International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies

http://TuEngr.com

PAPER ID: 11A11S

STOCHASTIC SEASONAL RAINFALL FORECASTING MODEL FOR WATER RESOURCES MANAGEMENT IN KLONG YAI RIVER BASIN, THAILAND

Kwanchanok Oonta-on [1], Uruya Weesakul [1*] and Nkrintra Singhrattna [2]

[1] *Department of Civil Engineering, Thammasat School of Engineering, Thammasat University, THAILAND.*
[2] *Department of Public Works and Town & Country Planning, Ministry of Interior, THAILAND.*

ARTICLEINFO

Article history:
Received 20 February 2020
Received in revised form 01 May 2020
Accepted 11 May 2020
Available online 26 May 2020

Keywords:
Stochastic hydrologic model; Climate change; Rainfall forecast model; Atmospheric variables; K-nearest neighbor

ABSTRACT

During the past decades, Thailand has faced flood and drought problems caused by the effects of global climate change which directly affected rainfall-runoff and water allocation for all demands. This research aims to study the relationships between rainfall in the study area and large-scale atmospheric variables at the different levels from surface level to 10-mb level at lead time 4-15 months. The predictors were identified to develop a 3-month rainfall forecast model to support water resources management. The model evaluations using the leave-one-out technique were evaluated by the goodness-of-fit of statistics and probability density function (PDF) of observed rainfall. The results revealed that the observed data could be preserved to the estimated data for the goodness-of-fit technique and the PDF technique, observed and predicted rainfall were classified into five categories (namely: extremely dry, dry, normal, wet and extremely wet respectively), represented the maximum efficiency in pre-monsoon season (May-June-July) and minimum model efficiency in monsoon season (August-September-October) approximately 42.5% and 31.7% respectively.

Disciplinary: Water Resources Engineering and Management, Environmental Climate Change, Sustainability.

1. INTRODUCTION

Over the past decades, many regions around the world have been exposed to severe and frequent natural disasters caused by global climate change such as the 2004 Indian Ocean earthquake and tsunami, extremely severe cyclonic storm Nargis in Myanmar during 2008, the 2010 Port-au-Prince earthquake in Haiti, an El Niño event caused a drought in some parts of Southern Africa in late 2015 and during 2018–20 Mozambique drought. Although some regions have not yet experienced extreme events or severe disasters, they have to face large-scale atmospheric variabilities such as

*Corresponding author (Yasser Ali Baeshen). Email: ybaeshen@kau.edu.sa

more floods and droughts. Hence, large-scale atmospheric variability also affects many areas of the world. Many studies reveal the relationship between large-scale atmospheric and rainfall variability (Loo et al., 2014; Mamombe et al., 2016; Barcikowska et al.; 2018; Sipayung et al., 2018). Also, the studies of El Niño and La Niña events impact on rainfall variability in many countries (Bridhikitti, 2017; Ueangsawat, 2015; Li et al., 2019).

Likewise, many regions of Thailand have experienced flood and drought problems caused by large-scale atmospheric variability. Its impacts on Thailand are prolonged droughts, decreased agricultural, severe flooding, health-related issues and the drought also challenges over water resource management within the region (Marks, 2011). Uncertainty of climate change was studied in Khon Kaen Province Thailand, the results showed that rainfall patterns were changed with longer high intense dry seasons following by shorter high intense rainy seasons. It has been found in that study that explicit consideration of climate change did not exist and that water availability patterns will continue the same as the past results (Friend and Thinphanga, 2018). Studies of impacts of climate change on rainfall in Thailand were done in various regions, for example, in the upper Chao Phraya river basin interannual variability of summer monsoon rainfall were analyzed (Singharattana et al., 2005; Singharattana et al., 2013) and the rainfall variability of four basins in the north, northeast, coast east and south of Thailand (Weesakul et al., 2013; Weesakul and Yodpongpiput, 2015; Weesakul and Oonta-on, 2015).

