Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 4, 3984-3998
https://doi.org/10.28919/jmcs/5818
ISSN: 1927-5307

PREDICTIVE ANALYTICS MODIFICATIONS IN WAVELET: CASE STUDY ON SONGKHLA LAKE BASIN RUNOFF

SOIKANOK SRISAWAT, KITIPOL NUALTONG, THAMMARAT PANITYAKUL,
KORAKOT WICHITSA-NGUAN JETWANNA*
Division of Computational Science, Faculty of Science,
Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

Abstract: Accurate short-term rainfall-runoff prediction is very important for flood mitigation and the safety of infrastructures in Southern Thailand. This study aims to utilize both analysis and prediction of the runoff forecast by combining the wavelet technique with regression and artificial neural network.

The daily rainfall and runoff data were collected from 1,031 days during January 2017 to October 2019 in Songkhla Lake Basin, Thailand. The performance of calibration and validation of the models is evaluated with appropriate statistical methods; coefficient of determination (R^2), root mean square error (RMSE) and Nash-Sutcliffe efficiency coefficient (ENS). The results of daily runoff series modeling indicated that the wavelet artificial neural network model performed the best among those models. This model showed the Coefficient of Determination, Nash-Sutcliffe Efficiency and Root mean Square Error in the value of 0.9999, 0.9998 and 0.0037, respectively. These values explained that the model can describe 99.99% of the variation of the current runoff in Songkhla Lake Basin.

Keywords: Rainfall; runoff; wavelet regression; wavelet artificial neural network.

2010 AMS Subject Classification: 62P12.

*Corresponding author
E-mail address: korakot.w@psu.ac.th
Received April 5, 2021

1. INTRODUCTION

Over the past 40 years, Thailand has been hit by many weather extremes. Floods are one of the most serious natural issues and present major social concern in Thailand that causes massive damage to lives and properties. Back in 2011, there was a great flood disaster that covered more than one-third of the provinces in Thailand that reduced 14% of Gross Domestic Product [11].

Runoff forecasting plays an important role in water management and flood prediction. Recently, there are a lot of papers explained that runoff is affected by many factors such as rainfall, land use, soil infiltration rates and others. The relationship between those factors and runoff is complicated and it is a non-linear relationship. Despite a non-linear relationship, none of a statistical model can describe the complexity of the relationship. There have been many attempts to find the most effective model for runoff forecasting.

Artificial neural network (ANN) models have been wildly used in the studies about the rainfall-runoff model [1, 2, 4, 12-16]. The neural network models are machine learning models based on studies of the brain and nervous system. Moreover, ANN has a flexible mathematical structure and has an advantage in various fields of science which is due to its can model both linear and non-linear without considering any assumptions as in the statistical models [3, 18].

In the past decade, wavelet transform has been successfully used with a highly non-linear model. Each study showed that the combination of wavelet transform and other models provided higher accuracy than normal models [1, 2, 4, 5, 9, 10, 14, 16]. A combination of a wavelet transform, artificial neural network and regression models are presented at a rainfall-runoff model. According to the Coefficient of Determination (R^2), Root Mean Square Error (RMSE) and Nash-Sutcliffe Efficiency (E_{NS}), it can be concluded that the artificial neural network model with wavelet transform is more efficient than the regular artificial neural network and regression model.

In this paper, purposing to compare the performance of regular regression and artificial neural network model with wavelet regression and artificial neural network for forecasting runoff in Songkhla Lake Basin, Thailand.

2. STUDY REGION AND DATASET

2.1 Study Area: Songkhla Lake Basin is located within 6°45′&8°00′ North latitude 99°30′&100°45′ East longitude. The approximate total area is 8,484.35 km^2 separated into the land area and lake area of 7,652.81 and 831.54 km^2 respectively. Songkhla Lake Basin has been influenced by the northeast monsoon that blows through the South China Sea and the Gulf of Thailand. As a result, the Songkhla Lake Basin receives water vapor, resulting in heavy rainfall from October to December. The average annual rainfall and runoff of Songkhla Lake Basin are 1,992 mm and 4,808 million m^3, respectively. If rain occurs consecutively with rainfall of 90.1 mm. or more within 24 hours, it can cause a flash flood. From the interpretation of THEOS and LANDSAT-5 TM satellite data, it was found that most of the flooding areas were lowland areas especially in the plain area around the basin. The map of Songkhla Lake Basin is shown in **Figure 1.**

