A Hybrid Seasonal Box Jenkins-ANN Approach for Water Level Forecasting in Thailand

Kittipol Nualtong¹, Thammarat Panityakul^{1,*}, Piyawan Khwanmuang¹, Ronnason Chinram¹,
Sukrit Kirtsaeng²

¹Faculty of Science, Prince of Songkla University, Hat Yai, 90110, Songkhla, Thailand

²Thai Meteorological Department, Bangna, 10260, Bangkok, Thailand

Received March 20, 2021; Revised April 26, 2021; Accepted May 23, 2021

Cite This Paper in the following Citation Styles

(a): [1] Kittipol Nualtong, Thammarat Panityakul, Piyawan Khwanmuang, Ronnason Chinram, Sukrit Kirtsaeng, "A Hybrid Seasonal Box Jenkins-ANN Approach for Water Level Forecasting in Thailand," *Environment and Ecology Research*, Vol. 9, No. 3, pp. 93 - 106, 2021. DOI: 10.13189/eer.2021.090301.

(b): Kittipol Nualtong, Thammarat Panityakul, Piyawan Khwanmuang, Ronnason Chinram, Sukrit Kirtsaeng (2021). *A Hybrid Seasonal Box Jenkins-ANN Approach for Water Level Forecasting in Thailand. Environment and Ecology Research*, 9(3), 93 - 106. DOI: 10.13189/eer.2021.090301.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Every year, many basins in Thailand face the perennial droughts and floods that lead to the great impact on agricultural segments. In order to reduce the impact, water management would be applied to the critical basin, for instance, Yom River basin. An importing task of management is quantitative prediction of water that is stated by water level. This study proposes the hybridized forecasting models between the stochastic approaches, seasonal autoregressive integrated moving average (SARIMA) models and machine learning approach, artificial neural network (ANN). The proposed hybrid model is called seasonal autoregressive integrated moving average and artificial neural network or SARIMANN model for average monthly water level (AMWL) time series of Yom River basin. The study period is from April 2007 to March 2020, over thirteen hydrological years. The forecasting performance is the minimum values of root mean squared error (RMSE) and mean absolute percentage error (MAPE) between SARIMA models, ANN models, and SARIMANN models. Results indicated that: The three models reveal the similarity of RMSE and MAPE for both four water level measurement stations for wet and dry seasons. The forecasting performance is the minimum values of RMSE and MAPE of three models. The SARIMA model is the best approach for Y.31 Station [Wet Season], Y.20 Station [Wet Season], Y.37 Station [Wet Season], Y.31 Station [Dry Season], Y.20 Station [Dry Season], and Y.1C Station [Dry Season], while the best

method for Y.37 Station [Dry Season] is ANN model, furthermore the SARIMANN model is the best approach for Y.1C Station [Wet Season]. All methods have delivered the similar results in dry season, while both SARIMA and SARIMANN are better than ANN in wet season by RMSE for all stations. Even though the downstream is affected by many disturbances, it is still more accurate than the upstream. This is the visible evidence to indicate that the stochastic based models, SARIMA and SARIMANN proposed in this study are appropriate for the high fluctuation series. Furthermore, the dry season forecasting is more accurate than the wet season.