In several years, the statistical model applied for rainfall forecast models are quite popular and widely used because it can be used easily and not complicated. Many statistical model have been developed for making predictions in different regions around the world. Various researches related to forecasting rainfall by using the relationship of large-scale atmospheric variables (LAV) and rainfall were also developed with the effort to offer the accuracy deterministic results as much as possible. For example, linear and non-linear models are also widely used and applied for seasonal forecasting (Ansari, 2013, Badr et al, 2015; Djibo et al, 2015; Hossain et al, 2018). The main limitation of the linear model is the assumption of linearly separable and assumes that the relationship is a straight-line. The non-linear model provides a better fit because it is both unbiased and produces smaller errors. An artificial neural network (ANN) is a non-linear model (Badr et al., 2013, Rasel et al., 2016). Disadvantages of the ANN models are complicated models, need more data to train and unexplained behavior of the network. Compared to an ANN model, the regression model is easier to explain and understand. The k-nearest neighbor algorithm (k-nn) is an interesting technique applied for hydrological models because it is a non-linear and non-parametric model.

In nature, the rainfall prediction cannot be specified exactly for one result because of the uncertainty of climate. So the stochastic modeling technique was selected to solve for the hydrologic problems rather than the statistic modeling technique. Singharattana et al. (2012) studied hydroclimate variability and long-lead forecasting of rainfall over Thailand by LAV (i.e. the surface air temperature (SAT), sea level pressure (SLP), surface zonal wind (SXW) and surface meridian wind (SYW) using the modified k-nn model. The results indicated the long-lead forecasting ability of the modified k-nn model for rainfall in the Ping River Basin is an impressive result that can be applied to manage water resources and develop long term plans for the reservoir. The modified k-nn was developed to forecast rainfall in the Mun and Chi River Basin, Thailand. Its performance reveals that the model can predict seasonal rainfall and providing sufficient information for appropriate cropping pattern planning in the area (Weesakul et al., 2014, Weesakul et al., 2016).

It is interesting to explore the capability of the modified k-nn model in forecasting seasonal rainfall in other river basins with different hydrological characteristics. This study aims to develop a modified k-nn model to forecast seasonal rainfall over Klong Yai River Basin using LAV as predictors.

