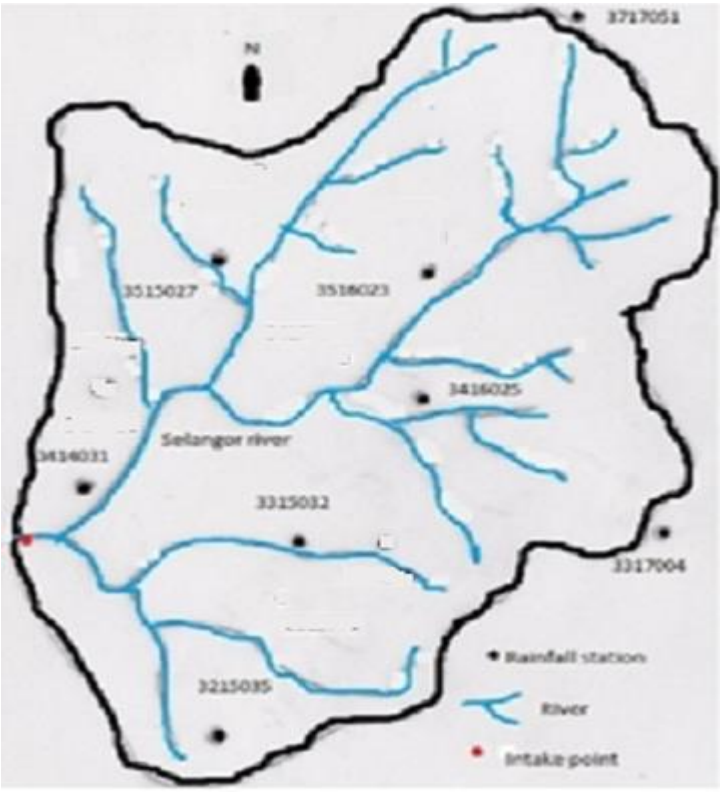

Figure 1. Rainfall Stations of Selangor Basin

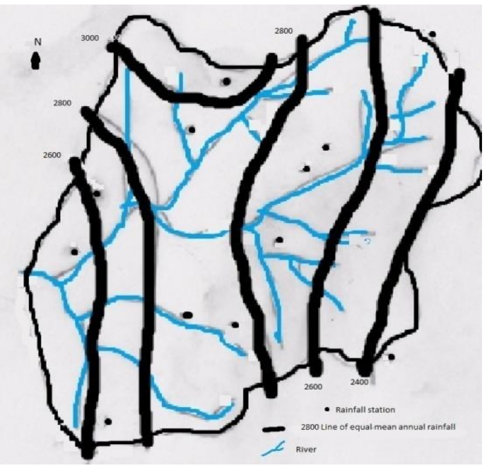

Figure 2. Mean Annual Rainfall of Selangor Basin

Table 1. Rainfall Records of Selangor Basin Used in this Study

Station no	Station name	Period of record
3215035	Ladang Strathairlie	1947-1997
3317004	Genting Sempah	1948-2000
3416025	Ldg Batang Kali	1947-1993
3414031	Selangor Tin Dredging	1947-1995
3516023	Kuala Kubu Hospital	1947-1997
3717051	Bt Fraser	1947-2000
3515027	Ldg Sg Beleta	1947-1994
3315032	Batang Kali	1947-1993

2.4. Trend of Rainfall Data

We use the trend /change detection software prepared for CRC for catchment Hydrology (Chiew and Siriwardena 2005) for testing the trend for the annual rainfall records. This software uses 11 different statistical tests to detect trend for time series samples. It includes parametric and non-parametric methods. Lag one autocorrelation function is also included to test randomness of a sample. For this study, we use the non-parametric Mann Kendall test, Sperman's Rho test, Turning point test, and auto-correlation test to test trend, randomness and independence of the data. Table 2 gives results at 5% significance level for the test. As most of the rainfall data of the stations passed the trend tests, the regional annual rainfall trend is not marked. Figures 3 and 4 show the trend test for two stations in the study area.

Table 2. Trend Test Results at 5% Significance Level

Station	Monn Kendall	Sperman's Rho	Turning point	Auto-correlation
3215035	NS	NS	NS	NS
3315032	NS	NS	NS	NS
3416025	S	NS	NS	NS
3515027	NS	NS	NS	NS
3515023	NS	NS	NS	S
3717051	S	S	NS	S
3317004	NS	NS	NS	NS
3414031	NS	NS	NS	NS