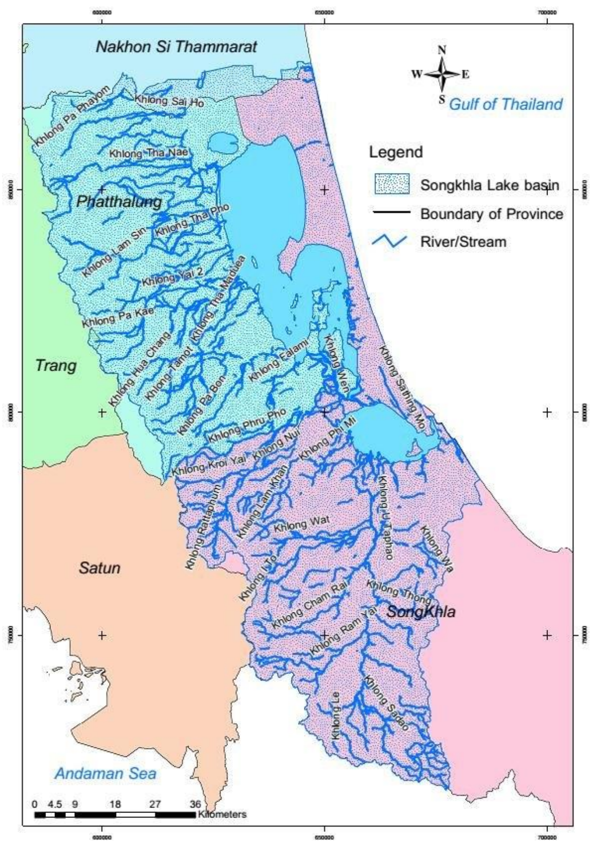


Figure 1. Map of Songkhla Lake Basin. Source: Cookey, et al. 2015 [7].

2.2 Data Collection: In this study, the daily rainfall and runoff time series had 1031 days from January 4th, 2017 to October 31st, 2019 in Songkhla Lake Basin, Songkhla Thailand that was obtained from Climate Information Service, Meteorological Department of Thailand [6] and Southern region irrigation hydrology center, Bureau of Water Management and Hydrology [17], respectively. The first 721 days (70% of data) were used for calibrating the model and the remaining 310 days (30% of data) were used for validation.

3. METHODOLOGY

3.1 Wavelet analysis

Wavelets are functions that use to decompose function. The wavelet transform decomposes a function into a family of wavelets, creates a representation of the function in both the time and frequency domain, thereby allowing efficient access of localized information about the function, so it is a mathematical tool useful for data analysis, its transform function with time domain to a frequency domain. The family of wavelets contains the dilated and translated versions of a mother function, which is called a mother wavelet. The scale and translation of wavelets determine how the mother wavelet dilates and translates along the time or space axis. A scale factor greater than one corresponds to a dilation of the mother wavelet along the horizontal axis, and a positive shift corresponds to a translation to the right of the scaled wavelet along the horizontal axis. There are two types of wavelet transforms: Continuous wavelet transforms and discrete wavelet transforms. Continuous wavelet transform (CWT) uses every possible wavelet over a range of scales and locations i.e. an infinite number of scales and locations. While the discrete wavelet transform (DWT) was developed by Ingrid Daubechies, uses a countable set of wavelets that is defined at a particular set of scales and locations. The method of discrete wavelet transform can be described by the multiresolution analysis as the following.

Let $\psi(x)$ be a mother wavelet function. For integer numbers, j and k, define $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ the operations of translation and dilation of the mother wavelet [8].

Where $\{\psi_{jk}(x),\ \textit{for all } j \textit{ and } k\}$ is an orthonormal basis of $L^2(\mathbb{R})$. For fixed j, define

$$V_j = \left\{ \sum_{k=-\infty}^{\infty} a_k \psi_{jk}(x) \ : \ \sum_{k=-\infty}^{\infty} |a_k|^2 < \infty \right\}.$$

A function $f \in L^2(\mathbb{R})$ can be expressed by

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \alpha_{jk} \psi_{jk}(x).$$

The α_{jk} are the wavelet transform coefficients of $f(x)$.

Every orthonormal mother has an auxiliary wavelet function $\phi \in L^2(\mathbb{R})$ called scaling wavelet function and $\{\phi_k(x) = \phi(x-k) : k \in \mathbb{Z}\}$ is an orthonormal set.

Define set V_0 by

$$V_0 = span\{\phi_k : k \in \mathbb{Z}\}.$$

Since V_0 is a subspace of $L^2(\mathbb{R})$, every function $f \in L^2(\mathbb{R})$ can be projected into the subspace V_0 so that the development of this projection in term of the moving scale function is an approximation of $f(x)$ such that

$$P_0 f(x) = f_0(x) = \sum_{k=-\infty}^{\infty} a_k \phi_k(x) \text{ where } a_k = \langle f, \phi_k \rangle.$$

In general, for each integer number j, we have the closed subspaces V_j of $L^2(\mathbb{R})$;

$$V_j = span\{\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k : \ j, k \in \mathbb{Z}\}.$$