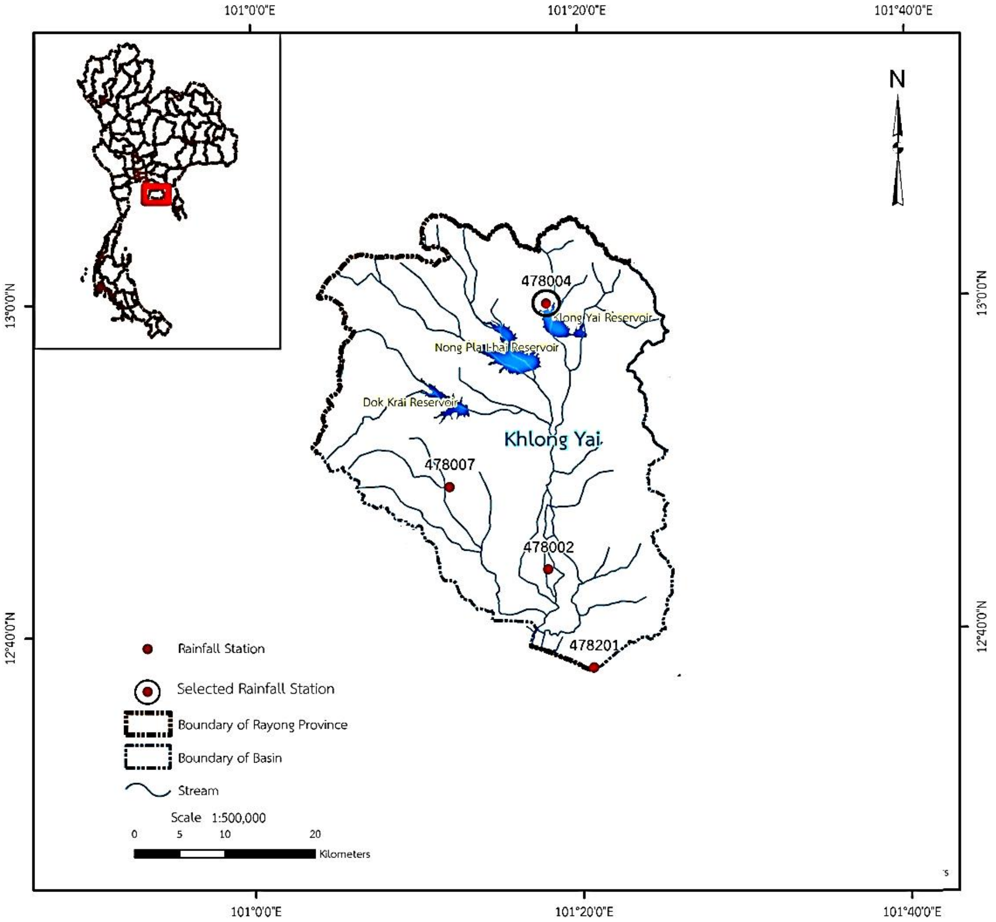

Figure 1: Klong Yai River Basin, Thailand.

2. STUDY AREA AND METEOROLOGICAL DATA

Klong Yai River Basin is a tributary of the Eastern River Basin, located on the east coast of the Gulf of Thailand between 12°38'-13°00'N latitude and 101°00'-101°40'E longitude covering an area of 1,804 km^2, as shown in Figure 1. Most areas of the Klong Yai River Basin are in Rayong province and some parts are in Chonburi province. Its topography is mostly a coastal plain. The upper part of the basin is a corrugated plateau and hill. The main river basin is the Rayong or Klong Yai River, which consists of many branches flowing into the Gulf of Thailand. The climate of the Klong Yai River Basin is tropical monsoon which is under the influence of southwest monsoon and northeast monsoon. Wet and dry season caused by seasonal shift in winds or monsoon; the tropical rain belt brings additional rainfall during the monsoon season. The transition period of the northeast to southwest monsoon starts from mid-February to mid-May. After that, it has heavy rain due to the influence of the southwest monsoon originating in the Indian Ocean from May to October. The average temperature since 1981-2017 equals 29°C. The winter season begins in November and ends in February influenced by the eastern monsoon and carries cool air mass from upper China to cover

*Corresponding author (Yasser Ali Baeshen). Email: ybaeshen@kau.edu.sa Volume 11 No.11 ISSN 2228-9860 eISSN 1906-9642 CODEN: ITJEA8 Paper ID:11A11S http://TUENGR.COM/V11A/11A11S.pdf DOI: 10.14456/ITJEMAST.2020.223

Thailand resulting in the average temperature in the winter since 1981-2017 equal to 26.9°C. Summer starts from March to April and the average temperature is 29.4°C.

The monthly cumulative rainfall calculated from the amount of daily rainfall data from four stations in the Klong Yai River Basin ranged from 1976-2017, collected and published by the Meteorological Department. The average monthly rainfall (Figure 2) shows the beginning and the preliminary peak of rainfall in May is 176.8 mm. and slightly decreases in July (the pre-monsoon season; MJJ). The average monthly rainfall for pre-monsoon ranged from 152.7-176.8 mm. The large quantities of rainfall are observed in August to October (ASO (the monsoon season)) is ranged from 136.1-236.0 mm. The amount of average monthly rainfall in the dry season, Nov-Jan (NDJ), and Feb-May (FMA) are ranged from 11.1-26.0 mm. and 31.3-85.7 mm. The average annual rainfall from four stations is 1333.9 mm. and the minimum and maximum of average annual rainfall are 812.0 mm and 1858.1 mm. was detected in 1979 and 1983, respectively.