S denotes significant
NS not significant

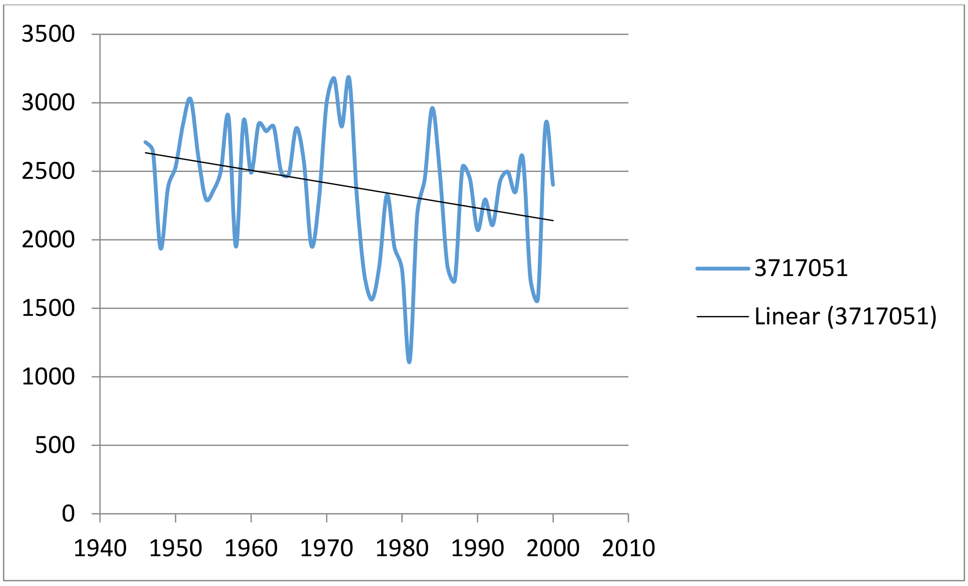

Figure 3. Annual Rainfall for Station 3717051 with Trend Line for 1946-2000

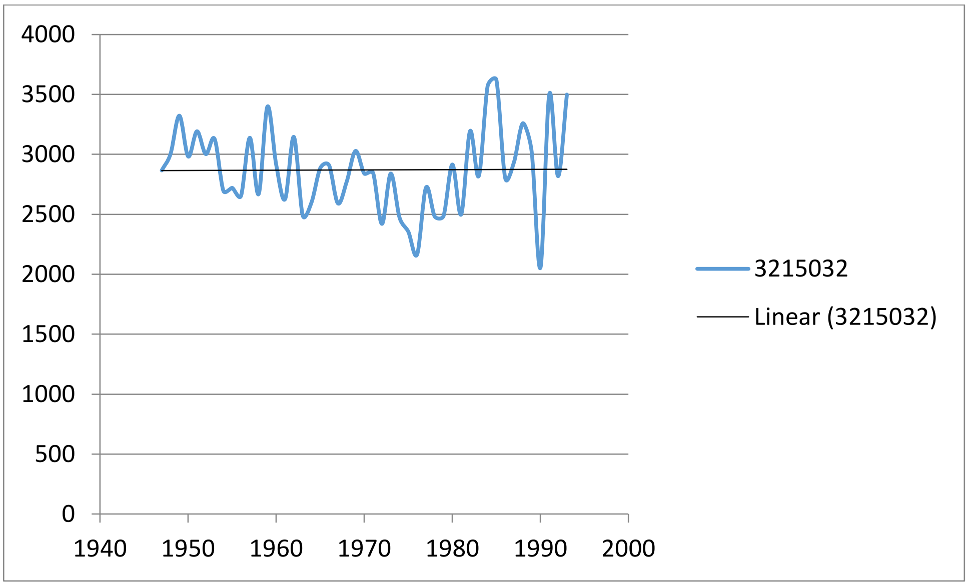

Figure 4. Annual Rainfall for Station 3215032 with Trend Line for 1946-1993

2.5. Outliers

Tests for outliers for the station records were carried out using annual rainfall data and the generalized Extreme Studentized Deviate (ESD) test. The ESD is a generalization of the Grubb's test (Zaiontz 2014). Results are presented in Table 3. The identified outliers are not excluded for further analysis unless there are strong hydrological and statistical evidences that they are real outliers.

2.6. Graphical Checks

Consistency of the annual rainfall data sometimes can be investigated using graphical plots. The annual rainfall of the adjacent stations are grouped and plotted as shown in Figures 5 to 7.Graphical plots show that stations 3414030,3315033 and 3414026 recorded low rainfall consistently for certain periods compared to adjacent stations. Station 3414029 has extremely low rainfall for 1991 to 1993.

