Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 3, 3102-3129
https://doi.org/10.28919/jmcs/5663
ISSN: 1927-5307

HIGH PERFORMANCE APPROACH FOR WATER LEVEL FORECASTING IN YOM RIVER BASIN OF THAILAND

THAMMARAT PANITYAKUL1[1], PIYAWAN KHWANMUANG[1], RONNASON CHINRAM[1],
PIJITRA JOMSRI[2,*]

[1]Algebra and Applications Research Unit, Division of Computational Science, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

[2]Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Dusit, Bangkok 10300, Thailand

Abstract: Analyses of average monthly water level (AMWL) time series (April 2007 - March 2020) at four water level measurement stations (Y.31, Y.20, Y.1C, Y.37) for wet and dry seasons in the Yom River basin of Thailand. Using Box-Jenkins method, eight best-fit seasonal ARIMA models for one hydrological year forecasting for wet and dry seasons of AMWL were selected; among from twenty-four possible models, by minimum values of AIC, SBIC RMSE, and MAPE. Besides, The comparisons with two benchmark models, Holt-Winters' Seasonal additive method, and Seasonal Naïve method were applied. Results indicated that: The four selected SARIMA models for wet seasons of Y.31, Y20, Y1C, and Y.37 are SARIMA(1,1,1)(1,0,0)[6], SARIMA(1,1,1)(1,0,0)[6], SARIMA(1,1,2)(1,0,0)[6], and SARIMA(1,1,1)(1,0,0)[6], respectively. While the models for dry seasons are SARIMA(0,1,1)(2,0,0)[6], SARIMA(1,1,1)(1,0,1)[6], SARIMA(1,1,1)(2,0,0)[6], and SARIMA(1,0,0)(1,0,1)[6]. The forecasting performance

[*]Corresponding author
E-mail address: pijitra.jo@ssru.ac.th
Received March 8, 2021

are the minimum values of RMSE and MAPE between SARIMA and the benchmark models. The SARIMA model is the best approach for Y.31 Station [Wet Season], Y.31 Station [Dry Season], and Y.37 Station [Dry Season], while the best method for Y.20 Station [Wet Season], Y.1C Station [Wet Season], Y.37 Station [Wet Season], Y.20 Station [Dry Season], and Y.1C Station [Dry Season] is Holt-Winters' seasonal additive method. The upstream station (Y.31 station) has higher accuracy than the downstream station (Y.37 station) due to human activities that disturb hydrological changes. Furthermore, the dry season forecasting is more accurate than the wet season.

Keywords: Box-Jenkins method; SARIMA model; Holt-Winters' method; Seasonal Naïve method; average monthly water level; Yom River basin.

2010 AMS Subject Classification: 62P12.

1. INTRODUCTION

Among 25 main basins of Thailand, The Yom River basin, located in Northern Thailand, with tropical wet and dry or even savanna climate throughout the year. Moreover, the Land Development Department reported in 2013 that Yom River basin has area of 907.75 km^2 with less than 4 droughts per ten year, 701.3 km^2 with 4-5 droughts per ten year, and 229.34 km^2 with over 5 droughts per ten year, which accounted for 49.38, 38.15, and 12.34 percentage of the total area with the repeated drought, and represents 3.77, 2.92, and 0.95 percentage of the Yom River basin area. The Yom River basin has area of 2,454.06 km^2 with less than 4 floods per ten year, 1,334.25 km^2 with 4-7 floods per ten year, and 541.90 km^2 with over 8 floods per ten year, which accounted for 56.67, 30.81, and 12.51 percentage of the total area with the repeated flood, and represents 10.21, 5.55, and 2.25 percentage of the Yom River basin area [1, 2, 3, 4]. The streamflow of a river is the integration of climatic factors and the precipitation. Changes of streamflow may caused by climate changing and the human activities disturbance. These lead to complication of hydrological modeling [4, 5, 6, 7]. Fouli et al., 2018 [8] analysed time series of total quarterly rainfall depth from 1964 to 2014 in Riyadh Region, KSA by seasonal autoregressive integrated moving average models. The best 3 models for 10-year forecasting were selected based on criteria of minimum values of variance, Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC). Puah et al., 2016 [9] analysed of the monthly and seasonal rainfall trends specifically from 1971 to 2012 depending on the availability of data within the Langat River Basin, Malaysia using the additive Holt–Winters method. Sopipan, 2014 [10] study of forecasting historical monthly rainfall data from April 2005 to March 2013 in Nakhon Ratchasima Province, Thailand. Using auto regressive integrated moving average (ARIMA) and multiplicative Holt–Winters method, which the mean absolute percentage error (MAPE), mean squared error (MSE) and mean absolute error (MAE) were used to measure the performance. Forecasts from both methods were found to be acceptable but ARIMA gave a better result for that case. Leo et al., 2016 [11] evaluate the potential of Random Forest, to make streamflow forecast at a 1-day lead time in the Pacific Northwest. Which benchmark the performance against simple Naïve and multiple linear regression models using the calculated Pearson correlation coefficient (r) between forecasted and observed values for each model.

This study propose the appropriate forecasting models for average monthly water level (AMWL) time series of Yom River basin in the Northern of Thailand. Using the Box-Jenkins method, best-fit seasonal autoregressive integrated moving average (SARIMA) models for one hydrological year forecasting for wet and dry seasons of AMWL compare forecasting model with two benchmark models, Holt-Winters' Seasonal additive method, and Seasonal Naïve method. The study period is from April 2007 to March 2020, over thirteen hydrological years.

2. STUDY REGION AND DATASET

Yom River basin located in Northern Thailand. It is river basin covers a surface area of approximately 24,046.89 km^2, between the latitude 14 50' N to 18 25' N and the longitude 99 16' E to 100 40' E. Yom River basin consisted of 11 major sub-river basins and covers administratively 11 provinces, 45 districts, 286 sub-district, and 2,028 villages, as shown in **Fig. 1.** Yom River originates from the Khun Yuam mountain in Phi Pan Nam Range in Pong district and Chiang Muan district, Phayao province, at high slope of 180-360 m(MSL). Then, the stream flow into a large lowland in Phrae province and continues down south through the lowland Sukhothai province. In this area, it flows parallel to the Nan River, at a lower slope of 50-180 m(MSL). Flowing through Phichit province it confluence with Nan River in Nakhon Sawan province at a low slope of 20-50 m(MSL). The length of the Yom River is approximately 735 km [4, 12].

In 2014, the Yom River basin received average annual precipitation of about 1,179 mm, and average annual runoff of about 5,261 million mm^3 and an average annual runoff of fewer than 2,500 mm^3 per year per person, which is less than average annual runoff in Thailand per year per person (3,496 m^3). In 2019, the Yom River basin is one large-sized reservoir and five medium-sized reservoirs with a total storage capacity of 295.62 million m^3 [1, 4, 12, 13].

For retrieving daily water level data in m(MSL) four water level measurement stations were selected over the length of Yom river: Ban Thung Nong [Y.31] station, Ban Huai Sak [Y.20] station, Ban Nam Khong [Y.1C] station, and Ban Wang Chin [Y.37] station. Data was collected from the Upper Northern Region Irrigation Hydrology Center (Royal Irrigation Department) from April 1, 2007 to March 31, 2020, over thirteen hydrological years or 4,649 days [14, 15, 16, 17, 18], as shown in **Fig. 1.** The AMWL data at the four previously listed water level measurement stations for the wet and dry seasons of Thailand was calculated, the wet season is from May to October and the dry season is from November to April.