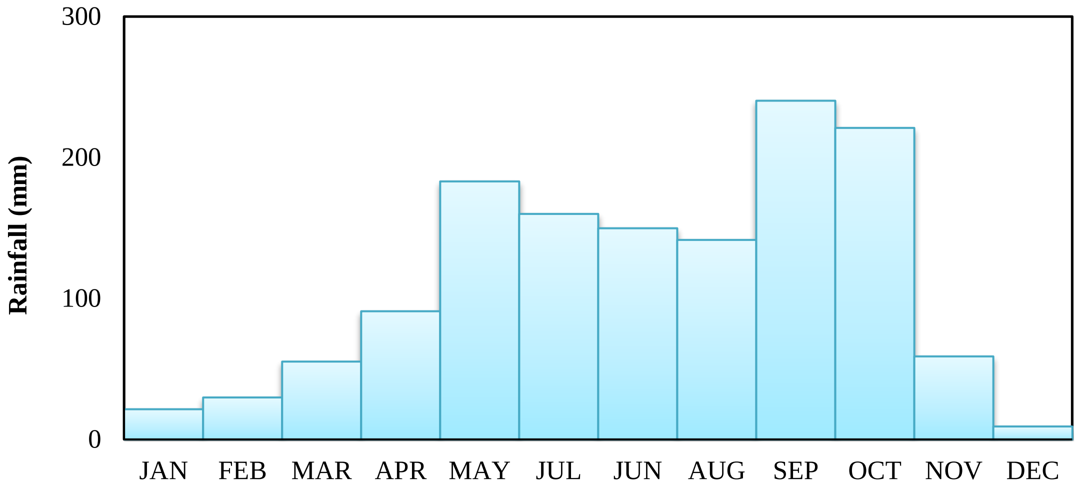

Figure 2: Averaged monthly rainfall from 1976-2017 in Klong Yai River Basin.

3. METHODOLOGY

3.1 DATA COLLECTION

There are two important data in this research consisted of the monthly rainfall data and the large-scale atmospheric variables (LAV). Very few rainfall stations are distributed over the Klong Yai River Basin as shown in Figure 1. The monthly rainfall data of four stations were collected by the Thailand Meteorological Department (TMD) cover the period 1976-2017. Data consideration was done under three conditions. The first one, the rainfall station should have located above the reservoir because this research aims to apply the rainfall forecast results for future water management such as finding the amount of water entering the reservoir for adequate water allocation. The second one is the number of completed rainfall data available at each station is not less than 95% and the last one is the consistency of duration is not less than 30 years. From Figure 1, the rainfall station above the reservoir is station 478004 with 4.0% incomplete data and forty-year rainfall duration consistency (1977-2017) while the incomplete data at station 478001, 478002, and 478007 are equal to 22%, 10.8%, and 4.5% respectively. Therefore, the data at station 478004 is suitable for forecast the 3-month rainfall in the Klong Yai River basin.

Large-scale atmospheric variables (LAV) is the global climate data are collected and provided by the National Center for Environmental Prediction (NCEP) which presented in Grid cell size 2.5°

latitude x 2.5° longitude around the world and at different heights, ranging from the surface level to the higher levels above the atmosphere. LAV data includes 13 atmospheric variables such as geopotential height (GH), zonal wind (u), meridional wind (v), air temperature (AT), sea surface temperature (SST), relative humidity (RH), specific humidity (SH), pressure (P), omega (O), sea level pressure (SLP), precipitation rate (PR), precipitable water (PW), outgoing long-wave radiation (OLR). The LAV data period from 1977-2014 was used in this study.

3.2 PREDICTOR IDENTIFICATION

3.2.1 CORRELATION MAPS OF THE LARGE-SCALE ATMOSPHERIC VARIABLES

Correlation map is the result of correlation analysis between the rainfall in the study area and the large-scale atmospheric variables (LAV) at different coordinates. The correlation map is an interactive plot that is analyzed and provided by the Earth System Research Laboratory (ESLR) of the National Oceanic and Atmospheric Administration (NOAA). The important condition to identify predictors based on the significant relationship which can be tested the hypothesis by Fisher's Transformation (z') method (Haan, 2002).

$$z' = 0.5 * \ln\left(\frac{1+r}{1-r}\right) \quad (1),$$

where z' is an approximately normally distributed, and is used to compare with z of the standard normal distribution to obtain the significant level (p) of correlation, r is the correlation coefficient between two independent data sets.