Therefore, every function $f \in L^2(\mathbb{R})$ can be projected into the subspace V_j so that the development of this projection in term of the moving scale function is an approximation of $f(x)$ such that

$$P_j f(x) = f_j(x) = \sum_{k=-\infty}^{\infty} a_k \phi_{jk}(x) \text{ where } a_k = \langle f, \phi_{jk} \rangle.$$

Define W_j the orthogonal complement of V_j in V_{j+1}. Hence $f \in L^2(\mathbb{R})$ can be written in term of an approximation subspace V_j in the j resolution which is formulated on scaling function or low pass filter and the orthogonal terms containing the finer details associated with W_j which is formulated on mother wavelet $\psi(x)$ or high pass filter. As in **Figure 2**, the multiresolution analysis builds a structure that requires an iterative application of scaling and mother wavelet, the function is split between the low-frequency approximation and the high-frequency details, respectively.

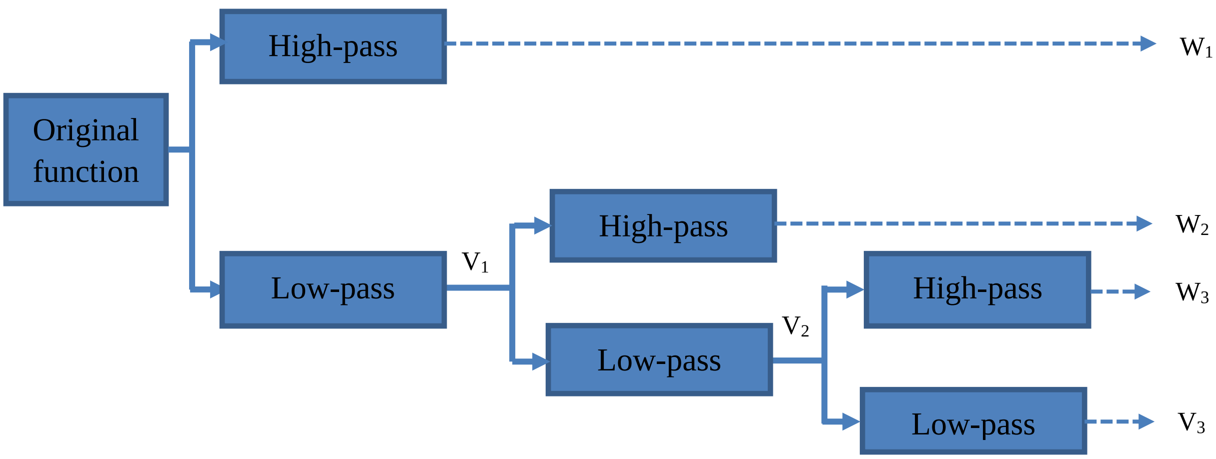


Figure 2. The function is split between the low-frequency approximation and the high-frequency details.

The Daubechies wavelets, based on the work of Ingrid Daubechies, is constructed from a class of wavelet function ψ_N, $N \in \mathbb{N} \setminus \{0\}$, that satisfy the following properties. 1) The collection $\psi_N(x-k), k \in \mathbb{Z}$ is an orthonormal system for fixed $N \in \mathbb{N} \setminus \{0\}$. 2) For each $N \in \mathbb{N} \setminus \{0\}$, ψ_N is compactly supported function and vanishing moments with number N. 3) For each $N \in \mathbb{N} \setminus \{0\}$, $\text{supp}(\psi_N) = [0, 2\text{N}-1]$. 4) $\exists \lambda > 0, \forall N \in \mathbb{N}, N > 2 : \psi_N \in C^{\lambda N}$. For large N one has $\lambda \approx 0.2$. In general, the Daubechies wavelets are chosen to have the highest number N of vanishing moments, in this case, dbN is the name type of Daubechies wavelet transform.

3.2 The Artificial Neural Network

The Artificial Neural Network is a technique of Artificial Intelligence that simulates the functioning of the nervous system in the human brain. The structure of the neural network is shown in the figure below.

