

# A Hybrid Wavelet-ARIMA Model for Standardized Precipitation Index Drought Forecasting

Alfa Mohammed Salisu and Ani Bin Shabri\*

Department of Mathematical Sciences, Faculty of Science  
Universiti Teknologi Malaysia (UTM)  
81200 Johor Bahru, Malaysia

\*Corresponding author: ani@utm.my

## Article history

Received: 7 January 2019

Received in revised form: 20 July 2020

Accepted: 20 July 2020

Published online: 1 August 2020

---

**Abstract** This paper proposes A Hybrid Wavelet-Auto-Regressive Integrated Moving Average (W-ARIMA) model to explore the ability of the hybrid model over an ARIMA model. It combines two methods, a Discrete Wavelet Transform (DWT) and ARIMA model using the Standardized Precipitation Index (SPI) drought data for forecasting drought modeling development. SPI data from January 1954 to December 2008 used was divided into two - (80%/20% for training/testing respectively). The results were compared with the conventional ARIMA model with Mean Square Error (MSE) and Mean Average Error (MAE) as an error measure. The results of the proposed method achieved the best forecasting performance.

**Keywords** ARIMA; Wavelet; SPI; Drought; Forecasting

**Mathematics Subject Classification** 91B76.

## 1 Introduction

To facilitate the comparison between ARIMA and Wavelet-ARIMA model, four SPI drought data series are used for forecasting. Drought forecasting over the world is an issue which has been carried out in different studies undertaken by different researchers by the application of different models to achieve their perceived goals. Forecasting drought using SPI data is very important in the management of water resources and planning for all creatures. Referring to the work of [1], time series forecasting has commonly been used in a broad range of scientific applications that includes meteorology and hydrology. Drought is described by [2] as one of natural phenomenon involving climate which is the first natural disaster in the world over which affects places and inflicting significant damages to both human beings and the environment in which they live. Drought as a normal feature of climate that occurs due to rainfall that is below average in a place that leads to the shortage of water, loss in economic activity and unexpected

reduction in precipitation over time, which is one of the most harmful natural disasters that affect human beings [3].

In the current and various studies, therefore, time series forecasting has commonly been used. Time series forecasting as one of the important research areas in analyzing the hydrological time series [4]. On the other hand, [5] found that time series forecasting is an act of predicting the future by understanding the past. Time series forecasting has become an important approach to drought forecasting whose application is widely used [6]. In the view of [7], forecasting techniques which have been developed with the performance evaluation in other to forecast future values of a time series are one of the areas in time series analysis. The principal aim of time series forecasting is to forecast future events based on known past data or an event.

SPI data series has a wide application for the description of drought in a different period. It is a non-fact that little efforts have been utilized to analyze the role played by the SPI used for the drought forecasting [8]. SPI is an expression that the actual rainfall can be view as a standardized departure from rainfall probability distribution function and has gained significance in recent time as a potential drought indicator which permits comparisons that cut across space of time [9].

SPI data series is therefore expected to serve as a data input in the ARIMA model to be assessed by using only precipitation SPI data which were applied for drought analysis. ARIMA model is a Box Jenkins methodology named after the original authors, [10] which seeks to transform any time series data to be stationary after which the data is then applied for forecasting using the past univariate time series process for the future forecast with some selection and diagnostic tools. ARIMA model is very popular due to its flexibility in representing several varieties of time series with the associated Box-Jenkins methodology [11,12,13].

Also, in a separate analysis, the ARIMA model was combined with a wavelet and thus, the discrete wavelet transforms (DWT) was used because of its simplicity and shorter time for computation purposes, which serves as an alternative in forecasting applications. The wavelet-ARIMA model can achieve a better forecasting accuracy than the traditional ARIMA model [14-17]. The existing forecasting ARIMA model was improved upon in the forecast of climate time series by using the wavelet transform [18].

The wavelet transformation has the basic objective of being able to do the analysis of the time series data involving time and frequency domain by decomposing the original time series in different frequency bands, because wavelets are tools which are important in time series forecasting with the use of wavelet functions. The wavelet transforms are mathematical functions which can be used for the analysis of time series with non-stationarities. This wavelet technique allows the use of long-time intervals for low-frequency information and short time intervals for high-frequency information that can reveal aspects of data like trends. Another merit of wavelet analysis is the flexibility of the choice of the mother wavelet in accordance with the characteristics of the time series that is investigated. Wavelet transform also has the advantage of allowing the study of various independent treatments on distinct time scales.

## 2 Methodology

Both ARIMA model and Wavelet have been found to be very effective in the areas of drought forecasting generally and has been so applied in many fields such as drought, flooding, river flow, and streamflow by applying different data sets like the rainfall which has its own limitations of being able to capture SPI data series. This study is motivated by the application of SPI data to obtain the best ARIMA model and to combine it with the wavelet to get a hybrid model for each of the SPI data series.

The main objective of this study is to combine wavelet with the ARIMA model for the drought forecasting using the SPI data series and to compare the traditional ARIMA model with the hybrid model (W-ARIMA model).

ARIMA which is a benchmark model based on the PACF and ACF of the time series was combined with the wavelet to obtain the Wavelet-ARIMA model which improved the ARIMA model which gave a better performance. To obtain this the SPI drought data was decomposed into  $D_{L-1}$  ( $L$  is the number of levels) of components which were used as inputs to the ARIMA model which made the errors obtained to be lower than that of the ARIMA model.

### 2.1 Standardized Precipitation Index (SPI)

As stated earlier, the data used here is the SPI data comprising SPI3, SPI6, SPI9, and SPI12 data series. SPI was developed in 1993 by [18] and his group and has been used in over 60 nations and applied as a drought indicator [19]. SPI was primarily developed and used as a tool for defining and monitoring drought. It permits the analyst to gauge the intensity of drought at a period. Being a tool for drought prediction, SPI is therefore, a tool for drought monitoring by the world meteorological organization [20, 21], for operational instruments [22]; [23] used for drought climatology in Europe [24]. [25] used SPI for the study of temporal and spatial variability of drought in Portugal. Equally, [26] determines the agreement among other indicators using regional SPI for analysis of coherent drought patterns in Europe.