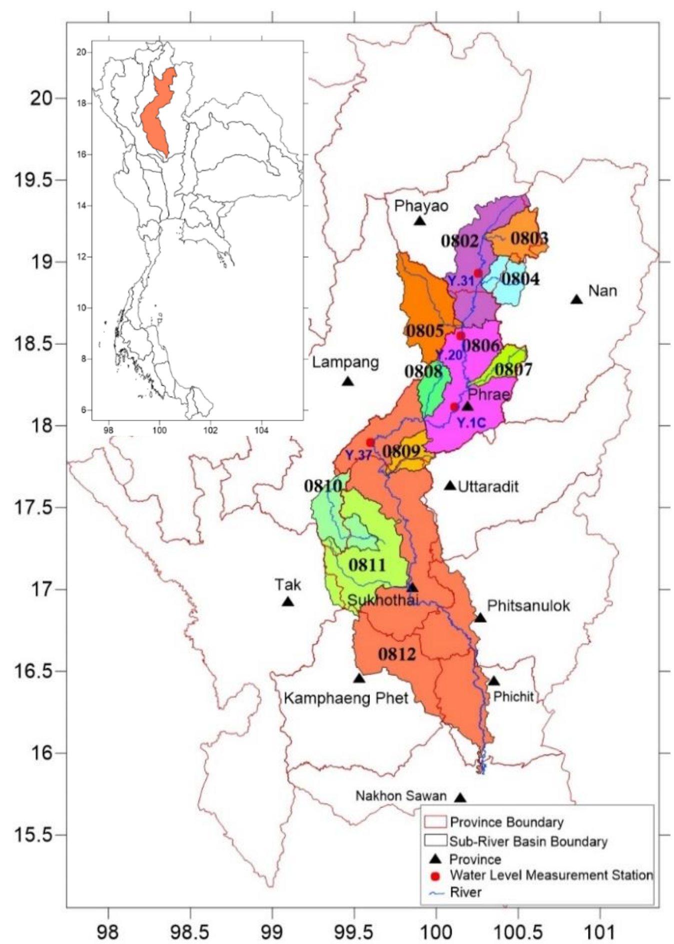

Fig. 1 Location of water level measurement station in Yom River basin.

3. METHODOLOGY

3.1 Seasonal Autoregressive Integrated Moving Average Model: SARIMA Model. The multiplicative autoregressive integrated moving average (ARIMA) model is a universal and widely used model in the time series analysis area. An ARIMA model is a combination of autoregressive (AR) and moving average (MA) parts with a number of differencing. Generally, ARIMA model is denoted by ARIMA(p, d, q) where p, d and q are non-negative integers. In this notation, the p-parameter refers to the autoregressive (AR) part, the d-parameter refers to the order of regular differencing (d) part, and the q-parameter refers to the moving average (MA) part [8, 19].

AR(p) is an autoregressive model of p-order and is represented by:

$$y_t = \sum_{j=1}^{p} \varphi_j y_{t-j} + \varepsilon_t \tag{1}$$

MA(q) is the moving average model of q-order and is represented by:

$$y_t = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{2}$$

Where y_t is the AMWL at time t

φ_j is the autoregressive parameter of j^{th}

θ_j is the moving average parameter of j^{th}

ε_t is independent random variable that represent the error term at time t

The combination between autoregressive model and moving average model is called mixed autoregressive moving average or ARMA(p, q) model and is represented by:

$$y_t = \varepsilon_t + \sum_{j=1}^{p} \varphi_j y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{3}$$

Introducing the back-shift operator (B), where $B^j y_t = y_{t-j}$ for j is a positive integer, the equation (3) may simply be written in terms of the back-shift operator. The ARMA(p, q) model is represented by:

$$\left(1-\sum_{j=1}^{p}\varphi_j B^j\right)y_t=\left(1+\sum_{j=1}^{q}\theta_j B^j\right)\varepsilon_t \tag{4}$$

The ARIMA model assumes that the d^{th} order differencing process follows the ARMA(p,q) model, where the d^{th} order differencing process means the d-times subtracted process. Generally, the d^{th} order differencing process y_t is written $(1-B)^d y_t$. Therefore, the ARIMA(p,d,q) model is represented by:

$$\left(1-\sum_{j=1}^{p}\varphi_j B^j\right)(1-B)^d y_t=\left(1+\sum_{j=1}^{q}\theta_j B^j\right)\varepsilon_t \tag{5}$$

Box and Jenkins then generalized the above model and developed the multiplicative seasonal ARIMA model or ARIMA$(p,d,q)\times(P,D,Q)S$ model is represented by:

$$(1-\Phi_1 B^S-\Phi_2 B^{2S}-\cdots-\Phi_P B^{PS})(1-B^S)^D y_t=\left(1-\Theta_1 B^S-\Theta_2 B^{2S}-\cdots-\Theta_Q B^{QS}\right)\alpha_t \tag{6}$$

Where S is the seasonality

D is the order of the seasonal differencing Process

P is the order of the seasonal Autoregressive polynomial

Q is the order of the seasonal Moving Average polynomial

α_t is the residuals

ARIMA(p,d,q) model is represented by:

$$\left(1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p\right)(1-B)^d\alpha_t=\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)\varepsilon_t \tag{7}$$

Solving the equation (7) for α_t and substituting the equation (6), get the multiplicative seasonal autoregressive integrated moving average model or SARIMA $(p,d,q)\times(P,D,Q)S$ model is represented by:

$$\Phi_P(B^S)\varphi_p(B)(1-B^S)^D(1-B)^d y_t=\Theta_Q(B^S)\theta_q(B)\varepsilon_t \tag{8}$$

Where $\varphi(B)$ is the regular Autoregressive polynomial of order p

$\Phi(B^S)$ is the seasonal Autoregressive polynomial of order P

$\theta(B)$ is the regular Moving Average polynomial of order q

$\Theta(B^S)$ is the seasonal Moving Average polynomial of order Q

$(1-B)^d$ is the d^{th} order differencing process y_t

$(1 - B^S)^D$ is the D^{th} order seasonal differencing process y_t

ε_t is independent random variable that represent the error term at time t

Box-Jenkins methodology used in time series analysis has three steps. The first step, model identification, which model should that time-series be used. The second step, model parameter estimation is specified in the first step. The third step, diagnostic testing is specified in the first step by using the estimation results in the second step as an evaluation. After getting the best-fit model, therefore, the model was used to forecast time series [19, 20].

3.1.2 Model Identification. Consider stationary data because Box-Jenkins methodology requires stationary time series data. A stationary time series of Y_t variable is one whose statistical properties such as mean, variance, autocorrelation etc., are all constant over time.

3.1.3 Model Parameter Estimation. After Identification the model in the first step, the parameters of the model are estimated using Conditional-Sum-of-Squares and Maximum Likelihood (CSS-ML).

3.1.4 Diagnostic Testing. The criteria for consideration is that the independent random variable (ε_t) in the selected subject must qualify as white noise, which is zero mean, variance constant over time, and is independent of other periods. Using the error obtained from the residual (e_t) in the second step as a test instead of ε_t variable, where $e_t = y_t - \hat{y}_t$, because ε_t variable cannot collect data. In practice, the independent random variable (ε_t) may qualify white noise in more than one model. The following ideas are required in determining the best-fit model:

1) The coefficients of all the selected subjects were statistically significant.

2) If those models with independent random variable (ε_t) qualify as white noise, a model with the lowest forecast error may be selected based on the Akaike's Information criterion (AIC) method [21] and considered in conjunction with Sawa's Bayesian information criterion (SBIC), by considering the model with the lowest of these values, it would be most suitable.

$$AIC = n\left(ln\left(\frac{SSE}{n}\right)\right) + 2p \tag{9}$$

$$SBIC = n\left(ln\left(\frac{SSE}{n}\right)\right) + \frac{2(p+2)n\,\hat{\sigma}^2}{SSE} - \frac{2n^2\,\hat{\sigma}^4}{SSE^2} \tag{10}$$

Where p is the number of estimated parameters in the model.

n is the number of observations of Average Monthly Water Level data.

SSE is the Sum of squared errors in the model.

$\hat{\sigma}^2$ is the error variance.

3.2 Holt's winter Seasonal method. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations: the level component, the trend component, and the seasonal component. The additive method is preferred when the seasonal variations are roughly constant through the series, the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component will add up to approximately zero [22, 23, 24]. The component form for the additive method is

$$\hat{y}_{t+h|t} = a_t + hb_t + c_{t+h-s(k+1)} \tag{11}$$

$$a_t = \delta(y_t - c_{t-s}) + (1-\delta)a_{t-1} + b_{t-1} \tag{12}$$

$$b_t = \lambda(a_t - a_{t-1}) + (1-\lambda)b_{t-1} \tag{13}$$

$$c_t = \eta(y_t - a_{t-1} - b_{t-1}) + (1-\eta)c_{t-s} \tag{14}$$

Where $\hat{y}_{t+h|t}$ is the AMWL at time $t+h$

a_t is an estimate of the level of the series at time t

b_t is an estimate of the trend (slope) of the series at time t

c_t is an estimate of the seasonal of the series at time t

δ is the smoothing parameter for the level, $0 \le \delta \le 1$

λ is the smoothing parameter for the level, $0 \le \lambda \le 1$

η is the smoothing parameter for the level, $0 \le \eta \le 1-\delta$

s is the frequency of the seasonality

k is the integer part of $\frac{h-1}{s}$

3.3 Seasonal Naïve method. The naïve method, naïve forecast is optimal when data follow a random walk, these are also called random walk forecasts. Set all forecasts to be the value of the last observation. The seasonal naïve method a similar method that is useful for highly seasonal data. Set each forecast to be equal to the last observed value from the same season of the year [24]. Formally, the forecast for time $n+h$ is written as

$$\hat{y}_{n+h|n} = Y_{n+h-s(k+1)} \tag{15}$$

Where $\hat{y}_{n+h|n}$ is the AMWL at time $t+h$

s is the frequency of the seasonality

k is the integer part of $\frac{h-1}{s}$

3.4 Accuracy and Performance. If the root mean square error (RMSE) and mean absolute percentage error (MAPE) values are small, the forecast value of AMWL is highly accurate.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2} \tag{16}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\frac{|x_t - y_t|}{x_t} \tag{17}$$

Where y_t is the forecast value of AMWL data of $t;\ t = 1, 2, \dots, n$

x_t is the observations of AMWL data of $t;\ t = 1, 2, \dots, n$

n is the number of observations of AMWL data.