Table 3. Results of Outlier Test

Station no	No of outliers
3215035	1
3315032	1
3317004	0
3414031	1
3416025	1
3516023	0
3516027	1
3717051	0

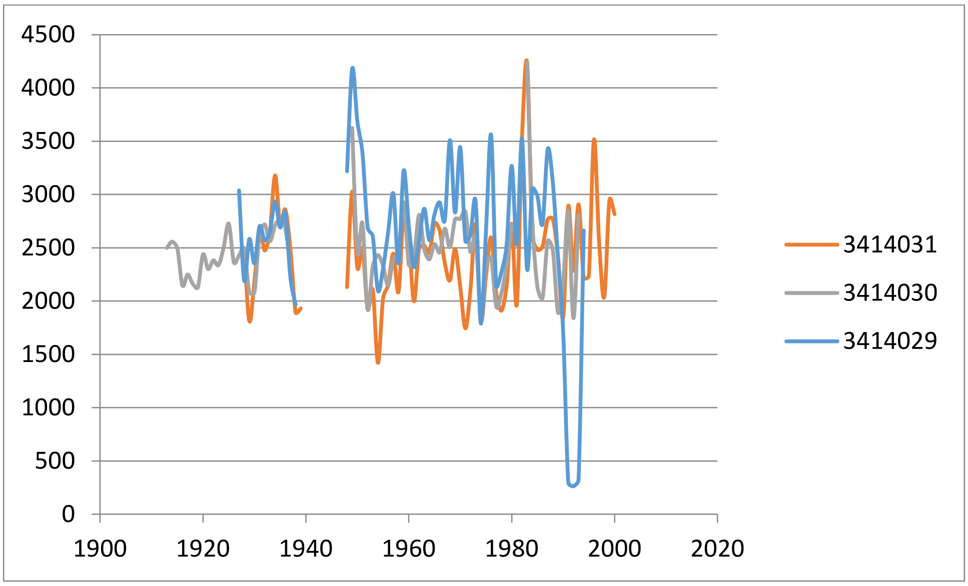

Figure 5. Graphical Plot of Annual Rainfall (a)

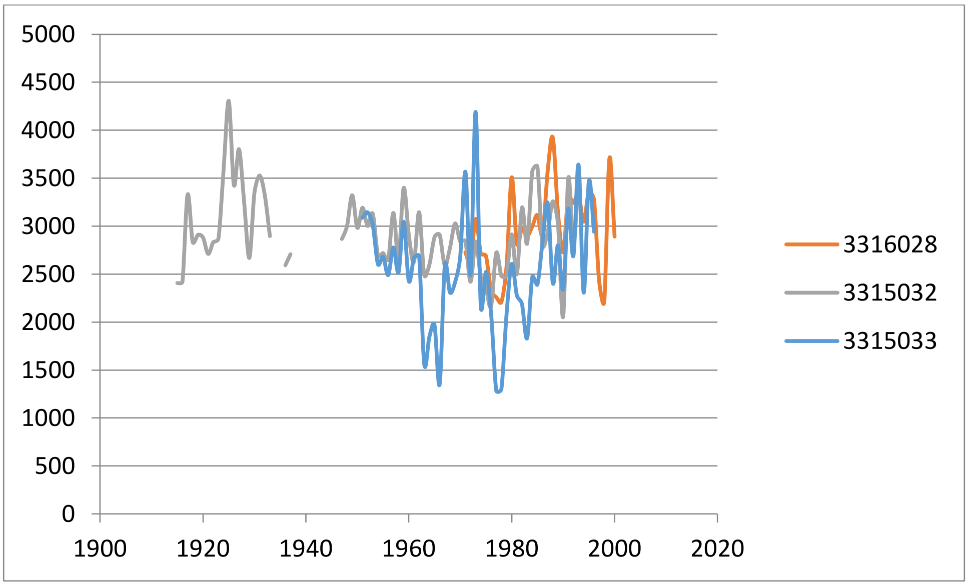

Figure 6. Graphical Plot of Annual Rainfall (b)