SPI is simple and uses only rainfall as input parameter: it is normalized and can be used to compare the drought severity for areas that have different climates, it uses different time scales and can be used in assessing different types of drought, it has a historical context, it is only based on precipitation, it provides early warning of the drought and it is less complex.

However, its disadvantages are: it has no temperature and it is based on preliminary data which can change. SPI is very important for the more effective assessment of drought, and the World Meteorological Organization [27] recommended the adoption of SPI to monitor the severity of drought events. In quantifying the deficit of precipitation, SPI has played a great role [28] view computation of SPI at different time scales from 1 to 48 months or more which can depend on the application in which short-term SPI can be used to detect agricultural drought and long-term SPI can be used for water supply management.

SPI is an index obtained from a long record of precipitation in each location which takes at least 30 years period. According to [29], SPI has been used to identify the meteorological drought or precipitation deficit. The SPI value has a range from  $-2$  to  $+2$  [30]. This is further explained in Table 1.

Table 1: The Different SPI Indexes

Stage	SPI Index
Normal	(0.99) to (−0.99)
Drought Watch	(−1.0) to (−1.49)
Drought Warning	(−1.5) to (−1.99)
Drought Emergency	(−2.0) to (<−2,0)

(Source: Mc Kee *et al.* [17])

## 2.2 Auto-Regressive Integrated Moving Average (ARIMA) Model

One of the most widely used time series models is the ARIMA model where its popularity is due to its statistical properties as well as the Box-Jenkins methodology in the building process [31]. With the ARIMA model, complex patterns in the data can also be used to generate forecasts [32]. In responding to the time series as a linear combination of its past values, the ARIMA model could predict a value or values [33].

Based on the approach of Box and Jenkins, ARIMA models for the SPI time series were developed in three steps: model identification, parameter estimation, and diagnostic checking. The details on the development of ARIMA models for the SPI time series can be found in the works of [34]. ARIMA time series, which is made stationary due to the differencing process, is referred to as the ARIMA model. ARIMA model is made up of three parameters namely  $p$  (order of autoregressive model),  $d$  (order of differencing) and  $q$  (order of moving average model). ARIMA models are one of the most important forecasting models that have been successfully applied in modeling and forecasting.

Non-seasonal ARIMA models are shown in equation (1).

$$X_t^1 = C + \phi_1 X_{t-1}^1 + \dots + \phi_p X_{t-p}^1 + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + \varepsilon_t \quad (1)$$

where  $X_t^1$  “differenced series” (It may have been differenced more than one time).

The “predictors” on the RHS include both lagged values of  $X_t$  and lagged errors. This is referred to as the ARIMA ( $p, d, q$ ) model in which

$p$  = order of the autoregressive part

$d$  = degree of first differencing involved

$q$  = order of moving average part

$\phi, \theta$  are polynomials of order such as  $\phi_p, \theta_q$ .

The ARIMA model in equation (1) can be written as

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)(1 - B)^d X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t \quad (2)$$

First-term referred to AR ( $p$ ), second term referred to  $d$  = differencing, RHS part referred to MA( $q$ ) shown in equation (2). Selecting the appropriate values for  $p$ ,  $d$ , and  $q$  can be difficult, hence, relevant software can be used.

### 2.3 Seasonal ARIMA Models

ARIMA models can model a wide range of seasonal data and it is formed by including additional seasonal terms in the ARIMA models like,  $ARIMA(p, d, q)(P, D, Q)_s$  which implies (Non-seasonal part of the model) (Seasonal part of the model) and  $s$  is the number of periods per season. The seasonal part of the model is made up of terms that are very similar to the non-seasonal components of the model, however, they involve backshifts of the seasonal period. For instance, an  $ARIMA(1, 1, 1)(1, 1, 1)_4$  model without a constant term is for quarterly data shown in equation (3).

$$(1 - \varphi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)X_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)e_t. \quad (3)$$

From the LHS, the first term is Non-seasonal AR(1); the second term represents Seasonal AR(1); the third term represents Non-seasonal difference; the fourth term represents Seasonal difference. Hence, from the RHS, the first term represents Non -seasonal MA(1); the second term represents Seasonal MA(1). It should be noted that the additional seasonal terms are simply multiplied by the non-seasonal terms.

Differencing is applied to time series data to make it stationary (which is a time series property whereby it does not depend on the time at which the data is observed). In differencing the data, the differences between consecutive observations are computed in equation (4).

$$Y_t = (1 - B)^d X_t. \quad (4)$$

Differencing the data eliminates the changes in the level of a time series data which eliminates the trends, seasonality and consequently stabilizes the mean of the series. Sometimes, it may be necessary to have a first or second differencing of the data to obtain a stationary time series. The first order ( $d = 1$ ) and the second order ( $d = 2$ ) differencing of time series are given in equation (5) and equation (6), respectively.

For  $d = 1$

$$Y_t = (1 - B)X_t = X_t - X_{t-1}. \quad (5)$$

For  $d = 2$

$$\begin{aligned} Y_t &= (1 - B)^2 X_t = (1 - 2B - B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2} \end{aligned} \quad (6)$$

Another method of differencing the data is by utilizing seasonal differencing involving the computation of the difference between an observation and the corresponding observation of the previous year given in equation (7).

$$Y_t = (1 - B^s)^D X_t, \quad (7)$$

where  $s$  is the number of seasons and  $D$  is seasonal difference. For  $D = 1$  and  $s = 4$

$$Y_t = (1 - B^4)X_t = X_t - X_{t-4}.$$

The differenced data is then used for the estimation of an ARIMA model.

## 2.4 Wavelet Analysis

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 (i.e. frequency) domains [35]. 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 [36].