Keywords SARIMA, ANN, SARIMANN, Average Monthly Water Level, Yom River Basin

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 an area of 907.75 km² with fewer than 4 droughts per ten years, 701.3 km² with 4-5 droughts per ten years, and 229.34 km² with over 5 droughts per ten years, 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 an area of 2,454.06 km² with fewer than 4 floods per ten year, 1,334.25 km² with 4-7 floods per ten years, and 541.90 km² with over 8 floods per ten years, 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 be caused by climate changing and the human activities disturbance. These lead to complication of hydrological modeling [4, 5, 6, 7]. Sopipan, 2014 [8] 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. Fashae et al., 2018 [9] are to compare the artificial neural network (ANN) and autoregressive integrated moving average (ARIMA) model of River Opeki discharge in Oyo State, Nigeria 1982 to 2010 and to use the best model to forecast the discharge of the river from 2010 to 2020. The performance of the two models was subjected to be based on the root mean square error (RMSE) and correlation coefficient. The result showed that ARIMA performs better considering the level of RMSE and higher correlation coefficient. Pongsiri, 2550 [10] are to compare accuracy forecasting the regressive integrated moving average (ARIMA) model, artificial neural network (ANN) model, and hybrid regressive integrated moving average and artificial neural network (ARIMA-ANN) model of daily closing price of Trans-Pacific Partnership (TPP) time series since 2005 to 2007. The accuracy of the three models was subjected to be based on the mean square error (MSE) and mean absolute error (MAE) and mean absolute percentage error (MAPE). The ARIMA model concept is suitable for time series with a linear structure, while some models in machine learning approach e.g. the ANN model was able to completely describe the non-linear structure, as well as the three-based threshold model [22], or nonlinear principal component analysis [21] for stochastic approach. Therefore, the hybrid ARIMA-ANN model composite model would be completely descriptive of time series data with a linear and non-linear structure. The result showed that the hybrid ARIMA-ANN model was able to provide more accurate forecasts short-term (7-day) than the ARIMA and ANN

models, but for a long-term forecast (30-day) the ANN model provides the most accurate forecast.

This study proposes the hybridized forecasting models between the stochastic approaches, seasonal autoregressive integrated moving average (SARIMA) models and machine learning approach, artificial neural network (ANN). The proposed hybrid model is called seasonal autoregressive integrated moving average and artificial neural network or SARIMANN model for average monthly water level (AMWL) time series of Yom River basin. The study period is from April 2007 to March 2020, over thirteen hydrological years. The forecasting performance is the minimum values of root mean squared error (RMSE) and mean absolute percentage error (MAPE) between SARIMA models, ANN models, and SARIMANN model.

2. Materials and Methods

2.1. Study Region and Dataset

Yom River basin is located in Northern Thailand. It is a river basin covering a surface area of approximately 24,046.89 km², 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 [4, 11], as shown in Figure 1.

In 2014, the Yom River basin received average annual precipitation of about 1,179 mm, and average annual runoff of about 5,261 million mm³ and an average annual runoff of fewer than 2,500 mm³ per year per person, which is less than average annual runoff in Thailand per year per person (3,496 m³). 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 m³ [1, 4, 11, 12].

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 were 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 [13, 14, 15, 16, 17], as shown in Figure 1. The AMWL data at the four previously listed water level measurement stations for the wet and dry seasons of Thailand were calculated, the wet season is from May to October and the dry season is from November to April.

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

2.2. Methodology

2.2.1. 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 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 part, and the q -parameter refers to the moving average (MA) part [18, 19]. The back-shift operator (B), where $B^j y_t = y_{t-j}$ for j is a positive integer. The multiplicative $SARIMA(p, d, q) \times (P, D, Q)^S$ model is represented by:

$$\Phi_P(B^S)\phi_p(B)(1-B^S)^D(1-B)^d y_t = \Theta_Q(B^S)\theta_q(B)\varepsilon_t \quad (1)$$

where $\phi(B)$ is the regular Autoregressive polynomial of order p , i.e. $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$;

$\Phi(B^S)$ is the seasonal Autoregressive polynomial of order P , i.e. $\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}$;

$\theta(B)$ is the regular Moving Average polynomial of order q , i.e. $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$;

$\Theta(B^S)$ is the seasonal Moving Average polynomial of order Q , i.e. $\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}$;

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

$(1-B^S)^D$ is the D^{th} order seasonal differencing process y_t and ε_t is independent random variable that represents the error term at time t .

The multiplicative SARIMA models are a form of the Box-Jenkins methodology, It requires stationary time series data of y_t and the independent random variable (ε_t) must qualify as white noise

2.2.2. Artificial Neural Network Model (ANN Model)

An artificial neural network (ANN) is the foundation of artificial intelligence (AI), which the piece of a computing system is designed to simulate the human brain analyzes and processes information. ANNs have self-learning capabilities and two properties: learning and recall. The multi-layer perceptron neural networks or feed-forward backpropagation neural network is the learning of neural network, utilizes a supervised learning technique, learning that requires the input set and the target output as a training pair. Usually, network teaching uses multiple training sets, during network teaching will generate actual output, which is different from the target output causes an error value. The network learns both data by adjusting the weight to reduce the difference value or the error value between actual output and target output to a minimum. The weight adjustment will be adjusted in small increments by the process of repeating the data one by one until the weights in the network converge. Its greatest strength is in non-linear solutions to illdefined problems [20]. The multi-layer perceptron neural networks are shown in Figure 2.