3.2.2 COMBINATION CASE OF PREDICTORS

To reduce the complexity of the rainfall forecast model, the optimal predictors' selection method is very important. In this study, generalized cross-validation (GCV) with the leave-one-out technique was used as a selection method. For m multiple independent variables, there are 2^m-1 cross-combination cases of variables. In this case, only one point was picked as the test set. The model was built on all the remaining, complementary points, and evaluates its error on the single-point held out. A generalization error estimate is obtained by repeating this procedure for each of the training points available, averaging the results. Each combination case is evaluated using this technique and the GCV is calculated by:

$$GCV = \frac{\sum_{i=1}^{n} \frac{e_i^2}{n}}{(1-m/n)^2} \quad (2),$$

where e_i is the residual error in each case, n is the total number of data points, and m is the number of variables.

3.2.3 RAINFALL FORECAST MODEL

Regression analysis is used to predict dependent variables by considering the relationship between a dependent variable and one or more independent variables. The most widely used regression analysis is the parametric approach based on the exact assumption of the function such as a linear regression in which its outcome can be estimated from a linear relationship. Each data point on the regression line has one residual. If all data points are on the regression line, the residual is equal to zero. The accuracy of the model depends on the data distribution. Unlike the parametric approach, the nonparametric approach does not require making assumptions or parameters. Therefore, the nonparametric approach is more flexible and accurate than the parametric approach. For this reason, the modified k-nearest neighbor model (k-nn) which is a nonparametric approach used for fit regression and forecast model. For fitting process, local regression fit, the minimum GCV was calculated from a couple of the neighbor size (k) and polynomial order (p) is almost always at order 1 or 2 because a low-order polynomial and the simple models can be easily fitted with the data more than higher-order polynomial.

$$GCV(k,p) = \frac{\sum_{i=1}^{n} \frac{e_i^2}{n}}{(1-m/n)^2} \quad (3),$$

where e_i is the residual, n is the total number of data points, and m is the number of variables.

The mean estimation of rainfall ($\bar{y}_1$, $\bar{y}_2$, $\bar{y}_3$,…, $\bar{y}_n$) and the residual (e_1, e_2, e_3,…, e_n) of independent variables (x_1, x_2, x_3,…, x_n) can be calculated from the developed local regression. The rainfall forecast ($\bar{y}_i$) at any point ($\bar{y}_i$) is made from a simple average of a small subset of nearby points ($\bar{y}_i$). The modified k-nn makes predictions based on the residual (e_i) of the k neighbors closest and the value of k can be computed in term of $\sqrt{n-1}$, n is the total number of data. The random selection of the residuals used the weight function which gives the most weight to the residual nearest the point of estimation and the least weight to the residual that is furthest away. The weight function can be calculated as

$$W(j) = \frac{1/j}{\sum_{i=1}^{k}(1/i)} \quad (4),$$

where $W(j)$ is the weight value of a neighbor of ($\bar{x}_i$) in j rank and k is the size of neighbors. To define closeness, the distance between point ($\bar{x}_i$) and (x_i) of the k neighbors must be calculated by the Euclidean distance equations:

$$d_i = \sqrt{(\bar{x}_i - x_i)^2} \quad \text{for univariate data} \quad (5),$$

$$d_i = \sqrt{\sum_{j=1}^{m}(\bar{x}_{i,j} - x_{i,j})^2} \quad \text{for multivariate data} \quad (6),$$

where i is 1, 2, 3, …, n, and m is the number of the identified predictors.