Wang and Ding [37], 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, [38], described wavelet analysis as a multi-decomposition analysis that provides information for time and frequency domains and gives useful decompositions of the original time series for the wavelet-transformed data in order to improve the power of a forecasting model. A wavelet is a tool in time series forecasting whose importance has been applied by many researchers. One of the basic objectives of wavelet transforms is to analyze the time series data. Wavelet transform can be categorized into a continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The former is not always used in forecasting because of its complex computational ability and the time needed for it [39]. In place of it, DWT is widely used in applications of forecasting to simplify numeric solutions. Therefore, it requires less time for computation which is simple to apply. DWT is given by equation (8)

$$\psi_{m,n}(t) = \frac{1}{\sqrt{s_0^m}} \psi\left(\frac{t - n\tau_0 s_0^m}{s_0^m}\right) \quad (8)$$

$\psi(t)$  is the mother wavelet,  $m$  and  $n$  are integers that control the scale and time respectively. The most common selections for the parameters are  $s_0 = 2$  and  $\tau_0 = 1$ . Mallet's theory has the original discrete time series  $x(t)$  that can be decomposed into  $a^{s_0}$  series of linearity independent approximation and detail signals by using the inverse DWT which is given by [40] as equation (9),

$$x(t) = T + \sum_{m=1}^M \sum_{t=0}^{2^{M-m-1}} W_{m,n} 2^{-(m/2)} \psi(2^{-m}t - n), \quad (9)$$

where  $W_{m,n} = 2^{-(m/2)} \sum_{m=1}^{N-1} \psi(2^{-m}t - n)x(t)$  is the wavelet coefficient for the discrete wavelet at scale  $s = 2^m$  and  $\tau = 2^m n$ .

In the analysis, DWT is mostly preferred in the forecasting problems due to its simplicity and ability to compute within a short period of time. Many researchers have undertaken studies in the field of water resources and hydrology using wavelet transforms which are based on data pre-processing. As a pre-processing tool, wavelet transforms provide a useful decomposition of the original time series, so that the pre-processed data can improve the ability of a forecasting model by capturing the information based on different resolution levels [41]. Wavelet transform (WT) analysis has become an ideal tool for the study of a measured non-stationary times series through the hydrological process. The DWT requires less time for computation and simple to implement. DWT scales and positions are usually based on powers of two (dyadic scales and positions). This is achieved by modifying the wavelet representation. [42] stated that the DWT operates two sets of functions: high-pass and low-pass filters. DWT is the best-known tool for

data analysis whose contribution to model hydrological resources can be seen in the last few years [43].

The main aim of this study is to compare traditional drought forecasting (ARIMA model) with a data pre-processed using wavelet transforms and the proposed drought forecasting model which is based on the wavelet transforms and ARIMA, known as Wavelet-ARIMA (W-ARIMA) model with SPI data series. The W-ARIMA model is obtained by the combination of two methods of DWT and ARIMA models. In W-ARIMA, the original drought SPI data series were decomposed into several sub time series components which serve as input to ARIMA in other to improve the accuracy of the model. To obtain several decomposition levels, the following formula by [44] is applied in equation (10).

$$L = \text{int}[\log(N)] \quad (10)$$

where  $L$  is the decomposition level and  $N$  is the number of the SPI data series. In this formula, the original SPI drought data series is decomposed into ten level components ( $A, D_1, D_2, \dots, D_{L-1}$ ) which stands for different frequency components of the original data. Each of these components plays a unique role in the original SPI data and has different effects on the original SPI drought. Instead of using D component separately, as an input model, the researchers employ added suitable D component which is more useful and capable of increasing the forecast performances of the hybrid models.

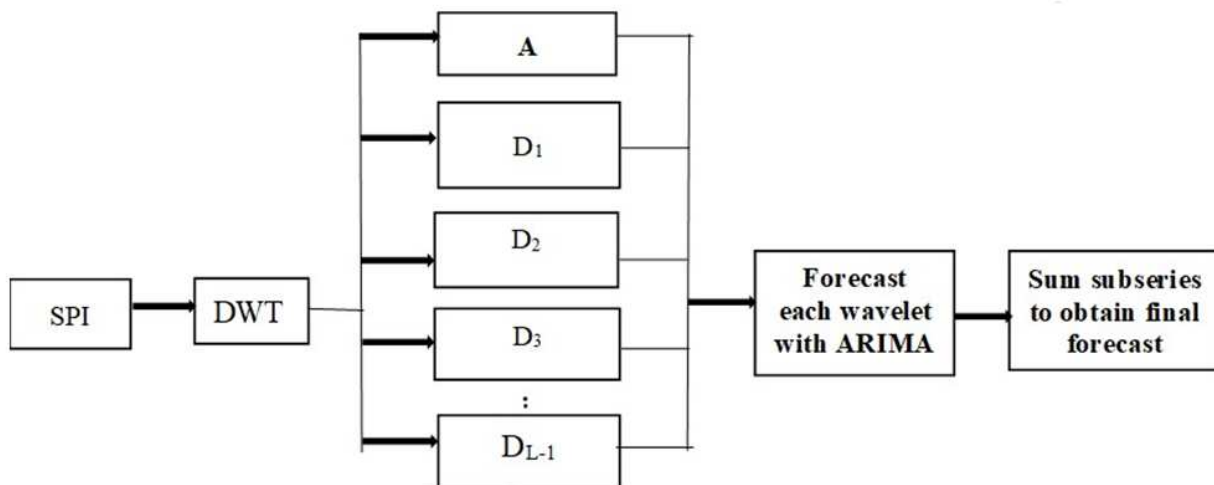


Figure 1: Framework of W-ARIMA Model

The structure in Figure 1 explains the development of W-ARIMA model in which each wavelet is used to forecast the ARIMA model. This begins with the input of the SPI data series to DWT by summing all the values for  $A, D_1, D_2, \dots, D_{L-1}$  in each of the data series to obtain SPIs for Wavelet-ARIMA model, which is then being input to the original data to obtain the MSE and MAE for both training and testing phases of the analysis.

## 2.5 Forecast Evaluation Methods

The criteria in judging the best model are how relatively small the models are in both the training and testing of the data series. This is needed in order to be able to quantify the

amount by which the estimator differs from the true (original) value. That is the reason why the measures with the smallest values are usually selected as the best model.