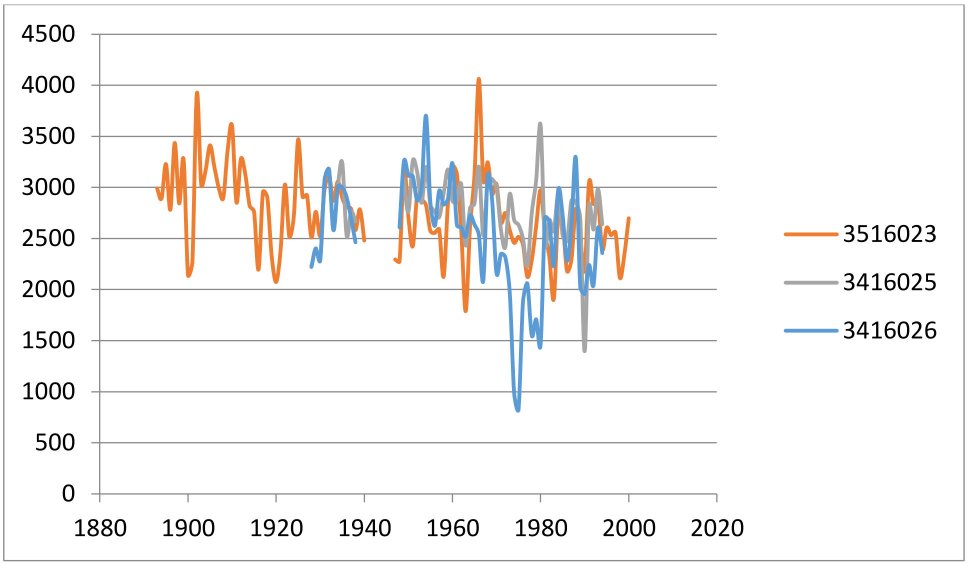

Figure 7. Graphical Plot of Annual Rainfall (c)

2.7. Methodology

The aim of the current study is to use sound scientific methods and reliable rainfall data to perform drought analysis, evaluation and forecasting so that the findings of the study can be used for formulating drought mitigation strategies by planners and water managers of the relevant authorities.

2.7.1. Calculation of Drought Indices: SPI is calculated based on monthly rainfall data. The SPI developed by McKee *et al.*, (1993) is based on rainfall alone. This makes its evaluation relatively easy compared to other drought indices. The other advantage of using SPI is that SPI is able to describe drought on multiple time scale.

The SPI drought index is computed by fitting a probability density function to the frequency distribution of rainfall summed over the time scale of interest. For example, the 3 month SPI at the end of March is calculated using the historical rainfall from January to March for all years. The Gamma probability distribution recommended by McKee et al. (1993) for use in SPI calculation has been widely used (Mishra *et al.*, (2006) , Belayneh *et al.*, (2013), Llyod Hughes and Saunders(2002), Morid *et al.*, (2007)). The calculation is performed separately for each month and for each desired location. The next step is to transform each probability function to the standardized normal distribution. The Gamma distribution as defined by its frequency or probability density function is:

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\alpha} \qquad \text{for x>0} \qquad (1)$$

Where $\alpha > 0$ *is* a shape factor, β >0 is a scale factor, and x>0 is the amount of precipitation. $\Gamma(\alpha)$ is the Gamma function which is defined as:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}\, e^{-y} dy \qquad (2)$$

To fit the distribution to the data requires α and β to be estimated. Edwards and Mckee(1997) propose estimating these parameters using the approximation of Thom(1958) for maximum likelihood as follows:

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right) \qquad (3)$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \qquad (4)$$

Where for n observations

A=$ln\bar{x}$ -$\frac{\sum \ln(x)}{n}$ (5)

The resulting parameters are then used to find the cumulative probability of an observed precipitation event for the given month and time scale as:

G(x)=$\int_0^x g(x)dx = \frac{1}{\hat{\beta}^{\hat{\alpha}}\Gamma(\hat{\alpha})} = \int_0^x x^{\hat{\alpha}-1}\, e^{-x/\hat{\beta}}dx$ (6)

If t is used for $x/\hat{\beta}$ the equation will reduce to incomplete Gamma function. McKee *et al.*, (1993) used an analytical method along with suggested software code from Press *et al.*, (1986) for solving the equation. For the undefined Gamma function at x=0 and a rainfall distribution may contain zeros, the cumulative probability becomes:

H(x)=q+(1-q)G(x) (7)

Where q is the probability of zero rainfall.

The cumulative probability H(x), is then transformed to the normal random variate , Z with mean zero and variance one, which is the value of SPI. Following steps suggested by Edwards and McKee (1997) the alternative approximate conversion is:

Z=SPI=-(t-$\frac{c_0+c_1t+c_2t^2}{1+d_1t+d_2t^2+d_3t^3}$) for 0<H(x)<0.5 (8)

Z=SPI=+(t-$\frac{c_0+c_1t+c_2t^2}{1+d_1t+d_2t^2+d_3t^3}$) for 0.5<H(x)<1.0 (9)

Where

$t = \sqrt{\ln[\frac{1}{(H(x))^2}]}$ for 0<H(x)<0.5 (10)

t= $\sqrt{\ln[\frac{1}{(1-H(x))^2}]}$ for 0.5<x<1.0 (11)

and

$c_0 = 2.515517$ $c_1 = 0.802853$ $c_2 = 0.010308$
$d_1 = 1.432788$ $d_2 = 0.189269$ $d_3 = 0.001308$

McKee *et al.*, (1993) suggest the classification system shown in Table 4 to define drought intensities resulting from the SPI. The criteria for a drought event for any of the time scale are also defined. Drought occurs when the SPI is continuously negative and reaches an intensity of -1.0 or less. Drought ends when the SPI becomes positive. Each drought event, therefore, has a duration defined by its beginning and end, and an intensity for each month that the event continues.