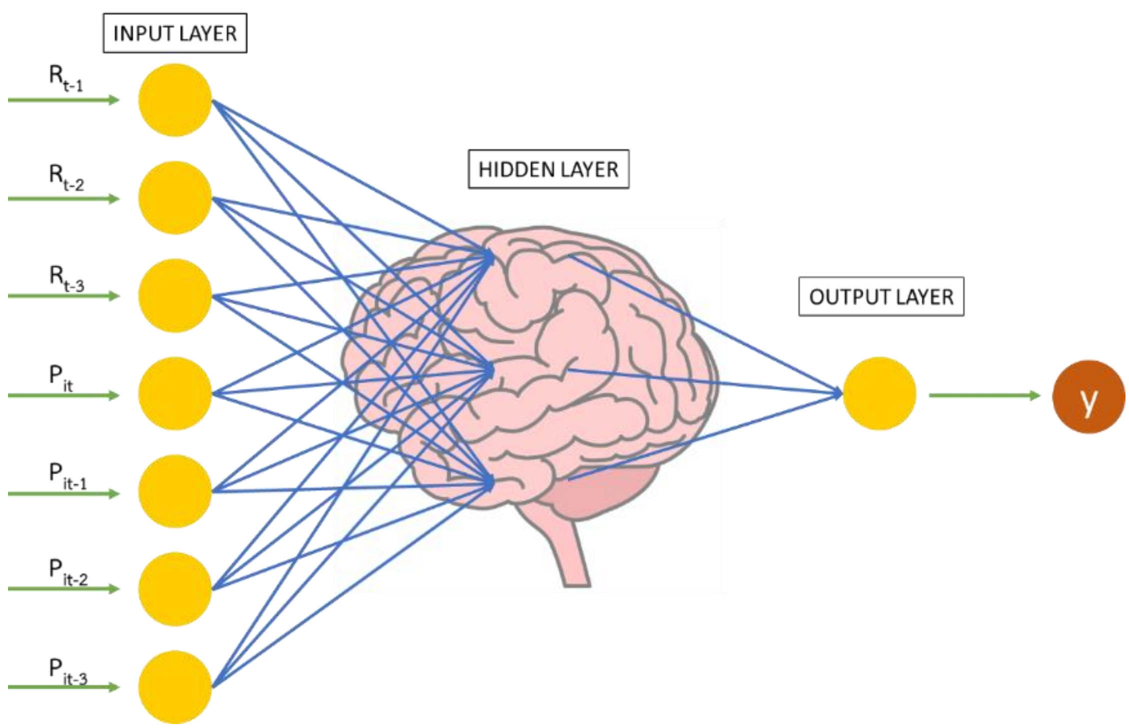


Figure 3. The structure of the neural network.

The structure of the neural network consists of 4 components as follows: 1) Input 2) Output 3) Hidden Layer and 4) Links (**Figure 3**). The artificial neural network is a model that simulates the function of the human nervous system. There will be several working units to process together. Each unit is submitted for use in calculating the results. The artificial neural network model is a self-modifying model during the learning process. It is used in modeling to describe the complex relationships between independent and dependent variables.

This study applies a single-layer neural network to this dataset, which has only one hidden layer that receives inputs from the previous layers. The outputs of the input layer are inputs to the hidden layer. Where the inputs to each node are weighted linear combination and then modified by nonlinear function before being output. For example, the inputs into the j^{th} -hidden node are linear combination such that

$$y_j = c_j + \sum_{i=1}^{7} w_{i,j} x_i$$

Then nonlinear function being the modification of this hidden layer, the useful nonlinear function is sigmoid,

$$\theta(y) = \frac{1}{1+e^{-y}}$$

We are learning from the data that gives the parameters c_1, c_2, c_3 and $w_{1,1}, \ldots, w_{7,3}$ the values of the weights. These are often set to be equal to 0.1. These weights are random values for initiation, then updated by the observed data. Generally, we determine the number of hidden layers and the number of nodes in each layer, but using cross-validation leads to the appropriate of both numbers.

3.3 Model Performance

The performance of models during calibration and validation were evaluated by using the statistical indices: Coefficient of Determination (R^2), Root Mean Square Error (RMSE) and Nash-Sutcliffe Efficiency (E_{NS}). The formula of each statistical index is presented below:

Coefficient of Determination (R^2):

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)\left(\hat{Y}_i - \bar{\hat{Y}}\right)\right]^2}{\left[\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right]\left[\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{\hat{Y}}\right)^2\right]}$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

Nash-Sutcliffe Efficiency (E_{NS}):

$$E_{NS} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

where Y_i and $\hat{Y}_i$ represent observed and predicted values at the time i, respectively,

$\bar{Y}$ and $\bar{\hat{Y}}$ represent the mean of observed and predicted values.

4. RESULTS AND DISCUSSION

In this study, input variables were rainfall of current day, past 1 day, past 2 days, past 3 days, and runoff of previous 1 day, 2 days and 3 days expressed as Pt, Pt-1, Pt-2, Pt-3, Qt-1, Qt-2, Qt-3, respectively. These variables were used to examine the regular regression model, regular artificial neural network model, wavelet regression and wavelet artificial neural network models. The structure of the regular and wavelet artificial neural network consisted of 12 nodes on the hidden layer with 5,000 epochs.

The original time series of daily runoff was decomposed into Details and Approximations by discrete wavelet transform algorithms. According to the trend of the series, Daubechies3 at level 5 was used as the mother wavelet. Consequently, D_1, D_2, D_3, D_4 and D_5 were detail components, and A_5 was approximation components. The decomposed components of details and approximation together with the original data of daily runoff were shown in **Figure 4-5**.