The cross-validation method is one of the methods of dividing data to perform performance testing from the variety of prediction models. The basic concept of cross-validation is to split data into multiple parts and use some of the data to predict others. The K-fold cross-validation method is certainly the most popular cross-validation method procedure. In K-fold cross-validation, firstly, the K equal-sized subsamples were partitioned from the original data. Secondly, we treat K−1 subsamples as the training data, i.e. the model fitting dataset, and the remaining subsample as the testing dataset, i.e. be compared with the predicted values from the training model. Finally, we repeat these process K times along with the K subsamples and find the average performance of K-results [25].

4. RESULTS AND DISCUSSION

4.1 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model. Time series plots of AMWL data of four water level measurement stations have relatively clear seasonal variations for the wet season and dry season. There is a peak of the AMWL every hydrological year. The peak and distribution of data in the dry season are less than the wet season, as shown in **Fig. 2** and **Fig. 3.**

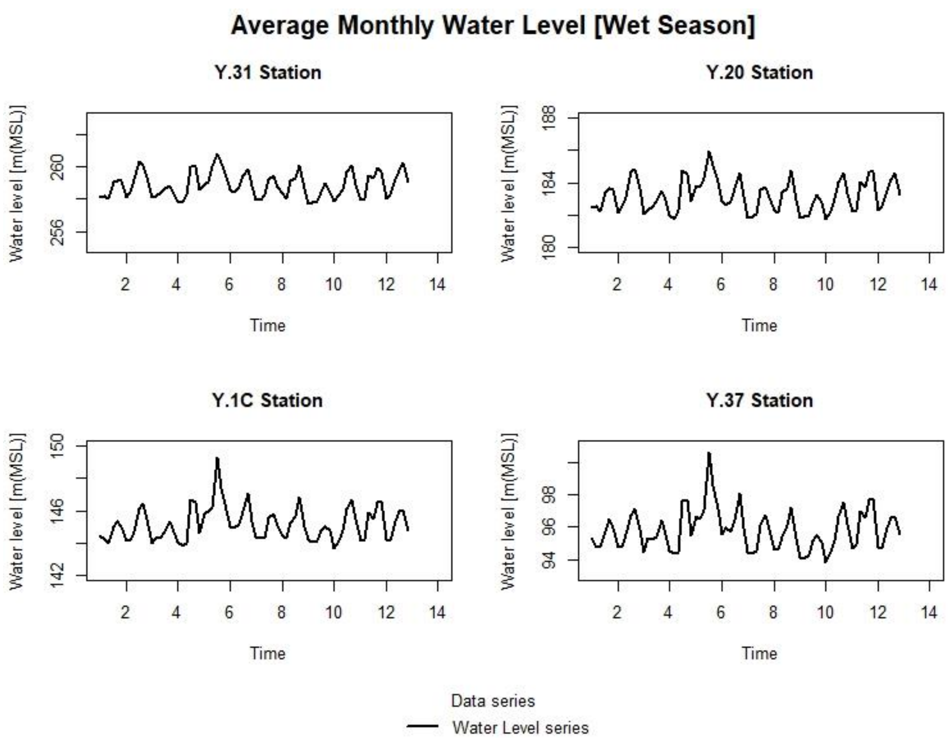

Fig. 2 Time series plots of AMWL data of all four water level measurement stations for wet season.

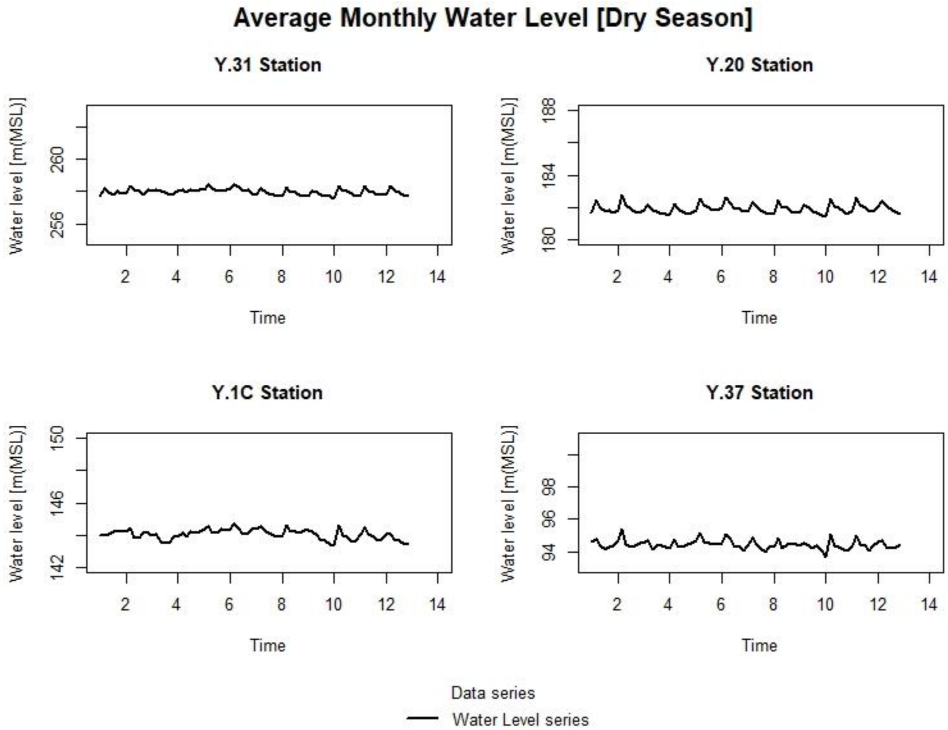

Fig. 3 Time series plots of AMWL data of all four water level measurement stations for dry season.

Consider the assumption of stationary of AMWL data of all four water level measurement stations for wet and dry seasons: The test of time series was constant mean with two-sample t-test method, the test of time series was constant variance with Bartlett's test method, and the test of time series was no autocorrelation with Ljung-Box test method. If time series is non-stationary, transformation with the first-order differencing process. As show the assumption of stationarity in **Table 1** and as show the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots with 99% confidence limits in **Fig. 4**. Though that have some positive and negative spikes of both ACF and PACF plots without 99% confidence limits, but the test of time series with Ljung-Box test method is non-significant (The p-value is more than the significance level = 0.01). Therefore, the time series is stationary.

Based on the twenty-four possible SARIMA models, there are eight selected best-fit SARIMA models of Average Monthly Water Level data of all four water level measurement stations for wet and dry seasons. Parameter estimates or coefficients of the SARIMA models shown in **Table 2,** assumptions of the residuals shown in **Table 3,** as shown the ACF plots with 99% confidence limits demonstrated in **Fig. 5,** and The performance and accuracy of the best-fit SARIMA models stated in **Table 4**

Table 1 Consider the assumption of stationary of AMWL data of all four water level measurement stations for wet and dry seasons.