Table 4. Drought Classification Based on SPI

SPI values	Class
>2	Extremely wet
1.5-1.99	Very wet
1.0-1.49	Moderately wet
-0.99-0.99	Near normal
-1—1.49	Moderately dry
-1.5—1.99	Severely dry
< -2	Extremely dry

The SPI-SL-6 program of World Meteorological Organization (2012) was used to compute the drought indices (SPI) using the average monthly basin rainfall starting from each month of the year for different time scales.

2.7.2. The MLP Neural Network Model: Artificial neural networks (ANN) are nonlinear and flexible massively parallel distributed information processing system that has certain characteristics similar to the biological neural networks of the human brain. For a number of nonlinear processing units, it is possible to train the neural networks to learn from experience and compute the complex functional relationships with accuracy. A number of neural networks has been proposed in the literature but the most commonly method used in hydrology for flood and drought forecasting is the feed forward multi-layer perceptron (MLP) model. As an example, a typical three layer feed forward MLP model is shown in Figure 8. For most drought forecasting studies carried out in the past two decades, the three layer MLP model was used with the input nodes consist the lagged SPI values and the output is the forecasted future value. In this particular MLP model, hidden nodes are used to process the information transmitted from the input nodes with a particular nonlinear transfer function. Previous studies (Mishra *et al.*, (2006), Belanah *et al.*, (2013) indicate that forecasting results for SPI using longer lead time ahead predictions are less accurate, therefore ,a drought prediction time horizon of one month is used in this study. In this context ,predictions are made one time step ahead using current and previous SPIs for all the subsequent simulation runs. This is also based on the requirements that forecasting will be reviewed on a monthly basis as more data will be available and the model can be reassessed for the coming month. For these reasons, the model considered is a single output MLP model. The network is processed through training, testing and validating stages in order to forecast the SPI indices using the input SPI data. Back propagation (BP) algorithm (Rumehaet 1986) is usually used to correct the weights of the interconnecting neuron. Back propagation (BP) uses the steepest gradient descent method to correct the weight of the interconnecting neuron. This method (BP) solves the interconnection of the processing of processing elements by adding hidden layers. For the learning process in the back propagation method, the interconnection weights are adjusted using the error convergence method to obtain a desired output from an input. The BP algorithm propagates the error at the output to the input layer through the hidden nodes to obtain the final output. The gradient technique is used to calculate the weight of the network and adjust the weight of the interconnections to minimize the output error.

BP uses the following equation to correct the weighting factor:

$$\Delta w_{ij}(n) = \alpha \Delta w_{ij}(n-1) - \varepsilon(\frac{\partial E}{\partial w_{ij}}) \tag{12}$$

Where $\Delta w_{ij}(n) and\ \Delta w_{ij}(n-1)$
$are\ weights\ interconnecting\ nodes\ i\ and\ j\ during\ the\ nth\ and\ (n-1)th\ steps$

α Is the momentum factor used to speed up training in flat regions of the error surface and helps to prevent oscillations in the weights.

A learning rate ε is used to increase the chance of avoiding the training process being trapped in a local minima instead of a global minima.

The number of neurons in the input and output layers are problem dependent and decided by the number of input and output variables in the MLP model. The size of hidden neurons is an important factor in solving the problems using MLP. There are no fixed rule in determining the number of hidden neurons required for the model and trial and error experiments are normally adopted to determine the hidden node that gives the model the best performance. However, empirical relationships between optimum hidden neurons and number of input and output elements were given by some authors *e.g.*, Mishra *et. al.*, (2006) used 2n+1 for estimating the number of hidden neurons.

Where n is the number of input neurons

As the specific number of hidden node and input node that will produce the best SPI cannot be fixed, we have to recourse to a trial and error procedure to determine the best combination of input and hidden nodes which will produce the best result in drought forecasting, and this method has been used in previous drought forecasting studies carried out by various authors. The number of input nodes are progressively increased from 1 to 20 and the hidden nodes increased from 1 to 2n+1, where n is the number of input nodes. The combination pair of input node and hidden node which produces the minimum relative root mean square error and highest correlation coefficient is accepted as the optimum network for the model.