Where y_n is the input data of input layer at node n^{th} ; $n = 1, 2, \dots, N$

x_m is the output data of hidden layer after weight adjustment using sigmoid-activation function at node m^{th} ; $m = 1, 2, \dots, M$

z_j is the actual output data of output layer after weight adjustment using linear -activation function at node j^{th} ; $j = 1, 2, \dots, J$

t_j is the target output data of output layer at node j^{th} ; $j = 1, 2, \dots, J$

e_j is the error of output layer at node j^{th} ; $j = 1, 2, \dots, J$

w_{nm} is the weight of network between input layer and hidden layer

w_{mj} is the weight of network between hidden layer and output layer

Figure 2. The multi-layer perceptron neural networks

Figure 3. The concept of input variables of SARIMANN models

2.2.3. Hybrid between Seasonal Autoregressive Integrated Moving Average and Artificial Neural Network Model (SARIMANN Model)

The SARIMA models are stochastic approach, which is suitable for linear variation with linear approach greatest seasonal and trends data but unable to manage non-linear data such as hydrological data. That opposes to ANN models which are a machine learning approach and suitable for non-linear situation. Therefore, we hybridize the seasonal autoregressive integrated moving average model and the artificial neural network model to deal with linear and non-linear situation simultaneously, which is called

SARIMANN models. The creation of the SARIMANN model uses the concept of using the input variables obtained from the best-fit SARIMA models and applying them to input data in the ANN model as shown in Figure 3.

Where the φ_p is the parameter of the autoregressive (AR) of order p , the Φ_P is the parameter of the seasonal autoregressive (SAR) of order P , the θ_q is the parameter of the moving average (MA) of order q , the Θ_Q is the parameter of the seasonal moving average (SMA) of order Q , the $(1 - B)^d$ is the differencing (a.k.a. Integrated, denoted by I) process y_t of order d , and the $(1 - B^S)^D$ is the seasonal differencing process (a.k.a. Seasonal

Integrated, denoted by SI) y_t of order D .

2.2.4. Accuracy and Performance

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

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|x_t - y_t|}{x_t} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - y_t)^2} \quad (3)$$

Where y_t is the forecast value of AMWL data of t since $t = 1, 2, \dots, n$; x_t is the observations of AMWL data of t and 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 processes K times along with the K subsamples and find the average performance of K-results [20].

3. Results

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 Figure 4 and Figure 5.

3.1. Hybrid between Seasonal Autoregressive Integrated Moving Average and Artificial Neural Network Model (SARIMANN Model)

Based on the twenty-four possible SARIMA models, there are eight selected best-fit SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons. Parameter estimates or coefficients of the SARIMA models are shown in Table 1.

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 1.

The input data and their weights of the ANN models of AMWL data of all four water level measurement stations for wet and dry seasons as shown in Table 2 and SARIMANN models as shown in Table 3. The weights of the network for the both models are shown in Figure 6, where the learning rate is 0.3 in 500 repetitions.

Figure 4. Time series plots of AMWL data of all four water level measurement stations for wet season.

Figure 5. Time series plots of AMWL data of all four water level measurement stations for dry season

Figure 6. The network weights of the ANN and SARIMANN models

Table 1. Parameter estimates of the best-fit SARIMA models of AMWL data of all four water level measurement stations for wet and dry seasons.