Instead of making for only one prediction of the most likely case, an ensemble of prediction is produced. Therefore, the number of ensembles of rainfall (N) was repeated many times until reach to achieve of rainfall forecasting ensemble; in this case, N is 300 ensembles.

3.2.4 EVALUATION OF MODIFIED K-NEAREST NEIGHBOR MODEL

The evaluation of the Modified k-nn model is based on the Leave-one-out technique. The regression between the independent variable (LAV) and the dependent variable (the 3-month rainfall from 1976-2017) in the Khlong Yai River Basin was developed. The leave-one-out technique was done at every point in this case total points are 39 points and after that leave 1 data point out from 38 training data. These data points are used as validation data. Repeat until complete 39 points at 12 different lead times.

(a) Goodness-of-fit technique

The goodness-of-fit technique is represented by the box plots. The box plot in each lag time during MJJ, ASO, NDJ, and FMA can be estimated by 300 simulations from the modified k-nn model. The criteria used to evaluate the model efficiency based on five standardizes of the statistics measurement including mean, median, standard deviation, interquartile range, and skewness. A box plot shows the quartile range of ensembles between the 25th percentile (QL) and 75 (QU). The whiskers are lines running from the box to the maximum and minimum values and the median is the line dividing the box. The solid line represents the statics observed data are plotted overlying the box plot of statics estimated data. The model efficiency depends on the consistency in statistics preserving between the observed and estimated data.

(b) Probability density function technique

The modified k-nn model evaluation used the probability concept to capture the Probability density function (PDF) of the historical 3-month rainfall data. The possible outcome of the probability indicates the chance that a particular event will occur. Probabilities can be expressed as proportions ranging 0-1. The historical 3-month rainfall data (MJJ, ASO, NDJ, and FMA), during 1977-2017 in the Klong Yai River Basin was considered and then classified into five categories at 20^{th}, 40^{th}, 60^{th}, and 80^{th} (namely: extremely dry, dry, normal, wet and extremely wet respectively). The probability can measure from

$$P(E)=\frac{N(E)}{N(S)} \quad (7),$$

where $P(E)$ is the probability of the possible outcomes, $N(E)$ is the number outcomes favorable and $N(S)$ is the total number of outcomes.

For a given year, 300 ensemble rainfall from the modified *k-nn* model was classified into five categories same as the classification of historical data and then calculate the probability of each category .After that, the probability of calculated rainfall each year could be identified as the possibility of rainfall interval by considered the maximum probability .Finally, compare the possibility categories between the historical rainfall and the calculated rainfall for example, in 1980, the probability of 3-month rainfall from historical data is dry and the category of rainfall from the model is also dry that means the model can be captured PDF and the model efficiency can be measured from the percentage of accuracy at that period.

4. RESULTS AND DISCUSSION

4.1 PREDICTOR IDENTIFICATION AND COMBINATION CASE

The correlation map, analyzed and published by NOAA, was used to identify the most influence large-scale atmospheric variables (LAV) on seasonal rainfall in the Klong Yai River Basin. The identified LAV were used to develop the 3-month rainfall forecast model. The correlation map demonstrated the statistical relationship between 3-month rainfall for four periods (MJJ, ASO, NDJ, and FMA) and 13 LAV for 12 different lead times ranging from 4-15 months covering potential regions such as study basin, the Indian Ocean, the Bay of Bengal, the Gulf of Thailand, South China Sea and the Pacific Ocean. The LAV in this research was selected based on the significant relationship at a 99% confidence level which was tested by Fisher's transformation from rainfall data 39 years (1976-2014). So the upper and lower bounds of the correlation coefficient (r) are ± 0.4. For example, the relationship between ASO rainfall and Zonal wind was found at a 30 mb level as shown in Figure 3. The correlation maps represented a higher Zonal wind over the study basin and nearby the sea, the Bay of Bengal, the Gulf of Thailand, and the Pacific Ocean, is associated with increasing ASO rainfall. The location, which has a high positive correlation is +0.4 and its long lead times 4-9 months over the Bay, is shown in the solid box.