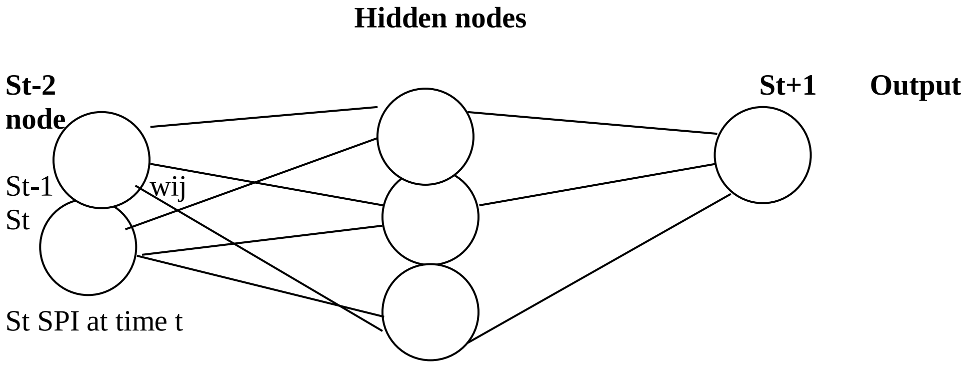

Figure 8. A Three Layer MLP Neural Network Method

The neural network add-in version 1.5 software developed by the University of Adelaide (2014) was used for the forecasting SPI index for a one month lead time.

The activation function used is the logical sigmoid function.

For training and validating purposes, data are normalized using the scaling method.

Input variables are selected using the partial mutual information selection option and data are split randomly with 60%, 20%, and 20% for training, testing and validating purposes using the program.

The data is trained using learning rate of 0.01 and momentum coefficient of 0.9.

The performances of the MLP network model in predicting the drought index are assessed using:

$$\text{RMSE}=\sqrt{1/p\sum[(X_m)_i - X_{s_i}]^2} \tag{13}$$

$$\text{MAE}=1/p\sum_1^p\left|(X_{m_i}) - (X_{s_i})\right| \tag{14}$$

Where m and s represent the observed and predicted SPI p= total number of observations.

MAE is the mean absolute error
RMSE is root mean square error

3. Results and Discussion

In this study ,SPI were derived for 3 ,6 and 9 month durations using average basin rainfall of Selangor basin and these are then used together with the MLP model to forecast the short and medium term droughts. The neural network excel add in version 1.5 software is used for data processing and subsequent mathematical calculations. In the

Selangor basin, severe droughts are usually persistent only for a few months and the design critical period for the existing reservoir is 9 months. Moreover, in the agricultural aspects, short term droughts for three to four months are critical and thus only short and medium term droughts are considered in this study. Drought studies are therefore focused on the 3 to 9 month time scales. Results of 3 to 9 month SPI (SPI-3, SPI -6, and SPI-9) are shown in Figures 9 to 11.

Representative SPI series for 3 to 9 months starting from January (*e.g.*, 3 month SPI starting from January to March) were used for forecasting the one month ahead SPI .The performance of the MLP model in predicting the SPI for different time scales are presented in Table 5.The results are presented in terms of correlation coefficient, MAE and RMSE for the validation dataset.

The predicted and observed SPI for 9 month duration are plotted in Figure 12.

It is noted that SPI with longer duration time scale can be predicted more accurately than shorter series data. In other words, the 9 month SPI values are predicted more accurately than those of 3 and 6 months using the MLP neural network model. These findings are consistent with conclusions made by other authors employing MLP model and SPI for drought forecasting.

Table 5. Comparison of Forecasting Measures Between Observed and Predicted Data

SPI series	r	MAE	RMSE
SPI-3	0.856	0.46	0.56
SPI-6	0.92	0.31	0.39
SPI -9	0.94	0.28	0.34

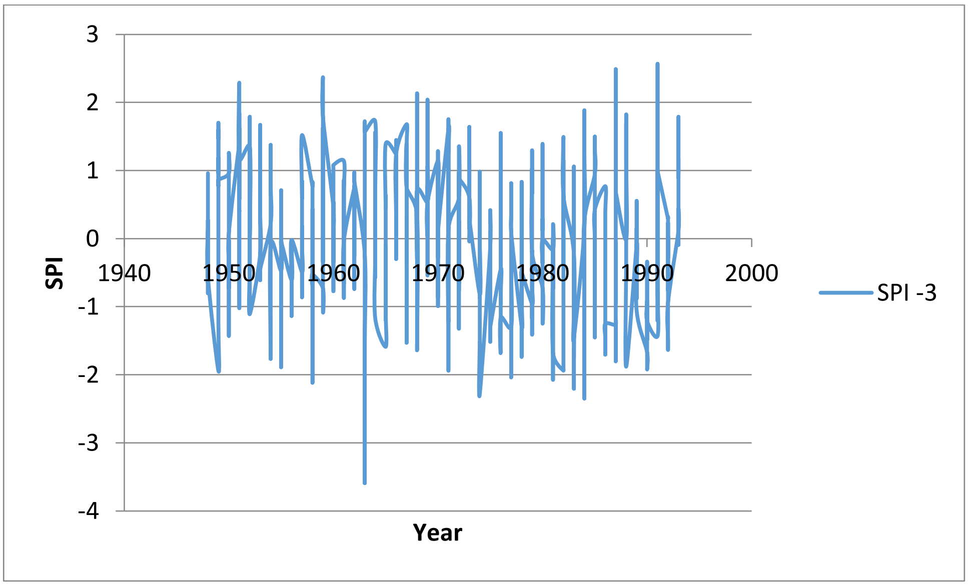

Figure 9. SPI Series of 3 Month Time Scale Based on Average Basin Rainfall of Selangor