As can be seen from **Table 1**, these statistical indices showed that wavelet transform can increase the performance of a forecasting model.

Wavelet artificial neural network model performed the best among those models. This model showed Coefficient of Determination, Nash-Sutcliffe Efficiency and Root mean Square Error in the value of 0.9999, 0.9998 and 0.0037, respectively. These values explained that the model can describe 99.99% of the variation of the current runoff in Songkhla Lake Basin.

Table 1. Statistical accuracy measures of models.

	Regression Model		Artificial Neural Network Model	
	Without wavelet	With wavelet	Without wavelet	With wavelet
R^2	0.9071	0.9999	0.9500	0.9999
E_{NS}	0.8185	0.9998	0.8798	0.9998
RMSE	0.1335	0.0045	0.1086	0.0037

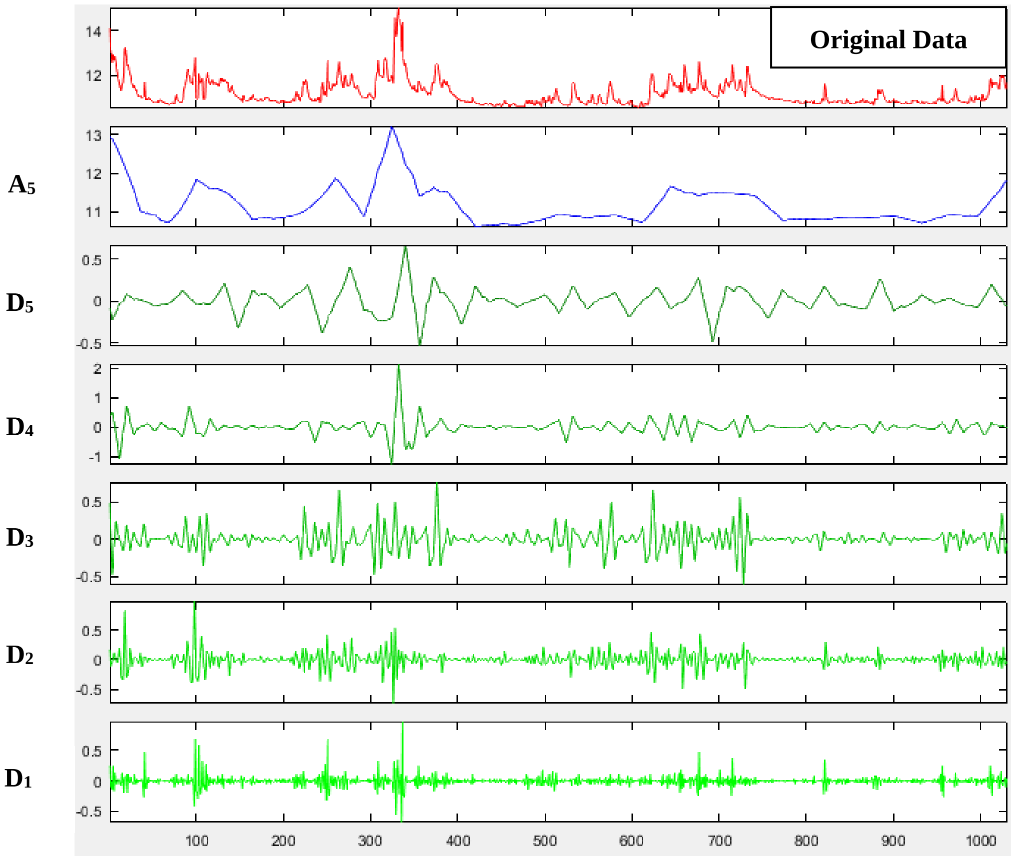


Figure 4. The decomposed components of details and approximation by discrete wavelet transform together with the original data of daily runoff.