	Two Sample t-test		**Bartlett test**		**Box-Ljung test**	
	t	**p-value**	**Bartlett's K-squared**	**p-value**	**X-squared**	**p-value**
Y.31 Station [Wet Season]						
$(\boldsymbol{d = 1})$	-0.2676	0.7898	0.2358	0.6273	2.7132	0.0995
Y.20 Station [Wet Season]						
$(\boldsymbol{d = 1})$	0.2133	0.8318	0.0011	0.9735	1.3747	0.2573
Y.1C Station [Wet Season]						
$(\boldsymbol{d = 1})$	0.0874	0.9306	1.1356	0.2866	0.1948	0.6589
Y.37 Station [Wet Season]						
$(\boldsymbol{d = 1})$	0.2611	0.7948	0.1530	0.6957	0.2699	0.6034
Y.31 Station [Dry Season]						
$(\boldsymbol{d = 1})$	2.2892	0.0251*	1.1401	0.2856	5.6360	0.0176*
Y.20 Station [Dry Season]						
$(\boldsymbol{d = 1})$	0.1042	0.9173	1.4739	0.2247	3.1279	0.0770
Y.1C Station [Dry Season]						
$(\boldsymbol{d = 1})$	-0.0218	0.9827	2.2306	0.1353	2.7425	0.0977
Y.37 Station [Dry Season]						
$(\boldsymbol{d = 0})$	2.1883	0.0320	0.0216	0.8831	5.5816	0.0182*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05

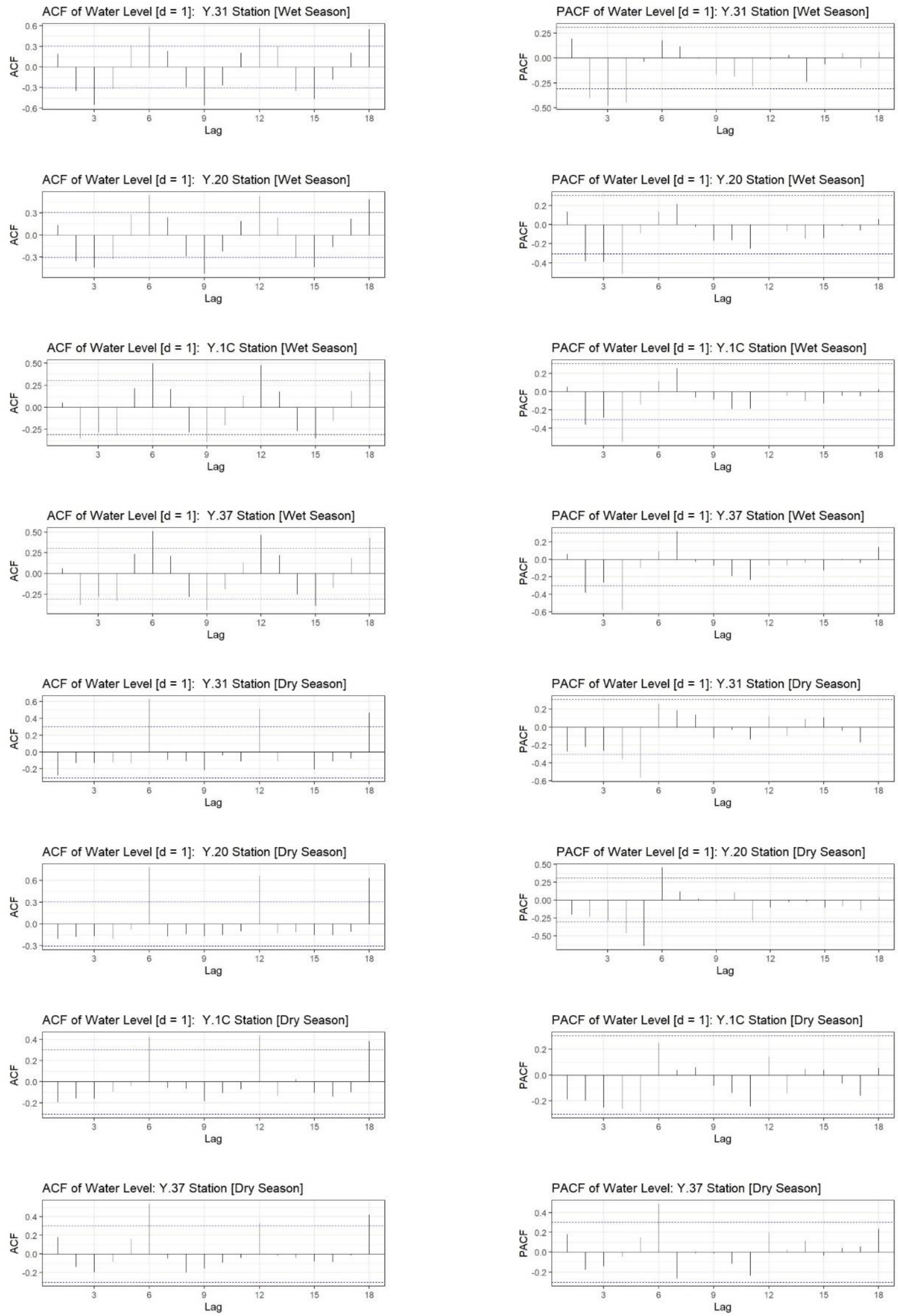

Fig. 4 The ACF and PACF plots with 99% confidence limits of AMWL data of all four water level measurement stations for wet and dry seasons.

Table 2 Parameter estimates of the SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons.

Model	Estimate (p-value)					
	AR (1)	**MA (1)**	**MA (2)**	**SAR (1)**	**SAR (2)**	**SMA (1)**
Wet Season						
Y.31 Station: SARIMA (1,1,1)(1,0,0)[6]	0.6457 (3.400×10^{-12}) ***	- 1.0000 ($< 2.2\times10^{-16}$) ***		0.6459 (3.386×10^{-12}) ***		
Y.20 Station: SARIMA (1,1,1)(1,0,0)[6]	0.5937 ($1.475\times10^{-9)}$ ***	- 1.0000 ($< 2.2\times10^{-16}$) ***		0.5934 (1.643×10^{-9}) ***		
Y.1C Station: SARIMA (1,1,2)(1,0,0)[6]	- 0.6800 (4.639×10^{-10}) ***	0.4789 (2.954×10^{-3}) **	- 0.5211 (7.471×10^{-4}) ***	0.6021 (1.155×10^{-9}) ***		
Y37 Station: SARIMA (1,1,1)(1,0,0)[6]	0.5570 ($3.984\times^{10-8}$) ***	- 1.0000 ($< 2.2\times10^{-16}$) ***		0.5411 (6.143×10^{-8}) ***		
Dry Season						
Y.31 Station: SARIMA (0,1,1)(2,0,0)[6]		- 0.6862 (5.354×10^{-13}) ***		0.4737 (2.168×10^{-5}) ***	0.3626 (2.9×10^{-3}) **	
Y.20 Station: SARIMA (1,1,1)(1,0,1)[6]	0.5021 (1.152×10^{-7}) ***	- 1.0000 ($< 2.2\times10^{-16}$) ***		1.0000 ($< 2.2\times10^{-16}$) ***		- 0.9817 ($< 2.2\times10^{-16}$) ***
Y.1C Station: SARIMA (1,1,1)(2,0,0)[6]	0.6582 (4.331×10^{-12}) ***	- 1.0000 ($< 2.2\times10^{-16}$) ***		0.2887 (7.6×10^{-3}) ***	0.3514 (2.9×10^{-3}) **	
Y37 Station: SARIMA (1,0,0)(1,0,1)[6]	0.3408 (2.5×10^{-3}) **			0.9991 ($< 2.2\times10^{-16}$) ***		- 0.9519 (8.989×10^{-6}) ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

Parameter estimates of the SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons were statistically significant (The p-value is less than the significance level = 0.01). Therefore, the parameter estimates were included in the model, as shown in **Table 2.**

Table 3 Consider the assumptions of the residuals of AMWL data of all four water level measurement stations for wet and dry seasons.