	Estimate (p-value)							
	Wet Season				Dry Season			
	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station
	SARIMA (1,1,1) (1,0,0) [6]	SARIMA (1,1,1) (1,0,0) [6]	SARIMA (1,1,2) (1,0,0) [6]	SARIMA (1,1,1) (1,0,0) [6]	SARIMA (0,1,1) (2,0,0) [6]	SARIMA (1,1,1) (1,0,1) [6]	SARIMA (1,1,1) (2,0,0) [6]	SARIMA (1,0,0) (1,0,1) [6]
AR (1)	0.6457 (3.400 $\times 10^{-12}$) ***	0.5937 (1.475 $\times 10^{-9}$) ***	-0.6800 (4.639 $\times 10^{-10}$) ***	0.5570 (3.984 $\times 10^{-10}$) ***		0.5021 (1.152 $\times 10^{-7}$) ***	0.6582 (4.331 $\times 10^{-12}$) ***	0.3408 (2.5 $\times 10^{-3}$) **
MA (1)	-1.0000 (<2.2 $\times 10^{-16}$) ***	-1.0000 (<2.2 $\times 10^{-16}$) ***	0.4789 (2.954 $\times 10^{-3}$) **	-1.0000 (<2.2 $\times 10^{-16}$) ***	-0.6862 (5.354 $\times 10^{-13}$) ***	-1.0000 (<2.2 $\times 10^{-16}$) ***	-1.0000 (<2.2 $\times 10^{-16}$) ***	
MA (2)			-0.5211 (7.471 $\times 10^{-4}$) ***					
SAR (1)	0.6459 (3.386 $\times 10^{-12}$) ***	0.5934 (1.643 $\times 10^{-9}$) ***	-0.6021 (1.155 $\times 10^{-9}$) ***	0.5411 (6.143 $\times 10^{-8}$) ***	0.4737 (2.168 $\times 10^{-5}$) ***	1.0000 (<2.2 $\times 10^{-16}$) ***	0.2887 (7.6 $\times 10^{-3}$) ***	0.9991 (<2.2 $\times 10^{-16}$) ***
SAR (2)					0.3626 (2.9 $\times 10^{-3}$) **		0.3514 (2.9 $\times 10^{-3}$) **	
SMA (1)						-0.9817 (<2.2 $\times 10^{-16}$) ***		-0.9519 (8.989 $\times 10^{-6}$) ***

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

Table 2. Weights of the ANN models of AMWL data of all four water level measurement stations for wet and dry seasons

	Weights							
	Wet Season				Dry Season			
	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station
Linear Node 0								
Threshold	0.1494	0.0934	-0.3972	-0.0964	0.0854	1.0284	0.1482	-0.1621
Node 1	1.0359	0.7389	2.9086	1.6453	2.4485	1.1895	4.1842	1.6887
Node 2	-0.7869	-0.7835	-0.6379	-0.5798	-0.8667	-1.9142	-1.7855	-1.2673
Node 3	0.7765	0.5542	1.3335	-0.3375	2.5791	-2.6691	1.5582	1.7122
Sigmoid Node 1								
Threshold	-1.5696	-1.3157	-3.2147	-3.0515	-4.6240	0.6590	-7.7099	-2.8729
AMWL _{t-1}	-0.5416	0.7506	0.1159	2.5133	3.2223	2.6723	-3.0610	-1.3120
AMWL _{t-2}	5.5071	1.2517	2.8811	0.1032	-0.4675	1.7150	-5.0924	-1.3865
AMWL _{t-3}	-0.2573	5.2971	1.1175	2.2680	0.8208	-0.5122	1.9657	-1.5255
AMWL _{t-4}	-1.4852	-1.2428	-4.6273	-5.7268	-0.1392	0.2092	-2.2274	-0.7883
AMWL _{t-5}	1.9705	0.8527	2.2212	1.6872	0.6219	1.3246	-1.2557	-0.2399
AMWL _{t-6}	4.1291	1.4816	1.6062	1.1200	2.7322	-1.3354	3.0677	-0.2952
Sigmoid Node 2								
Threshold	0.7619	0.3493	-2.2351	-2.0614	-3.4639	3.6275	-3.8208	-1.8865
AMWL _{t-1}	-6.4791	-6.8208	-4.3639	-4.6034	-2.6313	1.7624	-5.9928	-1.4595
AMWL _{t-2}	6.3179	5.3133	5.6453	3.2091	-1.1505	1.5220	0.0552	1.7130
AMWL _{t-3}	-1.5086	1.5733	1.0528	0.7059	0.2206	0.4403	1.6805	-0.0531
AMWL _{t-4}	-0.2468	0.1203	-1.0660	-0.5833	-1.6028	3.3420	-0.5489	1.3194
AMWL _{t-5}	-2.5332	-2.2248	-4.1954	-3.5902	-0.9167	-2.7299	-0.4376	0.1540
AMWL _{t-6}	3.2588	1.6231	-0.3913	0.8495	-3.4295	-6.9409	1.0961	0.1618
Sigmoid Node 3								
Threshold	-6.0669	-2.6830	-2.0891	-2.1890	-6.6147	-2.4214	-2.8988	-1.9035
AMWL _{t-1}	2.0517	1.8190	1.5501	0.9403	-2.1058	-1.3992	1.6166	1.8967
AMWL _{t-2}	-1.9365	3.2571	-1.0935	0.3358	-3.9984	2.7642	-1.3858	0.9088
AMWL _{t-3}	-7.5265	-2.5026	0.7810	0.5019	0.3163	-1.6852	1.5819	1.1689
AMWL _{t-4}	1.0565	-3.0205	-0.5790	-0.5462	-4.4778	1.7199	0.0639	1.4016
AMWL _{t-5}	-3.4877	2.8567	-0.4332	0.5286	0.5810	1.0267	0.5617	0.8979
AMWL _{t-6}	1.3809	2.1114	-0.9865	0.4235	2.0798	2.4989	2.3554	0.9549