Evaluation of the performance of each model was based on the Mean Square Error (MSE) and Mean Absolute Error (MAE) for both training and testing which is used for this study. All these performance evaluations are widely used in obtaining the results of time series forecasting [44]. These forecast evaluation methods are shown in equations (11) and (12) respectively.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^m (X_t - \hat{X}_t)^2 \quad (11)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^m |X_t - \hat{X}_t| \quad (12)$$

where  $\hat{X}_t$  is the predicted value,  $X_t$  is the actual value at time  $t$  and  $n$  is the number of predictions.

### 3 Results and Discussions

#### 3.1 Used Data

This study uses rainfall precipitation data from 15 stations in Peninsular Malaysia obtained from the Malaysia Department of Irrigation and Drainage and this corresponds to standardized precipitation index (SPIs) i.e. SPI3, SPI6, SPI9, and SPI12. The time series of each of the SPI values were computed. The reason for considering the total precipitation for running periods of 3, 6, 9 and 12 months is due to the classification of drought to be a short-term for SPI3, medium-term for SPI6 and SPI9 and long-term for SPI12. To develop the model, the data set which is made up of 624 months was obtained from January 1954 to December 2008. The SPI data was divided into two parts, which made up of training 80% (1954-1998) and testing 20% (1998-2008) for analysis.

#### 3.2 The ARIMA Model

The selection of parameters for the ARIMA models is based on the PACF and ACF of the time series. as soon as significant lags were obtained from PACF and ACF, ARIMA models with different combinations were then developed and the model with the lowest MSE and MAE as contained in Tables 1 and 2 were selected. Figure 2 and Figure 3 show the PACF and ACF plots.

Table 2 is made up of all the detail ARIMA model results using forecasting accuracy of MSE and MAE as the performance evaluation criterion which explains the detail results involving all the SPI data series. The results illustrated that SPI3 is a non-seasonal data while the rest three are seasonal data. Table 2 also illustrates the detailed result of each SPI data series.

Table 3 indicates the summary of selected best ARIMA model results extracted from Table 2, which are compared with the results of the W-ARIMA model in Table 4.

Figure 2 indicates all the 624 data sets for each of the SPI data used for the study.



Table 2: ARIMA Forecasting Accuracy Results for All the SPI Data

		Training		Testing	
Data	Model	MSE	MAE	MSE	MAE
SPI3	ARIMA (4, 0, 2)	0.432542193	0.5162857	0.316868681	0.4443989
SPI6	SARIMA (3, 0, 5) (2, 1, 2) <sub>6</sub>	0.089463502	0.2340574	0.438179921	0.4989562
	SARIMA (3, 0, 5) (1, 1, 2) <sub>6</sub>	0.089491799	0.2341256	0.427208647	0.4917703
	SARIMA (3, 0, 5) (1, 1, 0) <sub>6</sub>	0.144074068	0.2956569	0.25434774	<b>0.3643462</b>
	<b>SARIMA (3, 0, 5) (1, 1, 3)<sub>6</sub></b>	<b>0.087018215</b>	<b>0.2302998</b>	0.472565367	0.5199601
SPI9	SARIMA (2, 0, 4) (2, 1, 2) <sub>9</sub>	0.137031456	0.2881627	0.152586094	0.2958195
	SARIMA (2, 0, 4) (1, 1, 2) <sub>9</sub>	0.137512738	0.2896313	0.154443812	0.2974688
	SARIMA (2, 0, 4) (1, 1, 1) <sub>9</sub>	0.137562804	0.2898799	0.153798642	0.2972146
	<b>SARIMA (2, 0, 4) (2, 1, 3)<sub>9</sub></b>	<b>0.1366151</b>	<b>0.2880344</b>	<b>0.151674496</b>	<b>0.2955406</b>
SPI12	<b>SARIMA (2, 0, 5) (2, 1, 1)<sub>12</sub></b>	<b>0.093724883</b>	<b>0.2367253</b>	<b>0.085844453</b>	<b>0.2095234</b>
	SARIMA (2, 0, 5) (1, 1, 1) <sub>12</sub>	0.103938665	0.2498215	0.094129615	0.2189612
	SARIMA (2, 0, 5) (1, 1, 0) <sub>12</sub>	0.208200747	0.3515895	0.185092862	0.3171596
	SARIMA (2, 0, 5) (2, 0, 0) <sub>12</sub>	0.103499769	0.2518865	0.096818928	0.2384239

Table 3: The Summary of the Selected Best ARIMA Models

Data	TRAINING		TESTING	
	MSE	MAE	MSE	MAE
SPI3	0.4325422	0.5162857	0.3168685	0.4443989
SPI6	0.0866057	0.2302998	0.2543477	0.3643462
SPI9	0.1366151	0.2880344	0.1516745	0.2955406
SPI12	0.1034997	0.2518865	0.0858444	0.2095234