Model	**One Sample t-test**		**Bartlett test**		**Box-Ljung test**	
	t	**p-value**	**Bartlett's K-squared**	**p-value**	**X-squared**	**p-value**
Y.31 Station [Wet Season]						
SARIMA(1,1,1)(1,0,0)[6]	0.6121	0.5424	0.0170	0.8964	0.3092	0.5781
Y.20 Station [Wet Season]						
SARIMA(1,1,1)(1,0,0)[6]	0.4469	0.6563	0.0009	0.9760	0.0625	0.8026
Y.1C Station [Wet Season]						
SARIMA(1,1,2)(1,0,0)[6]	- 0.0501	0.9601	0.1311	0.7173	0.0269	0.8697
Y.37 Station [Wet Season]						
SARIMA(1,1,1)(1,0,0)[6]	0.2054	0.8378	0.2040	0.6515	0.0032	0.9552
Y.31 Station [Dry Season]						
SARIMA(0,1,1)(2,0,0)[6]	0.0659	0.9477	2.0019	0.1571	0.1788	0.6724
Y.20 Station [Dry Season]						
SARIMA(1,1,1)(1,0,1)[6]	0.7680	0.4450	0.7581	0.3839	0.3198	0.5717
Y.1C Station [Dry Season]						
SARIMA(1,1,1)(2,0,0)[6]	-0.4867	0.6280	4.2685	0.0388*	0.0754	0.7836
Y.37 Station [Dry Season]						
SARIMA(1,0,0)(1,0,1)[6]	-1.0335	0.3049	0.3221	0.5704	0.0294	0.8639

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05

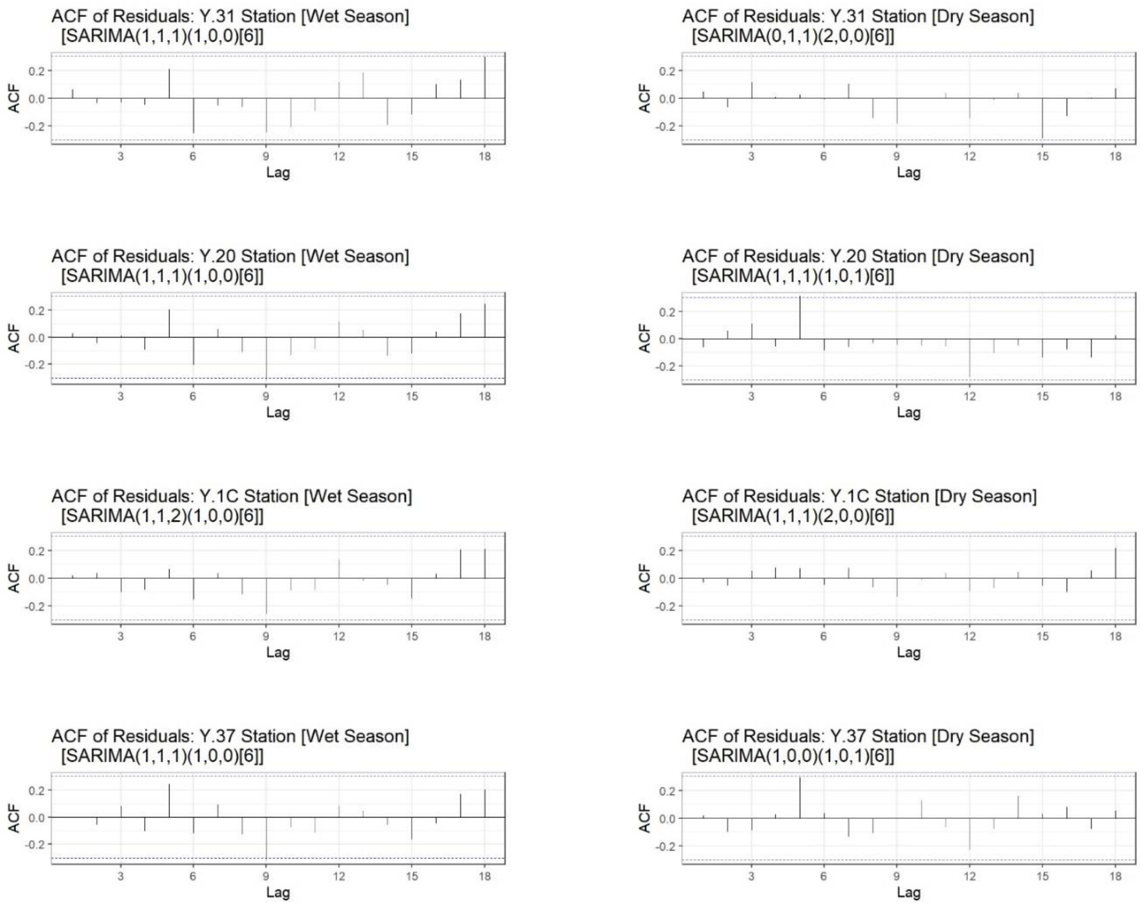

Fig. 5 The ACF of residuals plots with 99% confidence limits of the best-fit SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons.

Consider the assumptions of the residuals of AMWL data of all four water level measurement stations for wet and dry seasons: The test of time series was zero mean with one-sample t-test method, The test of time series was constant variance with Bartlett's test method, and the test of time series was no autocorrelation with Ljung-Box test method, as shown in **Table 3**. The residuals testing of all three methods showed non-significant (The p-value is more than the significance level = 0.01) and most positive and negative spikes of the ACF of the residuals plots within 99% confidence limits, as shown in **Fig. 5.** Therefore the residuals are as correct according to the assumption as white noise. Include The performance and accuracy of the best-fit SARIMA models as shown in **Table 4.**

Table 4 The relative performance and accuracy of the best-fit SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons.

Model	All Time Series Data				Validation Data			
	AIC	**BIC**	**RMSE**	**MAPE**	**AIC**	**BIC**	**RMSE**	**MAPE**
Y.31 Station [Wet Season]								
SARIMA(1,1,1)(1,0,0)[6]	116.78	125.83	0.5024	0.1559	36.43	39.77	0.4768	0.1506
Y.20 Station [Wet Season]								
SARIMA(1,1,1)(1,0,0)[6]	159.48	168.53	0.6791	0.2988	46.20	49.54	0.6505	0.2799
Y.1C Station [Wet Season]								
SARIMA(1,1,2)(1,0,0)[6]	175.40	186.72	0.7493	0.3913	50.76	54.93	0.7354	0.6000
Y.37 Station [Wet Season]								
SARIMA(1,1,1)(1,0,0)[6]	196.64	205.69	0.8833	0.7244	52.03	55.37	0.8141	0.6771
Y.31 Station [Dry Season]								
SARIMA(0,1,1)(2,0,0)[6]	-72.1	-63.05	0.1336	0.0382	27.45	31.99	0.3397	0.0971
Y.20 Station [Dry Season]								
SARIMA(1,1,1)(1,0,1)[6]	-66.58	-55.27	0.1207	0.0517	46.21	51.89	0.4653	0.1816
Y.1C Station [Dry Season]								
SARIMA(1,1,1)(2,0,0)[6]	-4.23	7.09	0.2120	0.1090	37.22	42.90	0.4201	0.2066
Y.37 Station [Dry Season]								
SARIMA(1,0,0)(1,0,1)[6]	-12.29	-0.91	0.1881	0.1453	48.41	54.30	0.5210	0.3739

The accuracy of best-fit SARIMA models, both the AMWL all 12 hydrological years time series and the AMWL time series with cross-validation method, the result showed that values of RMSE and MAPE during the dry season has higher than in the wet season all four water level measurement stations. Besides, The accuracy at Y.31 station, located upstream has higher than Y.37 station, located downstream. Due to hydrological changes that may be caused by human activities such as reservoirs, water use for agriculture or consumption of Y.37 station is greater.

The eight selected best-fit SARIMA models based on the twenty-four possible SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons as follows:

Y.31 Station [Wet Season]: SARIMA(1,1,1)(1,0,0)[6] model, all spikes of the ACF of the residuals plots within 99% confidence limits and residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = 116.78, SBIC = 125.83, RMSE = 0.5024, MAPE = 0.1559; Validation Data: AIC = 36.43, SBIC = 39.77, RMSE = 0.4768, MAPE = 0.1506)

Y.20 Station [Wet Season]: SARIMA(1 ,1 ,1) (1 ,0 ,0) [6] model, most spikes of the ACF of the residuals plots within 99% confidence limits, except negative spikes at lag 9, but residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = 159.48, SBIC = 168.53, RMSE = 0.6791, MAPE = 0.2988; Validation Data: AIC = 46.20, SBIC = 49.54, RMSE = 0.6505, MAPE = 0.2799)

Y.1C Station [Wet Season]: SARIMA(1,1,2)(1,0,0)[6] model, all spikes of the ACF of the residuals plots within 99% confidence limits and residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = 175.40, SBIC = 186.72, RMSE = 0.7493, MAPE = 0.3913; Validation Data: AIC = 50.76, SBIC = 54.93, RMSE = 0.7354, MAPE = 0.600)

Y.37 Station [Wet Season]: SARIMA(1,1,1)(1,0,0)[6] model, most spikes of the ACF of the residuals plots within 99% confidence limits, except negative spikes at lag 9, but residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = 196.64, SBIC = 205.69, RMSE = 0.8833, MAPE = 0.7244; Validation Data: AIC = 52.03, SBIC = 55.37, RMSE = 0.8141, MAPE = 0.6771)