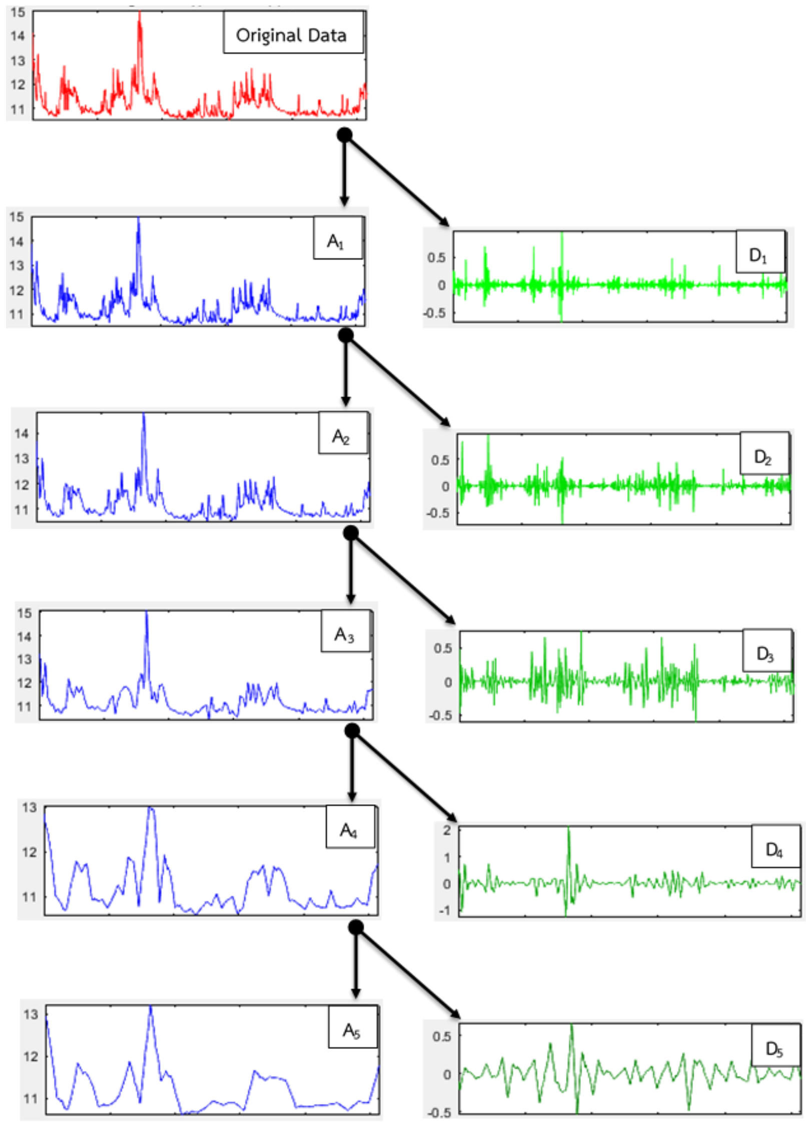


Figure 5. The order decomposed components of details and approximation by discrete wavelet transform together with the original data of daily runoff.

In the case of the wavelet regression model, there are assumptions that need to concern about. From **Figure 6**, the normal Q-Q plot and Shapiro-Wilk Normality Test can be explained that the residual of the model are not normally distributed (p-value = 0.0000) and from the scatter plot between fitted and residual value together with the value of Bartlett's Test of Homogeneity of Variances showed that the residuals have constant variance (p-value = 0.944). From these two assumptions might be explained those high statistical indices of the wavelet regression model.