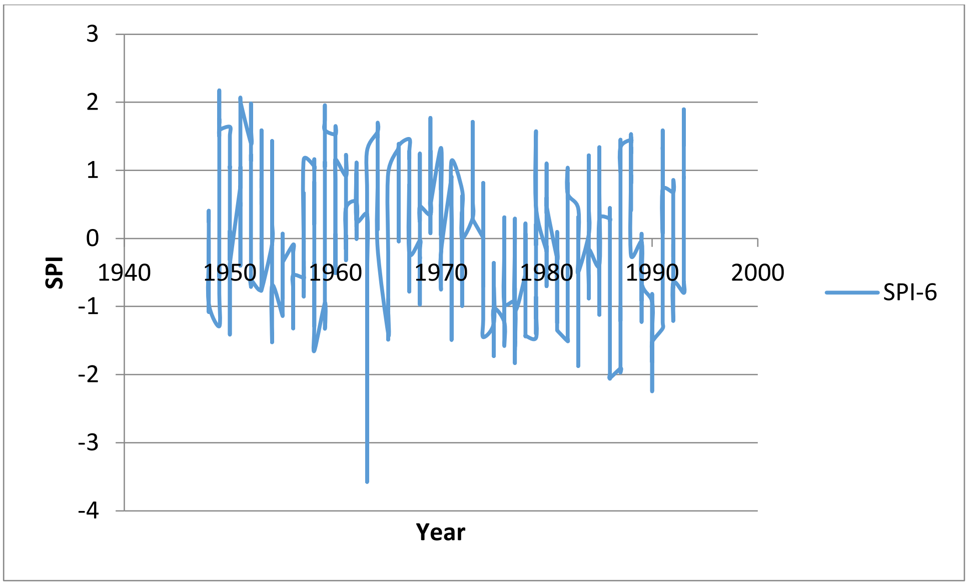

Figure 10. SPI Series of 6 Month Time Scale Based on Average Basin Rainfall of Selangor

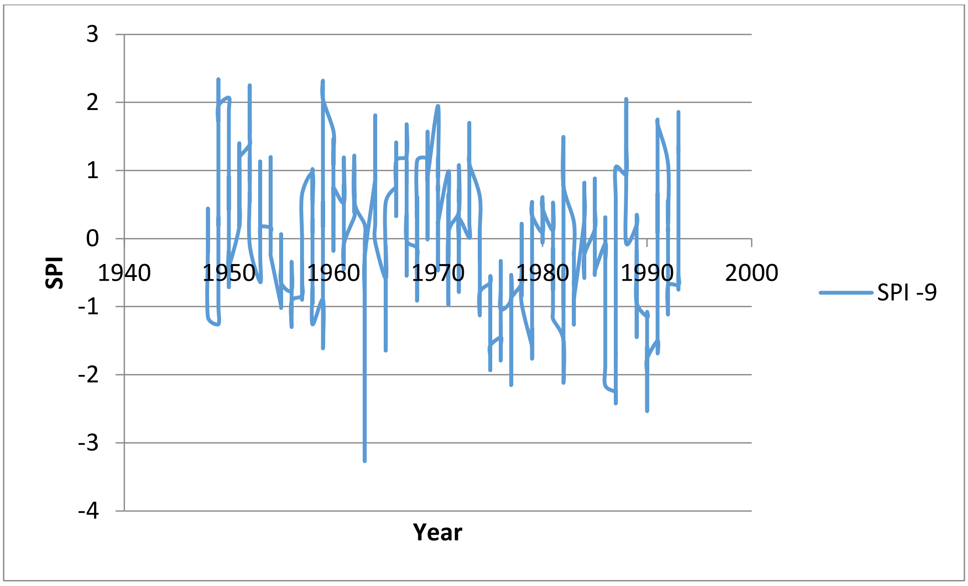

Figure 11. SPI Series of 9 Month Time Scale Based on Average Basin Rainfall of Selangor