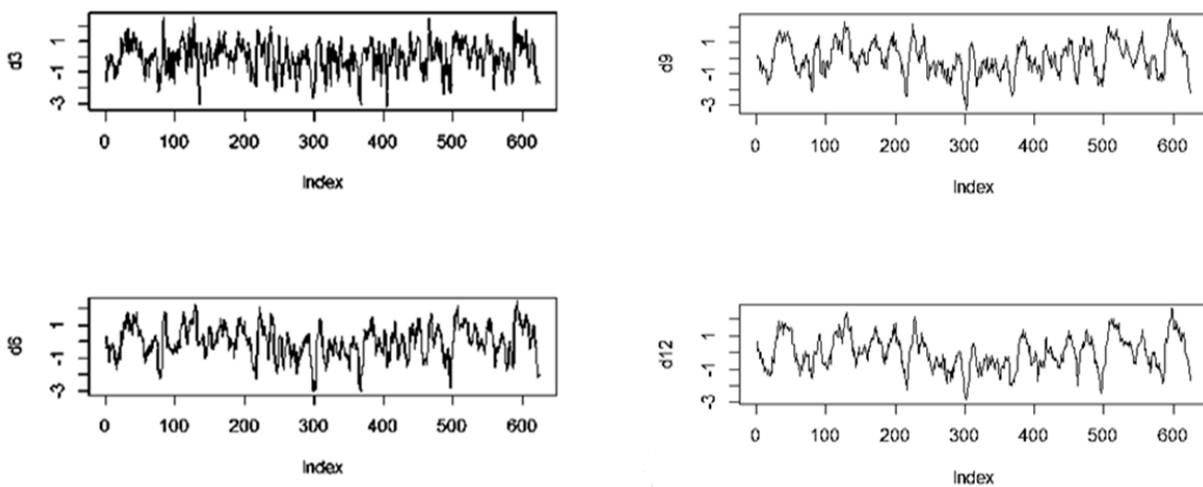


Figure 2: Time Series Plots of the SPI Datasets

1. Model identification: - this involves the selection of a model parameter  $(p, d, q)$  from the PACF and ACF plots.

Figure 3 indicates all the plots of PACF and ACF where models are identified for seasonality and SPI3 exhibits non-seasonality, other SPI data like SPI6, SPI9 and SPI12 are seasonal, hence, seasonal ARIMA model was carried out in the analysis. The orders of  $p, d, q$  is required to determine the best model based on PACF and ACF observed values and after careful observation of the plots of graphs. SPI3 which exhibits non-seasonal from the plot shows a possible ARIMA  $(p, d, q)$  model with  $p = 4$  and  $q = 2$ , the finally selected model after evaluation, was ARIMA  $(4, 0, 2)$  which was chosen. The residual ACF and PACF which has the best model are shown in Figure 3. the ACF and PACF both lie within confidence limits.

2. Model diagnostic Checking: - the third step is to carry out diagnostic checking which involves the standardized, ACF and LJ Box test residuals. Since the models have been fitted to the data, many diagnostic checks were carried out and if the models fit well, the residuals are expected to be uncorrelated with constant variance. However, in the development of model, it is usually assumed that the errors are normally distributed and for this reason, the residuals are expected to be normally distributed.

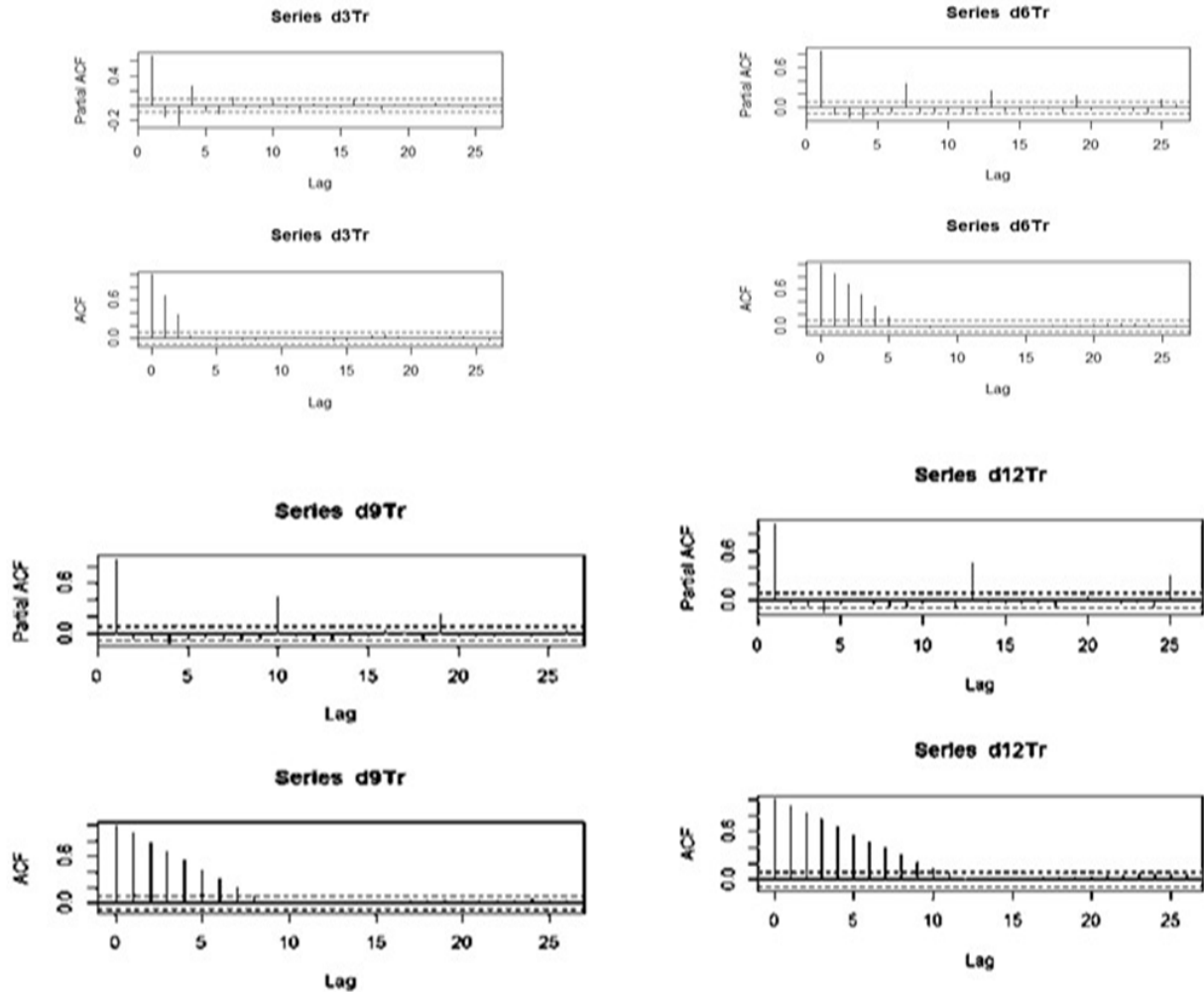


Figure 3: The PACF and ACF plots of the SPI Datasets

Standard checking of goodness-of-fit for ARIMA models is computed based on the PACF and ACF of the residuals as shown in Figure 4 and Figure 5. Since the residuals are normally distributed, they lie on a straight upward sloping line. In evaluating the proposed model, the ARIMA involving time series was used to model SPI3, PI6, SPI9 and SPI12 time series.