Y.31 Station [Dry Season]: SARIMA(0,1,1)(2,0,0)[6] model, all spikes of the ACF of the residuals plots within 99% confidence limits and residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = -72.1, SBIC = -63.05, RMSE = 0.1336, MAPE = 0.0382; Validation Data: AIC = 27.45, SBIC = 31.99, RMSE

= 0.3397, MAPE = 0.0971)

Y.20 Station [Dry Season]: SARIMA(1,1,1)(1,0,1)[6] model, most spikes of the ACF of the residuals plots within 99% confidence limits, except positive spikes at lag 5, but residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = -66.58, SBIC = -55.27, RMSE = 0.1207, MAPE = 0.0517; Validation Data: AIC = 46.21, SBIC = 51.89, RMSE = 0.4653, MAPE = 0.1816)

Y.1C Station [Dry Season]: SARIMA(1,1,1)(2,0,0)[6] model, all spikes of the ACF of the residuals plots within 99% confidence limits and residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = -4.23, SBIC = 7.09, RMSE = 0.2120, MAPE = 0.1090; Validation Data: AIC = 37.22, SBIC = 42.90, RMSE = 0.4201, MAPE = 0.2066)

Y.37 Station [Dry Season]: SARIMA(1,0,0)(1,0,1)[6] model, all spikes of the ACF of the residuals plots within 99% confidence limits and residuals testing of all three methods showed non-significant. Therefore, the residuals is white noise. (All Time Series Data: AIC = -12.29, SBIC = -0.91, RMSE = 0.1881, MAPE = 0.1453; Validation Data: AIC = 48.41, SBIC = 54.30, RMSE = 0.5210, MAPE = 0.3739)

Table 5 The equation of the best-fit SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons.

Model	Equation of the best-fit SARIMA model
Y.31 Station [Wet Season] SARIMA(1,1,1)(1,0,0)[6]	$y_t = e_t + 1e_{t-1} + 1.6457y_{t-1} - 0.6457y_{t-2} + 0.6459y_{t-6} - 1.0630y_{t-7} + 0.4171y_{t-8}$
Y.20 Station [Wet Season] SARIMA(1,1,1)(1,0,0)[6]	$y_t = e_t + 1e_{t-1} + 1.5937y_{t-1} - 0.5937y_{t-2} + 0.5934y_{t-6} - 0.9457y_{t-7} + 0.3523y_{t-8}$
Y.1C Station [Wet Season] SARIMA(1,1,2)(1,0,0)[6]	$y_t = e_t - 0.4789e_{t-1} + 0.5211\varepsilon_{t-2} + 0.32y_{t-1} + 0.6800y_{t-2} + 0.6021y_{t-6} - 0.5906y_{t-7} - 0.4094y_{t-8}$
Y.37 Station [Wet Season] SARIMA(1,1,1)(1,0,0)[6]	$y_t = e_t + 1e_{t-1} + 1.5570y_{t-1} - 0.5570y_{t-2} + 0.5411y_{t-6} - 0.8425y_{t-7} + 0.3014y_{t-8}$
Y.31 Station [Dry Season] SARIMA(0,1,1)(2,0,0)[6]	$y_t = e_t + 0.6862e_{t-1} + y_{t-1} + 0.4737y_{t-6} - 0.4737y_{t-7} + 0.3626y_{t-12} - 0.3626y_{t-13}$
Y.20 Station [Dry Season] SARIMA(1,1,1)(1,0,1)[6]	$y_t = e_t + 1e_{t-1} + 0.9817e_{t-6} + 0.9817e_{t-7} + 1.5021y_{t-1} - 0.5021y_{t-2} + 1y_{t-6} - 1.5021y_{t-7} + 1.5021y_{t-8}$
Y.1C Station [Dry Season] SARIMA(1,1,1)(2,0,0)[6]	$y_t = e_t + 1e_{t-1} + 1.6582y_{t-1} - 0.6582y_{t-2} + 0.2887y_{t-6} - 0.4787y_{t-7} + 0.19y_{t-8} + 0.3514y_{t-12} - 0.5827y_{t-13} + 0.2313y_{t-14}$
Y.37 Station [Dry Season] SARIMA(1,0,0)(1,0,1)[6]	$y_t = e_t + 0.9519e_{t-6} + 0.3408y_{t-1} + 0.9991y_{t-6} - 0.3405y_{t-7}$

The forecasting equation of the best-fit SARIMA models in **Table 5.** Including coefficients from **Table 2** with the historical AMWL time series, which is different from each station. The historical time series was based on the parameter of the SARIMA$(p, d, q) \times (P, D, Q)S$ model in equation (8).

4.2 Comparisons of Forecasting Performance. Comparisons of best-fit SARIMA models and the two benchmark models reveal the similarity of RMSE and MAPE for both wet season forecasting (six months: May 2019 – October 2029) and dry season forecasting (six months: November 2019 – April 2020). The forecasting performance are the minimum values of RMSE and MAPE between SARIMA and benchmark models. The SARIMA model is the best approach for Y.31 Station [Wet Season], Y.31 Station [Dry Season], and Y.37 Station [Dry Season], (the RMSE value of 0.9447, 0.1480, and 0.3093; the MAPE value of 0.3170, 0.0556, and 0.2809), while the best method for Y.20 Station [Wet Season], Y.1C Station [Wet Season], Y.37 Station [Wet Season], Y.20 Station [Dry Season], and Y.1C Station [Dry Season] (the RMSE value of 1.1315, 1.1641, 0.8440, 0.1319, and 0.0905; the MAPE value of 0.5340, 0.6717, 0.7715, 0.0602, and 0.0586) is Holt-Winters' seasonal additive method. As shown in **Table 6** and **Fig. 6.** However, the Box-Jenkins method is a time series analysis by using the stochastic process such that yields the forecast values without overfitting problem.

Forecasting of the three method of AMWL data of all four water level measurement stations for wet season as shown in **Fig. 7 and** dry season as shown in **Fig. 8.**

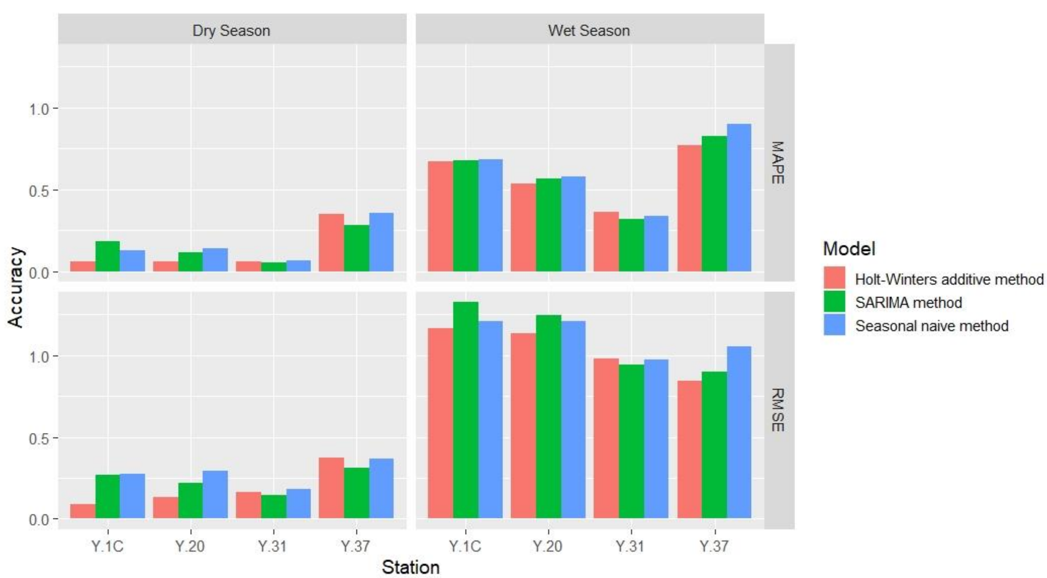

Fig. 6 Comparisons of accuracy forecasting AMWL values of all four water level measurement stations for wet and dry seasons.