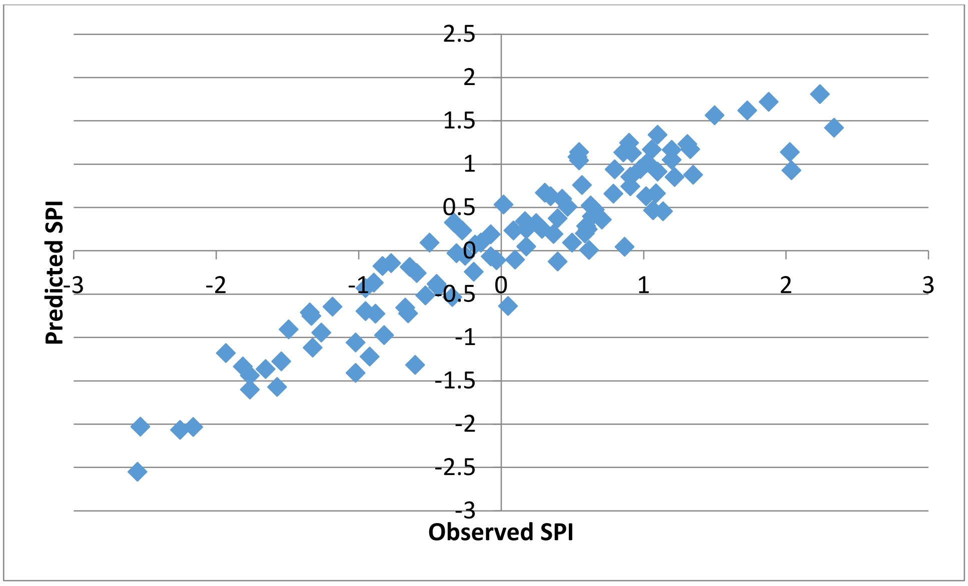

Figure 12. Observed and Predicted SPI-9

4. Conclusion

The purpose of this study is to use established scientific methods and available hydrological data to identify, monitor, and forecast droughts for the planning, management and formulating drought strategies to reduce and mitigate the adverse effect of drought impacts in Selangor basin. Standardized precipitation index (SPI) has been used as a conventional tool to identify and monitor drought occurrences. To achieve the aims, we use average long term monthly rainfall data for eight stations covering both the dry and wet seasons from Selangor river basin to derive the SPI values for durations of 3 to 9 months. These drought indicators, which are time series derived from rainfall data together with the multi-layer artificial neural networks model were used for drought forecasting for the basin. Forecasting were made for SPI with a one month ahead lead time as forecasting accuracy is reduced for longer lead times. This has been shown by studies carried out elsewhere. Our finding indicates that more accurate predictions are achieved using SPI of longer durations, *i.e.* 6 and 9 months. This is consistent with findings of studies by others.

References

[1] A. Belayneh and J. Adamowski, "Forecasting using new machine learning method", Journal of water and land development. Drought, no. 18 (1-v1), **(2013)**, pp. 3-12.

[2] F. Chiew and L. Siriwardena, "Trend user guide", www.toolkit.net.au/trend, **(2005)**.

[3] D. C. Edwards and T. B. McKee, "Characteristics of 20th century drought in the United States at multiple time scale", Colorado State University, Fort Collins, Climatology report, CO., USA, No 97-2, **(1997)**.

[4] B. Llyod–Hughes and M. A. Saunders, "A drought climatology for Europe", International Journal of Climtol, no. 225, **(2002)**, pp. 1571-1592.

[5] T.B. McKee, N. J. Doesken and J. Kliest, "The relationship of drought frequency and duration to time scales", Proceedings in the English Conference on Applied Climatology, Anaheim, CA American Meteorological Society, **(1993)** January 17-32.

[6] A. K. Mishra and V. R. Desain, "Drought forecasting using feed forward recursive neural network ,Ecological modeling", vol. 198, no. 2, **(2006)**, pp. 117-138.

[7] S. Morid, V. Smakhin and K. Bagherzadeh, "Drought forecasting using artificial neural networks and time series of drought indices", International Journal of climatology, vol. 27, no. 15, **(2007)**, pp. 2103-2111.

[8] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vettrling, "Numerical recopies", Cambrige University Press, UK, **(1986)**.

[9] D. E. Rumelhart, G. E. Hilton and R. J. William, “Learning representations by back propagation errors Nature”, **(1986)**, pp. 313, 533-536.

[10] School of Engineering, Adelaide University, Neural networks add in, version 1.5, **(2014)**.

[11] H. C. S. Thom, “A note on Gamma distribution”, Monthly weather rev, vol. 86 ,**(1958)**, pp. 117-122.

[12] World Meteorological Organization, Standard Precipitation Index, User guide , **(2012)**.

[13] C. Zaiontz, “Real statistics in Excel”, http//www.real-statistics.com, **(2014)**.

[14] D. Hong and K.A. Hong, “Drought Forecasting Using MLP Neural Networks”, DCA **(2015)**.