### 3.3 Wavelet-ARIMA (W-ARIMA) Model Results

The hybrid of wavelet and ARIMA model is obtained by the combination of two methods: discrete wavelet transforms (DWT) and ARIMA model. In the W-ARIMA, the original SPI data series were decomposed into several data series components which were input to ARIMA to improve the accuracy of the model. For this, the original SPI drought data series is decomposed into ten level components (A, D1, D2, D3, D4, D5, D6, D7, D8,  $D_{L-1}$ ) which stands for different frequency components of the original data up to  $D_{L-1}$ . Where  $L$  is the number of levels which is 9

The forecasting ability of the ARIMA model was improved upon by the application of Wavelet analysis with respect to all the SPI as shown in Table 4. It also improved the forecasting performance measures.

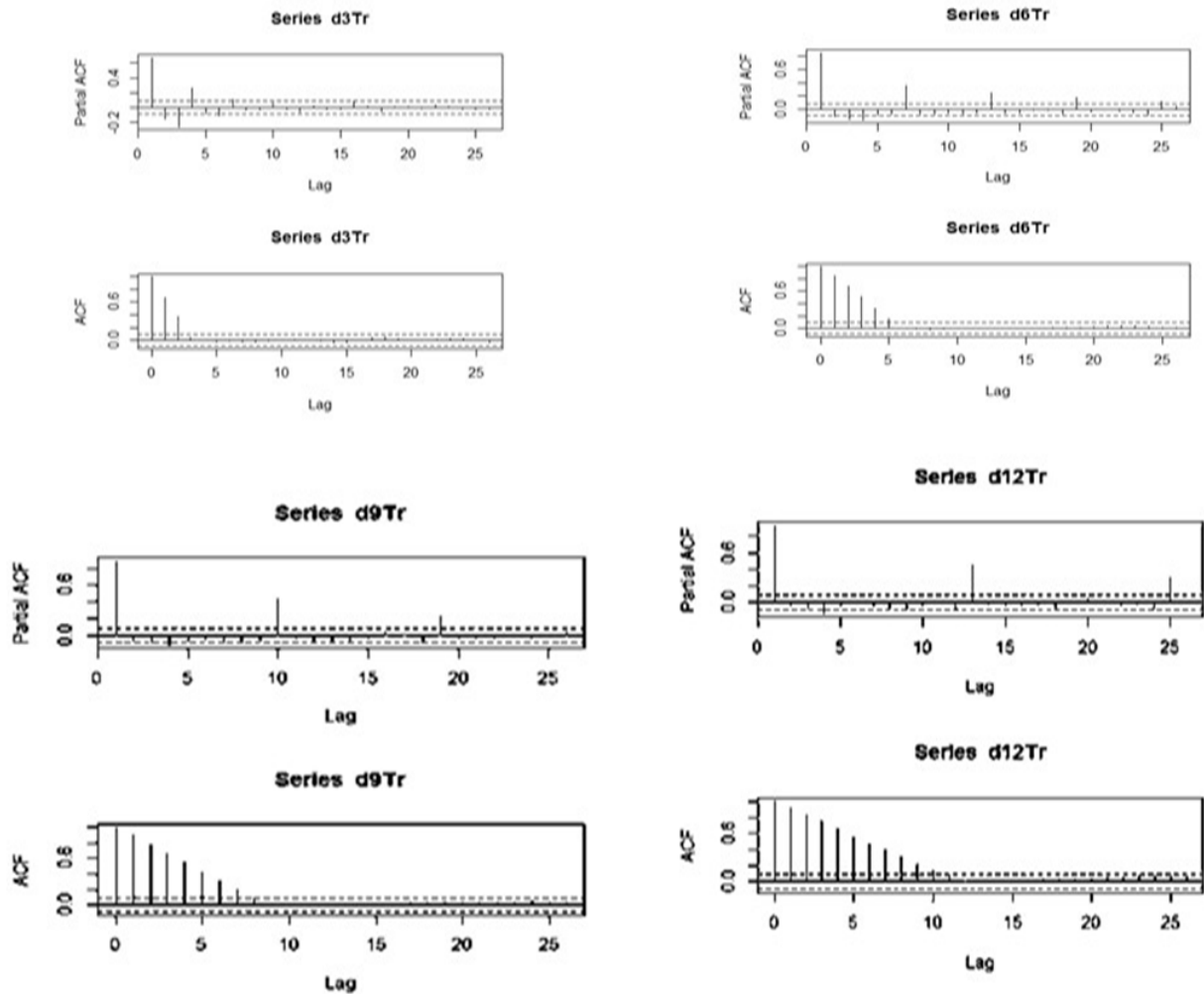


Figure 4: The Plots of Standardized, ACF and P-values of Residuals for all SPIs

Table 4 indicates the measures of goodness of fit which were used to evaluate the forecast performance of all the models.