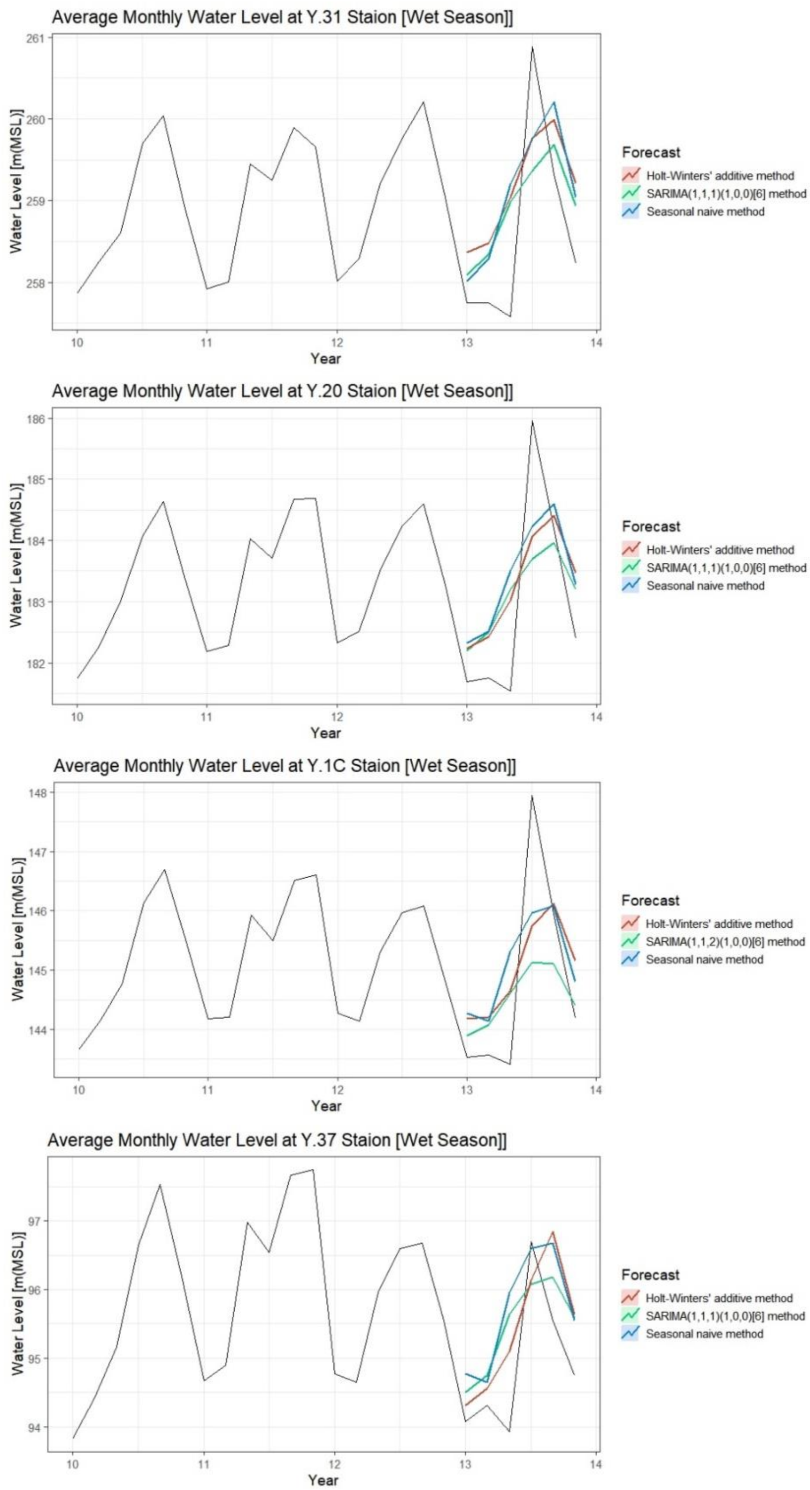

Fig. 7 Forecasting of the three methods of AMWL data of all four water level measurement stations, show only last three hydrological years and forecasting one hydrological year of wet season.

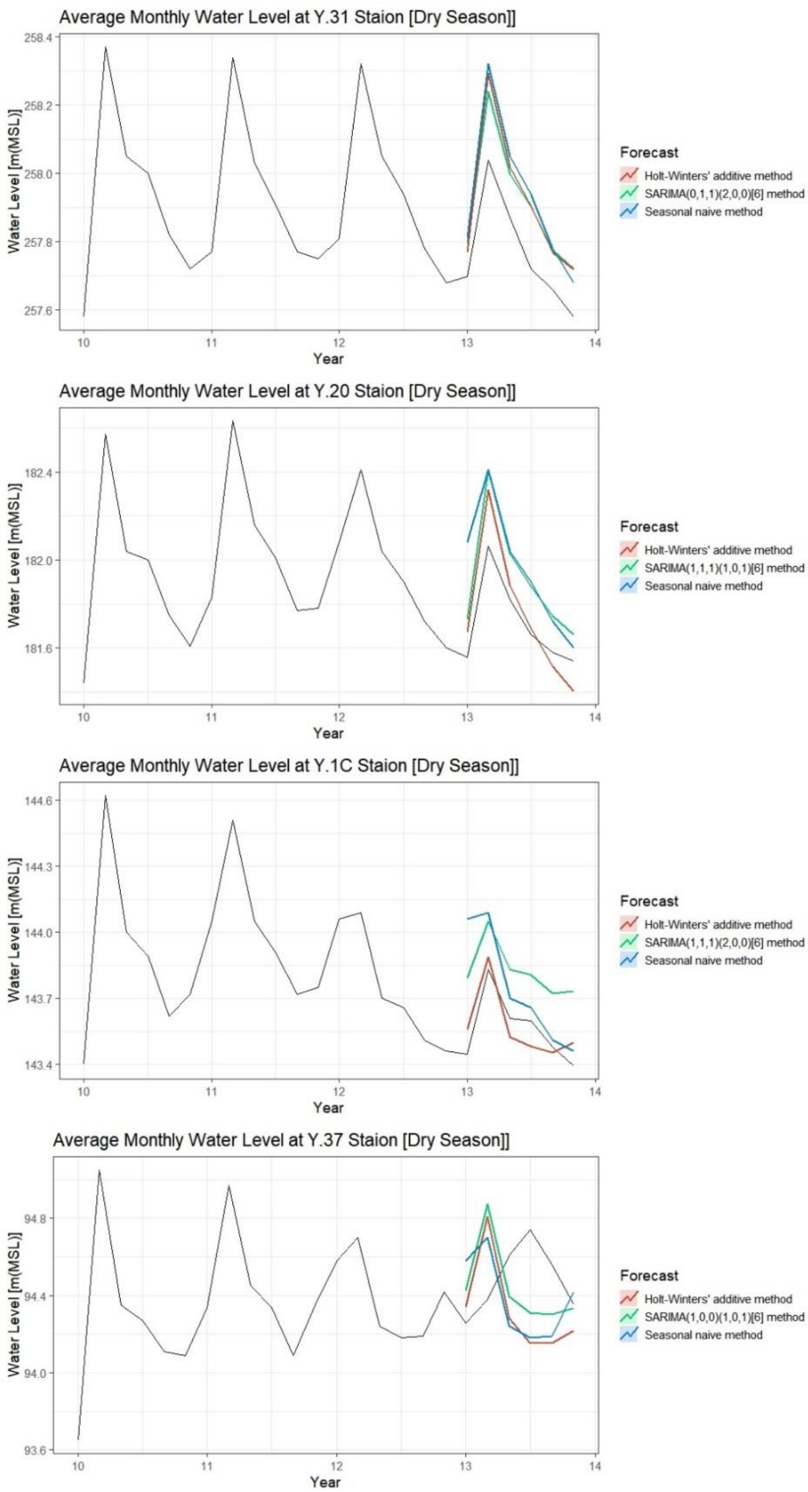

Fig. 8 Forecasting of the three methods of AMWL data of all four water level measurement stations, show only last three hydrological years and forecasting one hydrological year of dry season.

Table 6 Comparison of forecasting performance and accuracy of AMWL data of all four water level measurement stations for wet and dry seasons.