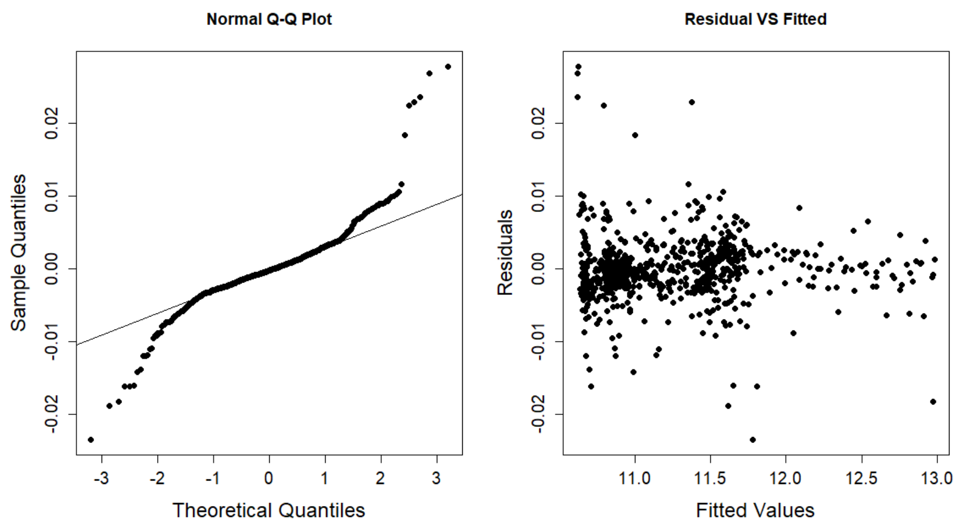
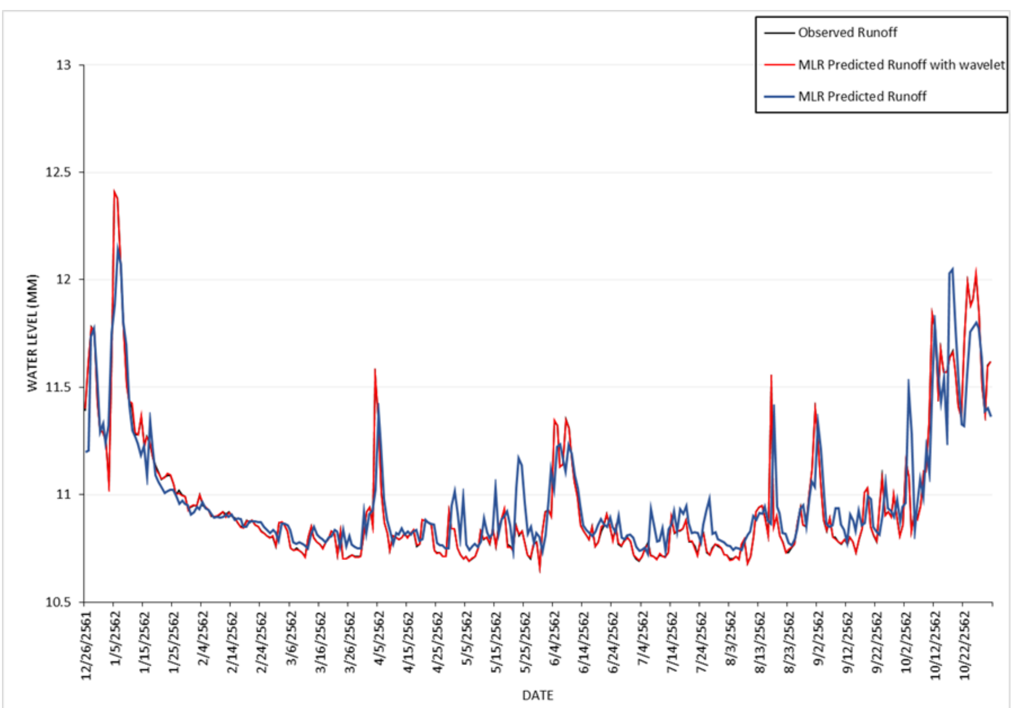
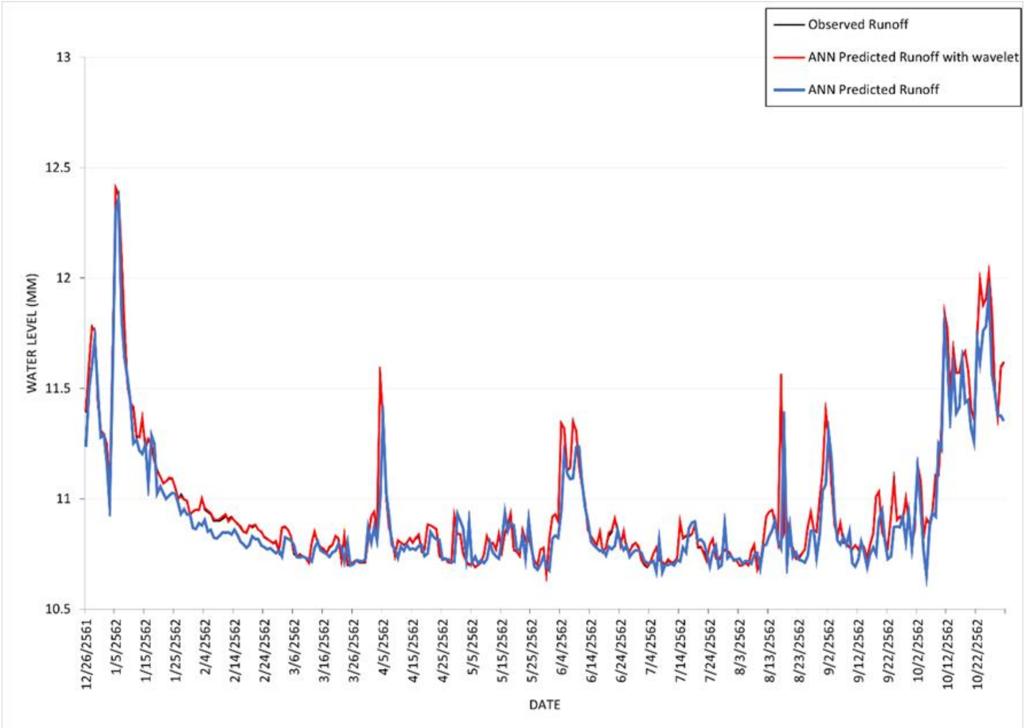


Figure 6. Normal Q-Q plot and scatter plot of the residuals of the wavelet regression model.

Figure 7 shows the comparison between the predicted values and observed values of daily runoff of regression models (a) and artificial neural network models (b). The line of predicted values of the daily runoff in Songkhla Lake Basin (Red) from the models that combined with wavelet transform is placed nearly close to the observed values of the daily runoff. Moreover, the predicted values from regular regression and artificial neural network models (Blue) are placed in a little longer distance from the observed values of the daily runoff in Songkhla Lake Basin.