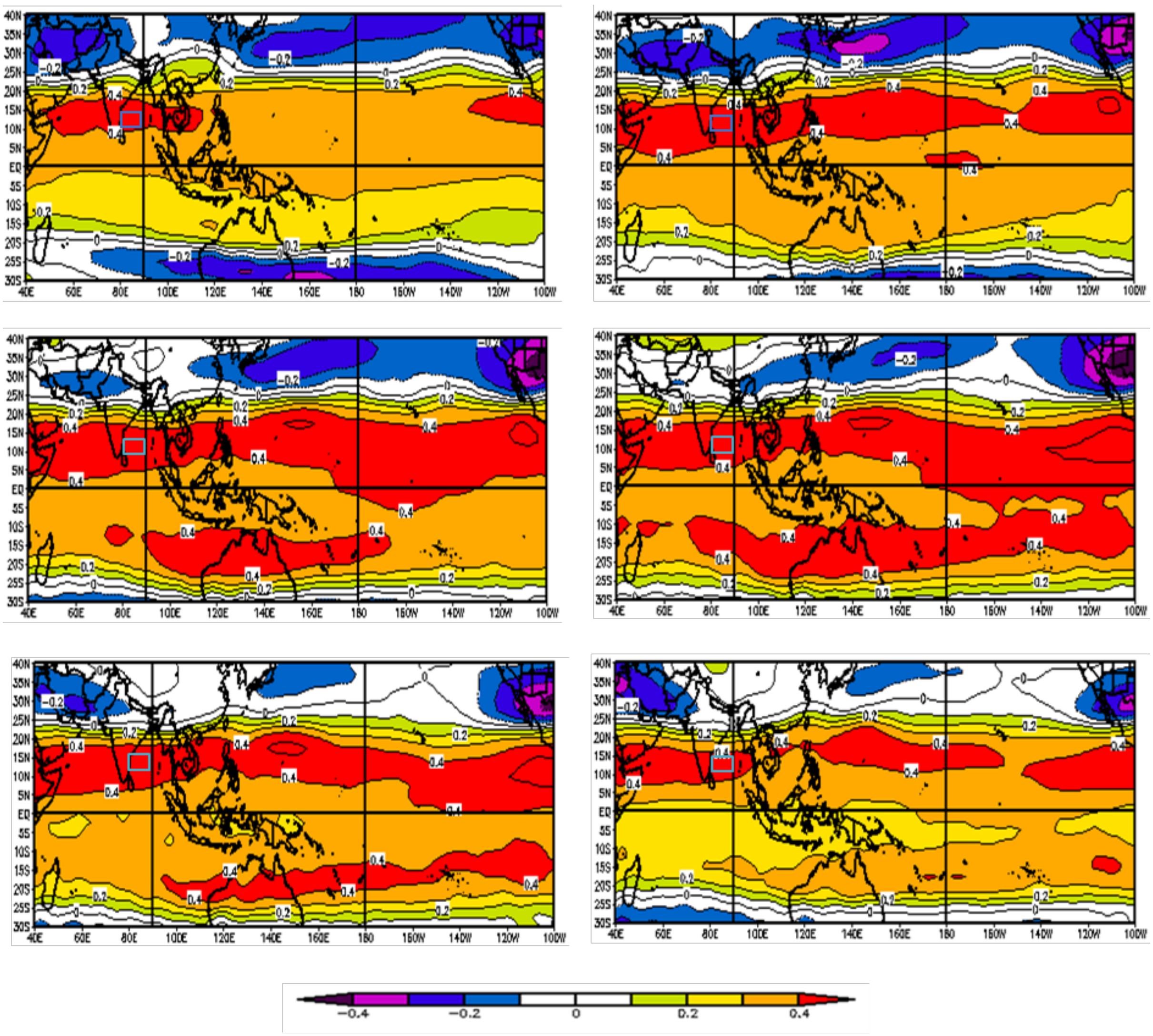

Figure 3: The correlation maps between ASO rainfall at a lead time 4-9 months, respectively from left-right and top-bottom and Zonal wind at 30 mb level