Season	Station	Relative Performance		Accuracy					
		AIC	BIC	RMSE			MAPE		
		SARIMA	SARIMA	SARIMA	B1	B2	SARIMA	B1	B2
Wet Season	**Y.31**	116.78	125.83	0.9447	0.9773	0.9745	0.3170	0.3600	0.3385
	Y.20	159.48	168.53	1.2475	1.1315	1.2061	0.5637	0.5340	0.5798
	Y.1C	175.4	186.72	1.3236	1.1641	1.2098	0.6797	0.6717	0.6814
	Y.37	196.64	205.69	0.8961	0.8440	1.0543	0.8244	0.7715	0.8989
Dry Season	**Y.31**	-72.1	-63.05	0.1480	0.1611	0.1809	0.0556	0.0583	0.0655
	Y.20	-66.58	-55.27	0.2162	0.1319	0.2947	0.1126	0.0602	0.1401
	Y.1C	-4.23	7.09	0.2690	0.0905	0.2768	0.1830	0.0586	0.1297
	Y.37	-12.29	-0.91	0.3093	0.3725	0.3651	0.2809	0.3489	0.3544

B1 is Holt-Winters' Seasonal additive method

B2 is Seasonal Naïve method

5. CONCLUSIONS

The AMWL time series from April 2007 to March 2020, at four water level measurement stations (Y.31, Y.20, Y.1C, Y.37) for wet and dry seasons in the Yom River basin had analyzed. The following conclusions and recommendations can be obtained from this study:

5.1 The four selected SARIMA models of four water level measurement stations (Y.31, Y.20, Y.1C, Y.37) for wet seasons are: SARIMA(1,1,1)(1,0,0)[6], SARIMA(1,1,1)(1,0,0)[6], SARIMA(1,1,2)(1,0,0)[6], and SARIMA(1,1,1)(1,0,0)[6], respectively. The four selected seasonal ARIMA models for dry seasons are: SARIMA(0,1,1)(2,0,0)[6], SARIMA(1,1,1)(1,0,1)[6], SARIMA(1,1,1)(2,0,0)[6], and SARIMA(1,0,0)(1,0,1)[6], respectively. These models were selected from more than 24 models based on criteria that included minimum values of AIC and SBIC.

5.2 Comparisons of best-fit SARIMA models and the two benchmark models reveal the similarity of RMSE and MAPE for both wet season forecasting (six months: May 2019 – October 2029) and dry season forecasting (six months: November 2019 – April 2020). The forecasting performance are the minimum values of RMSE and MAPE between SARIMA and benchmark models. The SARIMA model is the best approach for Y.31 Station [Wet Season], Y.31 Station [Dry Season], and Y.37 Station [Dry Season], (the RMSE value of 0.9447, 0.1480, and 0.3093; the MAPE value of 0.3170, 0.0556, and 0.2809), while the best method for Y.20 Station [Wet Season], Y.1C Station [Wet Season], Y.37 Station [Wet Season], Y.20 Station [Dry Season], and Y.1C Station [Dry Season] (the RMSE value of 1.1315, 1.1641, 0.8440, 0.1319, and 0.0905; the MAPE value of 0.5340, 0.6717, 0.7715, 0.0602, and 0.0586) is Holt-Winters' seasonal additive method. However, the Box-Jenkins method is a time series analysis by using the stochastic process such that yields the forecast values without overfitting problem.

5.3 The upstream station (Y.31 station) has higher accuracy than the downstream station (Y.37 station) due to human activities (reservoirs, water use for agriculture and consumption) that disturb hydrological changes. Human activities and climatic changes cause data uncertainty in forecasting.

5.4 The dry season forecasting is more accurate than the wet season all four water level measurement stations due to the seasonal variation of the AMWL time series during the wet season is more complicated than the dry season and the magnitude of data dispersion in wet season is wider than dry season.

ACKNOWLEDGEMENT

The authors would like to thank the Meteorological Department and the Royal Irrigation Department for valuable data of this paper.

The authors would like to thank Suan Sunandha Rajabhat University for scholarship support.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] Office of the National Water Resources, The National Water Resources Management Strategies (2015-2026), Retrieved from http://www.onwr.go.th/?page id=3684, (2015).

[2] Department of Water Resources, Water Situation: Drought area, Retrieved from http://data.dwr.go.th/dwr/watersituation/drought/index, (2013).

[3] Department of Water Resources, Water Situation: Flood Area, Retrieved from http://data.dwr.go.th/dwr/watersituation/deluge/index, (2013).

[4] P. Khwanmuang, R. Chinram, T. Panityakul, The hydrological and water level data in Yom River basin of Thailand, J. Math. Comput. Sci. 10 (2020), 3026-3047.

[5] Q. Zhang, C-Y. Xu, Y.D. Chen, J. Jiang, Abrupt behaviors of the streamflow of the Pearl River basin and implications for hydrological alterations across the Pearl River Delta, China, J. Hydrol. 377 (2009), 274-283.

[6] Q. Zhang, X. Gu, V.P. Singh, X. Chen, Evaluation of ecological instream flow using multiple ecological indicators with consideration of hydrological alterations, J. Hydrol. 529 (2015), 711-722.

[7] Q. Zhang, X. Gu, V. P. Singh, C-Y. Xu, D. Kong, M. Xiao, X. Chen, Homogenization of precipitation and flow regimes across China: changing properties, causes and implications, J. Hydrol. 530 (2015), 462-475.

[8] H. Fouli, R. Fouli, B. Bashir, O.A. Loni, Seasonal Forecasting of Rainfall and Runoff Volumes in Riyadh Region, KSA, KSCE J. Civil Eng. 22 (2018), 2637-2647.

[9] Y.J. Puah, Y.F. Huang, K.C. Chua, T.S. Lee, River catchment rainfall series analysis using additive holt–Winters method, J. Earth Syst. Sci. 125 (2016), 269-283.

[10] N. Sopipan, Forecasting Rainfall in Thailand: A Case Study of Nakhon Ratchasima Province, Int. J. Environ. Ecol. Geol. Marine Eng. 8 (2014), 712-716.

[11] T.P. Leo, L. Lifeng, O.F. Andrew, Evaluation of Random Forest for short-term daily streamflow forecast in rainfall and snowmelt driven watersheds, preprint, https://doi.org/10.5194/hess-2020-305.

[12] Hydro - Informatics Institute (Public Organization), Implementation of data collection and analysis of data on

the 25 river basin data warehouse of the system development project, and modeling flood and drought (Yom River basin), Retrieved from http://www.thaiwater.net/web/attachments/25basins/08-yom.pdf, (2012).

[13] Bureau of Water Management and Hydrology (Royal Irrigation Department), Water allocation and cultivation plan for irrigation area during the dry season in 2019 - 2020, Retrieved from http://water.rid.go.th/hwm/wmoc/planing/dry/manage water2562-63.pdf, (2019).

[14] Upper Northern Region Irrigation Hydrology Center (Royal Irrigation Department), Average daily of water level - water content data, (2020).

[15] Upper Northern Region Irrigation Hydrology Center (Royal Irrigation Department), Station history: Hydrological station of Yom River (Y.1C), Ban Nam Khong, Pa Maet sub-district, Mueang Phrae district, Phrae province, (n.d.). Retrieved from https://www.hydro-1.net/Data/STATION/History/Y1c.pdf

[16] Upper Northern Region Irrigation Hydrology Center (Royal Irrigation Department), Station history: Hydrological station of Yom River (Y.20), Ban Huai Sak, Tao Pun sub-district, Song district, Phrae province, (n.d.). Retrieved from https://www.hydro-1.net/Data/STATION/History/Y20.pdf

[17] Upper Northern Region Irrigation Hydrology Center (Royal Irrigation Department), Station history: Hydrological station of Yom River (Y.31), Ban Thung Nong, Sa sub-district, Chiang Muan district, Phayao province, (n.d.). Retrieved from https://www.hydro-1.net/Data/STATION/History/Y31.pdf

[18] Upper Northern Region Irrigation Hydrology Center (Royal Irrigation Department), Station history: Hydrological station of Yom River (Y.37), Ban Wang Chin, Wang Chin sub-district, Wang Chin district, Phrae province, (n.d.). Retrieved from https://www.hydro-1.net/Data/STATION/History/Y37.

[19] R. Nau, Statistical forecasting: notes on regression and time series analysis. Retrieved from https://people.duke.edu/~rnau/411home.htm, (2016).

[20] P. Rangkakulnuwat, Time series analysis for economics and business, Chulalongkorn University Press, Bangkok, 2013.

[21] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control. 19 (1974), 716–723.

[22] C. E. Holt, Forecasting seasonals and trends by exponentially weighted averages, Carnegie Institute of Technology, Pittsburgh USA. O.N.R. Memorandum No. 52 (1957).

[23] P.R. Winters, Forecasting sales by exponentially weighted moving averages, Manage. Sci. 6 (1960), 324–342.

[24] R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice (2018). Retrieved from https://otexts.com/fpp2/

[25] S. Arlot, A survey of cross validation procedures for model selection, Stat. Surveys. 4 (2010), 40-79.