(a)



(b)

Figure 7. Comparison of runoff in validation period predicted by regression models (a) and artificial neural network models (b).

5. CONCLUSIONS

In this study, both of the regression model and the artificial neural network model were found as a model that can reveal the complex relationship between runoff and rainfall. When wavelet transform was used, it was able to significantly improve the performance of those models. Therefore, in terms of adoption of both models in a timely manner provide an alert or warning in the event of flash flooding to the residents of Songkhla Lake Basin.

This study used only two influenced factors, the researcher recommended to investigate other influenced factors to runoff such as water flow rate, soil water absorption rate, evapotranspiration, land-use, etc. to optimize forecasting for long terms.

From this study, regarding the efficiency of the daily runoff forecast model in the Songkhla Lake Basin, noted that if the predictive models obtained from this study are used in organizations dealing with water resource management and flooding. It was found that the both of wavelet regression model and wavelet neural network model providing a very high performance. The practical use should be within the scope of consideration in a conjunction with hydrological problem management experts.

ACKNOWLEDGMENT

Thanks to the Climate Information Services for daily rainfall data and the Water crisis prevention center for daily runoff data in Songkhla Basin, Thailand, to Mr.Sukrit Kirtsaeng for suggesting the study area and to Ms.Ratchaya Weerakarn for editing this article. And thanks to the established program of Smart Analytics for SME and Community (SASMEC).

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] A. Solgi, V. Nourani, A. Pourhaghi, Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system (Case Study: Verayneh Station, Nahavand), Adv. Civil Eng. 2014 (2014), 279368.

[2] A.H. Nury, K. Hasan, Md.J.B. Alam, Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh, J. King Saud Univ.-Sci. 29 (2017), 47–61.

[3] A. Kurenkov, A brief history of neural nets and deep learning, Retrieved from https://www.skynettoday.com/overviews/neural-net-history. (2015).

[4] C.A.G. Santos, P.K.M.M. Freire, G.B.L. Silva, R.M. Silva, Discrete wavelet transform coupled with ANN for daily discharge forecasting into Três Marias reservoir, Proc. Int. Assoc. Hydrol. Sci. 364 (2014), 100–105.

[5] C. Chou, Enhanced accuracy of rainfall–runoff modeling with wavelet transform, J. Hydroinformatics. 15 (2013), 392–404.

[6] Climate information services, Daily rainfall data in Songkhla Basin. (2017), Retrieved from https://www.tmd.go.th/cis/main.php

[7] P.E. Cookey, R. Darnswadi, C. Ratanachai, Governing Songkhla Lake Basin, Thailand: the requirement for the adoption of adaptive integrated lake basin management. In: Conference Proceedings of the 1st National and 2nd International Conference on Ecotourism and Social Development for ASEAN Community (ESDA 2015), 22-23 January 2015, Muang District, Surathani Province, Thailand.

[8] I. Daubechies, Ten lectures on wavelets, SIAM, Philadelphia, 1992.

[9] J. Adamowski, H.F. Chan, S.O. Prasher, V. NathSharda, Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data, J. Hydroinformatics, 14 (2012), 731- 744.

[10] J. Lu, A. Li, Monthly Runoff Prediction Using Wavelet Transform and Fast Resource Optimization Network (Fron) Algorithm, J. Phys.: Conf. Ser. 1302 (2019) 042005.

[11] 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).

[12] P.M. Chavan, S.K. Ukarande, Comparison of rainfall-runoff model using ANN and MLR. Int. J. Recent Adv. Eng. Technol. 5(2017), 2347-2812.

[13] P.S. Kumar, T.V. Praveen, M.A. Prasad, Artificial neural network model for rainfall-runoff-a case study, Int. J. Hybrid Inform. Technol. 9 (2016), 263-272.

[14] P.K. de M.M. Freire, C.A.G. Santos, G.B.L. da Silva, Analysis of the use of discrete wavelet transforms coupled with ANN for short-term streamflow forecasting, Appl. Soft Comput. 80 (2019), 494–505.

[15] P. Kumari, P. Kumar, P.V. Singh, Rainfall-runoff modelling using artificial neural network and adaptive neural-fuzzy inference system. Indian J. Ecol. 45 (2018), 281-285.

[16] R. Venkata Ramana, B. Krishna, S.R. Kumar, N.G. Pandey, Monthly rainfall prediction using wavelet neural network analysis, Water Resour. Manage. 27 (2013), 3697–3711.

[17] Southern region irrigation hydrology center, Bureau of Water Management and Hyrology, Daily runoff data in Songkhla Basin, Retrieved from http://hydro-8.com/ (2017).