Table 3. Weights of the SARIMANN models of AMWL data of all four water level measurement stations for wet and dry seasons

	Weights							
	Wet Season				Dry Season			
	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station
Linear Node 0								
Threshold	0.0393	0.0003	0.1540	0.0933	0.7204	1.1860	-0.3559	-0.1626
Node 1	1.7603	1.1384	-1.3507	0.1886	-1.2869	-2.2986	1.1903	0.5037
Node 2	-0.9396	-1.0489	-1.2822	-1.1374	-0.7399	-0.5553	2.6910	-1.1827
Node 3	2.2005	1.7861	-0.8550	-0.1687	-1.2089	-1.7066	1.7950	3.0640
Sigmoid Node 1								
Threshold	-1.8394	-2.5142	-2.4211	-2.8144	-2.6892	-2.4460	-1.6151	-1.5586
0/AMWL _{t-1}								2.9992
0/AMWL _{t-6}	1.9283	1.4968	0.0664	0.5790	-1.4051	3.0081	4.3745	5.0042
0/AMWL _{t-12}					1.2434		-2.9507	
1/AMWL _{t-1}	0.4532	0.4805	-1.1228	-0.1449		2.0422	2.9640	
0/RES _{t-6}						-1.0014		2.1169
1/RES _{t-1}	-0.7487	0.1120	1.4661	0.2092	1.2531	-1.9468	-1.6293	
1/RES _{t-2}			0.9633					
Sigmoid Node 2								
Threshold	1.8728	0.0047	-2.4521	-0.7574	-2.8149	-2.6016	-4.9368	-3.1347
0/AMWL _{t-1}								-2.1851
0/AMWL _{t-6}	2.8957	0.3468	0.0124	-1.5131	-0.7311	0.1576	-1.2989	-1.4632
0/AMWL _{t-12}					0.5743		4.9074	
1/AMWL _{t-1}	-10.2971	-7.2319	-0.9860	-2.2438		0.6384	0.1385	
0/RES _{t-6}						-0.3432		-0.4407
1/RES _{t-1}	-0.8767	1.7358	1.3275	0.7987	0.7811	0.4208	0.2847	
1/RES _{t-2}			0.8943					
Sigmoid Node 3								
Threshold	-2.9161	-2.5650	-1.4840	-2.8647	-0.5016	-0.7649	-1.7879	-4.3288
0/AMWL _{t-1}								-2.0127
0/AMWL _{t-6}	1.3745	1.8639	-2.8224	0.3167	-2.0426	-5.3370	0.4385	3.0204
0/AMWL _{t-12}					-1.6690		0.4474	
1/AMWL _{t-1}	-0.1713	-2.4052	-3.3351	-0.4573		-1.2588	1.1706	
0/RES _{t-6}						0.8794		-4.6688
1/RES _{t-1}	-1.1639	-0.5578	-1.0053	0.2717	-0.9486	0.7940	-0.9258	
1/RES _{t-2}			-1.6456					