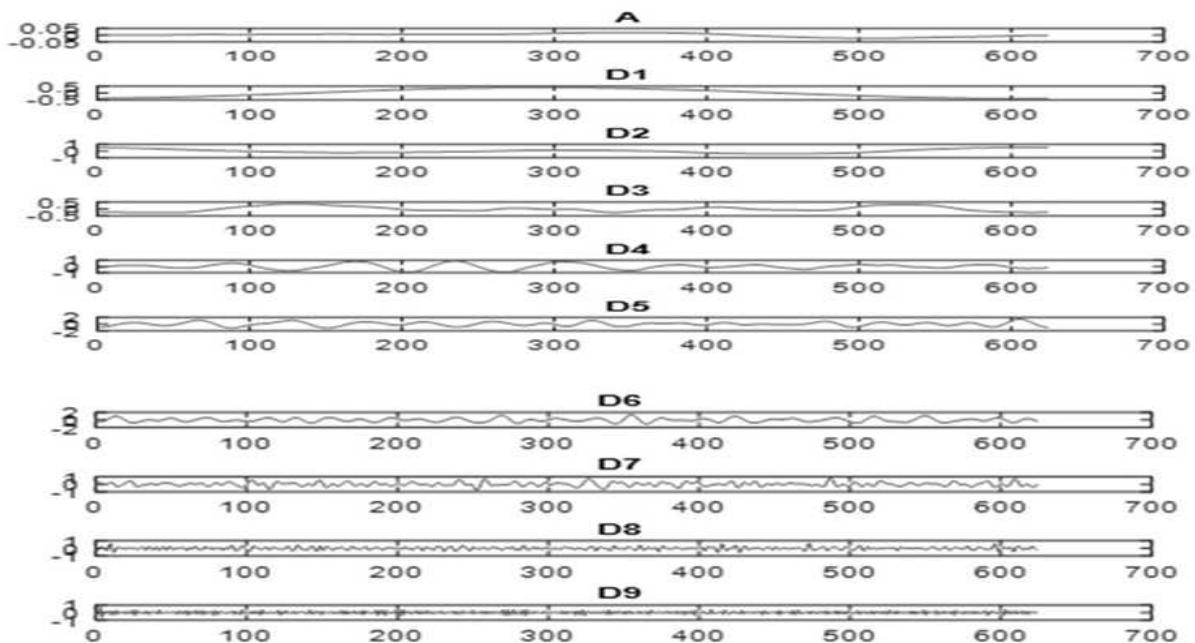
The implication of these results of the Wavelet-ARIMA model showed that the SPI drought data fits the model has improved the result of the ARIMA model.

## 4 Conclusion

In this study, wavelet transforms and ARIMA model were combined to develop a hybrid model for forecasting the SPI drought data. Firstly, the ARIMA model was used to model the SPI data series without preprocessing the data. Secondly, the proposed hybrid model was obtained by combining ARIMA with wavelet transforms and was used to capture the multi-scale features of the SPI data series used to decompose the data. In the study too, the new SPI data series obtained with the addition of effective wavelet components which was employed as an output to the ARIMA model used to forecast the SPI drought. The performance evaluation of the proposed W-ARIMA model which was based on MSE and MAE was then compared with the ARIMA model, which indicated a great improvement in SPI

Table 4: The Results of W-ARIMA Models for All the SPI Data

	TRAINING		TESTING	
	MSE	MAE	MSE	MAE
SPI3	0.212966	0.347828	0.451984	0.532811
SPI6	0.142051	0.293843	0.103373	0.258408
SPI9	0.060979	0.181762	0.043217	0.160485
SPI12	0.052434	0.169528	0.042803	0.148414

Figure 5: The DWT plots for the SPI12 Time Series at Different Decomposition Levels (A, D1, D2, . . . D<sub>L-1</sub>)

drought modeling and produced a better forecast result than the individual ARIMA model. The researchers then concluded that the forecasting capability of W-ARIMA was found to be improved upon when the wavelet transformation technique was used for data preprocessing. The decomposed periodic components obtained from the DWT technique were found to be very effective in getting accurate forecasts when it was employed as inputs in the ARIMA model. The result for the forecasts is an indication that the W-ARIMA model provides an alternative forecasting model to other models and serves as a potential hope as a new method to be used in SPI drought forecasting.

## References

- [1] Shah, V. Ravi, Bharadiya, Nitin, Manekar, Drought index computation using standardized precipitation index (SPI) method for surat district Gujarat. *J. Hydrol.* 2010. 391(1–2): 202–216.
- [2] Shabri, A. A hybrid wavelet analysis and adaptive neuro-fuzzy inference system for drought forecasting. *Appl. Math. Sci.* 2014. 8(139): 6909–6918.
- [3] Dalwadi. The wavelet tutorial. *Internet Resour. [http://engineering.rowan.edu/polikar/WAVELETSWT\\_tutorial.html](http://engineering.rowan.edu/polikar/WAVELETSWT_tutorial.html)*. 2016. 1–67.
- [4] Shijin, L. I. Procedia engineering,” *Appl. Math. Sci.* 2012.
- [5] Bhaskaran, S. Time series data analysis for long term forecasting and scheduling of organizational resources – few cases. 2012. 41(12): 4–9.
- [6] Zhang, L. and Zhou, T. Drought over East Asia: A review. *Journal of Climate* . 2015. 28(8): 3375–3399.
- [7] Cordeiro, C. and Neves, M. Forecasting time series with Boot. EXPOS procedure. *Revstat*, 2009. 7 (2): 135–149.
- [8] Bonaccorso, B., Cancelliere, A. and Rossi, G. Probabilistic forecasting of drought class transitions in Sicily (Italy) using Standardized Precipitation Index and North Atlantic Oscillation Index. *J. Hydrol.* 2015. 526: 136–150.
- [9] Kumar, D. N., Reddy, M. J. and Maity, R. Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence Ttechniques. 2007. 16(4): 307–322.
- [10] Box, G. E. P. and Jenkins, G. M. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco. 1-14. 1976. 1–14.
- [11] Hamzacebi, C. Improving artificial neural networks’ performance in seasonal time series forecasting. *Information Sciences: an International Journal*. 2008. 178 (23): 4550–4559.
- [12] Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model. 2003. 50: 159–175.
- [13] Wei W., Jiang, J., Liang, H., Gao, L., Liang, B., Huang, J., Zang, N., Lio, Y., Yu, J., Lai, J., Qin, F., Su, J., Ye., Li., Chen, H. Application of a combined model with an autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng county, China. *PLoS One*, 2016. 11 (6): 1–13.
- [14] Kirschgässner, G. and Wolters, J. *Introduction to Modern Time Series Analysis*. Springer-Berlin Germany. 2007.
- [15] Szolgayová, E., Arlt, J. Blöschl, G. and Szolgay, J. Wavelet-based deseasonalization for modeling and forecasting of daily discharge series considering long-range dependence. *Journal of Hydrology and Hydromechanics*. 2014. 62(1): 2014. 24–32.
- [16] Udom, P. and Phumchusri, N. A comparison study between time series model and ARIMA model for sales forecasting of the distributor in the plastic industry 2014. *IOSR Journal of Engineering*. 2014. 4(2): 32–38.