3.2. Comparison of Forecasting Performance

Comparisons of the three models reveal the similarity of RMSE and MAPE for both wet season forecasting (six months: May 2019 – October 2019) and dry season forecasting (six months: November 2019 – April 2020). The forecasting performances are the minimum values of

RMSE and MAPE of three models. The SARIMA model is the best approach for Y.31 Station [Wet Season], Y.20 Station [Wet Season], Y.37 Station [Wet Season], Y.31 Station [Dry Season], Y.20 Station [Dry Season], and Y.1C Station [Dry Season] (the RMSE value of 0.9447, 0.12475, 0.8961, 0.1480, 0.2162, and 0.2690; the MAPE value of 0.3170, 0.5637, 0.8244, 0.0556, 0.1126, and 0.1830), while

the best method for Y.37 Station [Dry Season] (the RMSE value of 0.2290; the MAPE value of 0.1823) is ANN model, furthermore the SARIMANN model is the best approach for Y.1C Station [Wet Season] (the RMSE value of 1.3181;

the MAPE value of 0.6021). As shown in Table 4 and Figure 7. Forecasting of the three methods of AMWL data of all four water level measurement stations for wet season are shown in Figure 8 and dry season as shown in Figure 9.

Table 4. Comparison of forecasting performance and accuracy of AMWL data of all four water level measurement stations for wet and dry seasons.

	Accuracy							
	Wet Season				Dry Season			
	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station	Y.31 Station	Y.20 Station	Y.1C Station	Y.37 Station
RMSE								
SARIMA	0.9447	1.2475	1.3236	0.8961	0.1480	0.2162	0.2690	0.3093
ANN	1.6177	1.5114	1.6293	1.5814	0.3108	0.3076	0.6789	0.2290
SARIMANN	1.1767	1.3575	1.3181	1.0208	0.2212	0.3585	0.4480	0.2851
MAPE								
SARIMA	0.3170	0.5637	0.6797	0.8244	0.0556	0.1126	0.1830	0.2809
ANN	0.5543	0.7270	1.0390	1.4291	0.1049	0.1198	0.4278	0.1823
SARIMANN	0.3721	0.5988	0.6021	0.8953	0.0835	0.1769	0.3067	0.2746

Figure 7. Comparisons of forecasting accuracy AMWL values of all four water level measurement stations for wet and dry seasons

Figure 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 wet season

Figure 9. 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

4. Conclusion and Discussion

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 been analyzed. The following conclusions and recommendations can be obtained from this study:

Comparisons of the three models reveal the similarity of RMSE and MAPE for both four water level measurement stations for wet and dry seasons.

Comparisons of the three models reveal the similarity of RMSE and MAPE for both wet season forecasting (six months: May 2019 – October 2020) and dry season forecasting (six months: November 2019 – April 2020). The forecasting performance is the minimum values of RMSE and MAPE of three models. The SARIMA model is the best approach for Y.31 Station [Wet Season], Y.20 Station [Wet Season], Y.37 Station [Wet Season], Y.31 Station [Dry Season], Y.20 Station [Dry Season], and Y.1C Station [Dry Season], while the best method for Y.37 Station [Dry Season] is ANN model, furthermore the SARIMANN model is the best approach for Y.31 Station [Wet Season].

In the dry season, all methods are similarly performed, while in the wet season SARIMA and SARIMANN explicitly overcome ANN method by RMSE for all the upstream through downstream stations. The downstream station (Y.37 station) has more accuracy than the upstream station (Y.31 station) despite the downstream affected by human activities (reservoirs, water use for agriculture and consumption) that disturb hydrological changes that cause data uncertainty and hard to handle in forecasting. This is the clarified evidence to indicate that the stochastic based models, SARIMA and SARIMANN which proposed in this study are suitable for handle with high fluctuation series

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.

The possible further study in hybridized-manner for handling the extreme values which often nonstationary may be Tree-based threshold model [22].