- [17] Mckee, T. B., N. J. Doesken, N. J. and Kliest, J. The relationship of drought frequency and duration to time scales. In *Proc. Eighth Conf. of Applied Climatology, Anaheim, CA, Amer. Meteor. Soc.* 1993: 179184.
- [18] Svoboda, M. D. and Hayes, M. J. Appropriate application of the standardized precipitation index in arid locations and dry seasons. 2007. *International Journal of Climatology*. 2007. 27(1): 65–79.
- [19] Paulo, A. A. and Pereira, L. S. Prediction of SPI drought class transitions using Markov chains. *Water Resour. Manag.* 2007. 21 (10): 1813–1827.
- [20] W. Meteorological and Organization. WMO, *WMO Statement on the Status of the Global Climate in 2015 Report*. WMO, UNIC Canberra. 2016
- [21] Wilhite, D. A., Hayes, M. J., Knutson, C. and Smith, K. H. Planning for drought: moving from crisis to risk management. *J. Am. Water Resour. Assoc.* 2000. 36 (4): 697–710.
- [22] Hayes, M., M. Svoboda, D. LeComte, Redmond, K. and Pasteris, P. Drought monitoring: new tools for the 21st century. In Wilhite, D. (Ed). *Drought and Water Crises: Science, Technology, and Management Issues*. Boca Raton, FL: CRC Press, 2005. 5369.
- [23] Nielson-Gammon, J. W. The 2011 Texas drought. *Texas Water J.* 2012. 3(1): 59–95.
- [24] Santos, C. A. G. Morais, B. S. and Silva, G. B. L. Drought forecast using an artificial neural network for three hydrological zones in San Francisco River basin. *IAHS Publication*. 2009. 333. 302–312.
- [25] Barker, L. J., Hannaford, J., Chiverton, A. and Svensson, C. From meteorological to hydrological drought using standardized indicators. *Hydrol. Earth Syst. Sci.* 2016. 20 (6): 2483–2505.
- [26] Belayneh, A., Adamowski, J. and B. Khalil. Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. *Sustain. Water Resour. Manag.* 2016. 2 (1): 87–101.
- [27] Zhou H. and Liu, Y. SPI based meteorological drought assessment over a humid basin: Effects of processing schemes. *Water (Switzerland)*. 2016. 8 (9): 1–16.
- [28] Zhang, X., Peng, Y., Zhang, C. and Wang, B. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting some experiment evidence. *J. Hydrol.*, 2015. 530: 137–152,
- [29] Mishra A. K. and Singh, V. P. Drought modeling - A review. *J. Hydrol.* 2011. 403 (1–2):157–175.
- [30] Nourani, V., Alami, M. T. and Vousoughi, F. D. Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. *J. Hydrol.* 2015. 524: 255–269.
- [31] Dong, B., Mao, Y., Dinov, I. D., Tu, Z. and Shi. Y. Wavelet-based representation of biological shapes wavelet-based representation of biological. *Advances in Visual Computing: 5th International Symposium, ISVC. 2009, Las Vegas, NV, USA, November 30 - December 2*. 2009: 955-964.
- [32] Shabri, A. and Samsudin, R. Fishery landing forecasting using wavelet-based autoregressive integrated moving average models. *Math. Probl. Eng.* 2015. ID 969450: 1-9.
- [33] Kisi, O. and Cimen, M. A wavelet-support vector machine conjunction model for monthly stream flow forecasting. *J. Hydrol.* 2011. 399 (1–2): 132–140.

- [34] Adamowski, J. F. Development of a short-term river flood forecasting method for snow melt-driven floods based on wavelet and cross-wavelet analysis. *Journal of Hydrology*. 2008. 353(3-4): 247-266.
- [35] Moosavi, V., Vafakhah, M., Shirmohammadi, B. and Behnia, N. A wavelet-ANFIS hybrid model for ground water level forecasting for different prediction periods. *Water Resour. Manag.* 2013. 27 (5): 1301–1321.
- [36] Kisi, O. Wavelet regression model for short-term stream flow forecasting. *J. Hydrol.* 2010. 389 (3–4): 344–353.
- [37] Wang, W. and Ding, J. Wavelet network model and its application to the prediction of hydrology. *Nature and Science of Sleep*. 2003. 1(1): 67-71.
- [38] Dawson, C. W., Abrahart, R. J. and See, L. M. HydroTest: a web-based toolbox of evaluation metrics for the standardized assessment of hydrological forecasts. *Environ. Model. Softw.* 2007. 22 (7): 1034–1052.
- [39] Shabri, A. A hybrid wavelet analysis and adaptive neuro-fuzzy inference system for drought forecasting. *Appl. Math. Sci.* 2014. 8 (139): 6909–6918.
- [40] Shabri, A., & Samsudin, R. Fishery landing forecasting using wavelet-based autoregressive integrated moving average models. *Mathematical Problems in Engineering*. 2015. <https://doi.org/10.1155/2015/969450>
- [41] Adamowski, J. F. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. 2008. 247–266. <https://doi.org/10.1016/j.jhydrol..02.013>
- [42] Moosavi, V., Vafakhah, M., Shirmohammadi, B., & Behnia, N. A. Wavelet-ANFIS Hybrid Model for Groundwater Level Forecasting for Different Prediction Periods. *Water Resources Management*. 2013. 27(5): 1301–1321. <https://doi.org/10.1007/s11269-012-0239-2>
- [43] Kisi, O. Wavelet regression model for short-term streamflow forecasting. *Journal of Hydrology*. 2010. 389(3–4): 344–353. <https://doi.org/10.1016/j.jhydrol.2010.06.013>
- [44] Dawson, C. W., Abrahart, R. J., & See, L. M. HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*. 2007. 22(7): 1034–1052. <https://doi.org/10.1016/j.envsoft.2006.06.008>