However, this proposed method, SARIMANN is suitable for wet season as well as SARIMA and better than ANN definitely for both upstream and downstream especially in downstream. So, the implication of this proposed model is for the highly disturbance situation of any stream.

But in dry season the proposed method has quite similar performance as SARIMA and ANN for both upstream and downstream, this might be the limitation of the proposed model. So, the prospective researcher should try other approaches for instance frequency domain time series

analysis or fuzzy time series analysis for the future study.

REFERENCES

- [1] Office of the National Water Resources, The National Water Resources Management Strategies (2015-2026), Online available from http://www.onwr.go.th/?page_id=3684
- [2] Department of Water Resources, Water Situation: Drought area, Online available from <http://data.dwr.go.th/dwr/watersituation/drought/index>
- [3] Department of Water Resources, Water Situation: Flood Area, Online available from <http://data.dwr.go.th/dwr/watersituation/deluge/index>
- [4] P. Khwanmuang, R. Chinram, T. Panityakul. The Hydrological and Water Level Data in Yom River Basin of Thailand, *Journal of Mathematical and Computational Science*. Vol.10, No.6, 3026-3047, 2020.
- [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*, Vol.377, No.3-4, 274-283, 2009.
- [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*. Vol.529, No.3, 711-722, 2015.
- [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*. Vol.530, 462-475, 2015.
- [8] N. Sopipan. Forecasting Rainfall in Thailand: A Case Study of Nakhon Ratchasima Province, *International Journal of Environmental, Ecological, Geological and Marine Engineering*, Vol.8, No.11, 712-716, 2014.
- [9] O. A. Fashae, A. O. Olusola, I. Ndubuisi, C. G. Udomboso. Comparing ANN and ARIMA model in predicting the discharge of River Opeki from 2010 to 2020, *River Research Application*, Vol.35, No.2, 169-177, 2018.
- [10] P. Siripanich. Time series forecasting using a combined ARIMA and artificial neural network model, Online available from <http://www.sure.su.ac.th/xmlui/bitstream/handle/123456789/11717/Fulltext.pdf?sequence=1&isAllowed=y>
- [11] 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), Online available from <http://www.thaiwater.net/web/attachments/25basins/08-yom.pdf>
- [12] Bureau of Water Management and Hydrology (Royal Irrigation Department), Water allocation and cultivation plan for irrigation area during the dry season in 2019 - 2020, Online available from <http://water.rid.go.th/hwm/wmoc/planning/dry/manage%20water2562-63.pdf>

- [13] Upper Northern Region Irrigation Hydrology Center (Royal Irrigation Department), Average daily of water level - water content data, Online available from <http://hydro-1.rid.go.th/>
- [14] 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, Online available from <https://www.hydro-1.net/Data/STATION/History/Y1c.pdf>
- [15] 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, Online available from <https://www.hydro-1.net/Data/STATION/History/Y20.pdf>
- [16] 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, Online available from <https://www.hydro-1.net/Data/STATION/History/Y31.pdf>
- [17] 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, Online available from <https://www.hydro-1.net/Data/STATION/History/Y37.pdf>
- [18] H. Fouli, R. Fouli, B. Bashir, O. A. Loni. Seasonal Forecasting of Rainfall and Runoff Volumes in Riyadh Region, KSA, KSCE Journal of Civil Engineering, Vol.22, No.7, 2637-2647, 2018.
- [19] R. Nau, Statistical forecasting: notes on regression and time series analysis. Online available from <https://people.duke.edu/~rnau/411home.htm>
- [20] P. N. TAN, M. Steinbach, V. Kumar. Introduction to data mining, Pearson Education Limited, U.S.A., 2014.
- [21] Solimun, Fernandes, A.A.R., Cahyoningtyas, R.A., The implementation of nonlinear principal component analysis to acquire the demography of latent variable data (A study case on brawijaya university students), Mathematics and Statistics 8(4), pp. 437-442, 2020. DOI: 10.13189/ms.2020.080410.
- [22] Shihabuddin, A., Ali, N., Adam, M.B., Tree-based threshold model for non-stationary extremes with application to the air pollution index data, Mathematics and Statistics, 7(4A), pp. 58-64, 2019. DOI: 10.13189/ms.2019.070708.