Table 1: Summary of the identified predictor from correlation maps in the Klong Yai basin

Season of rainfall	Atmospheric variables	Location		level (mb)	Lead time (month)
		Latitude	Longitude		
MJJ	GH	7.5-12.5°N	180-185°E	250	8-10
	u1	30-35°N	180-185°E	100	4-7
	u2	0-5°N	175-180°E	100	11-13
	AT1	25-30°N	175-180°E	400	4-6
	AT2	0-5°N	210-217.5°E	400	9-13
	RH1	2.5-7.5°N	135-140°E	925	10-14
	RH2	2.5-7.5°N	210-215°E	925	13-15
	SH1	0-5°N	180-185°E	500	8-15
	SH2	12.5-17.5°N	205-210°E	500	4-6
	O	10-15°N	207.5-212.5°E	500	4-6
	SLP	7.5-12.5°S	205-210°E	Surface	12-14
	PW	5-10°S	180-190°E	Surface	8-12
ASO	U	10-15°N	85-95°E	30	4-9
	V	12.5-20°N	180-190°E	10	11-15
	AT	12.5-17.5°N	157.5-167.5°E	850	12-15
	SH1	7.5-15°N	120-130°E	400	4-8
	SH2	0-5°S	110-120°E	400	10-14
NDJ	GH	0-5°N	135-140°E	600	4-11
	U	2.5-7.5°N	230-235°E	700	6-11
	V	10-15°N	207.5-212.5 °E	150	5-11
	AT	2.5°S-2.5°N	250-260 °E	Surface	4-9
	SST	0-5°N	255-260 °E	Surface	4-9
	RH	0-5°N	255-260 °E	400	6-11
	SH	0-5°N	110-115 ° N	400	5-15
	P	0-5°N	275-250 °E	Surface	5-8
	SLP	2.5-7.5°S	255-260°E	Surface	4-7
	PR	2.5°S-2.5°N	255-260°E	Surface	5-9
	PW	2.5-7.5°N	255-260°E	Surface	4-8
	OLR	0-5°N	255-260°E	Surface	6-11
FMA	GH	0-5°N	240-250 °E	850	7-11
	U	2.5-7.5°N	162.5-167.5 °E	850	4-11
	v1	0-5°N	155-160 °E	200	4-9
	v2	5-10°N	92.5-97.5 °E	200	13-15
	AT	10-15°N	110-120 °E	50	6-13
	SST	7.5°-12.5°	145-150 °E	Surface	4-9
	RH	5-10°N	175-180 °E	400	4-11
	SH1	2.5°S-2.5°N	217.5-222.5 °E	400	4-9
	SH2	5-10°N	95-100 °E	400	10-12
	P	0-5°S	220-225 °E	Surface	4-9
	O	5-10°N	155-160°E	925	4-12
	SLP	0-5°S	207.5-212.5 °E	Surface	4-9
	PR	0-5°N	202.5-207.5°E	Surface	4-8
	PW	7.5°S-2.5°N	105-115°E	Surface	4-8
	OLR	0-5°S	202.5-207.5°E	Surface	4-9

The summary results of the identified predictors from the correlation map at a 99% confidence level were shown in Table 1. The results indicated that the significant relationships covered the different areas depending on the period of 3-month rainfall at a different height, and the long lead times. The significant relationship for the monsoon season (i.e. MJJ and ASO rainfall) and the selected predictors have long lead times ranging from 3-8 lead times. The predictors for pre-monsoon rainfall (MJJ) are identified over the equatorial eastern Pacific Ocean which its relationships are the positive and negative correlation. There are five predictors such as v, SST, P, PR, and OLR have significant correlations at every level but short lead times. In addition to the above predictors, a significant relationship can be found with 3-8 months lead times. For example, the relationship between MJJ rainfall and GH at the level 250 mb can be found over the equatorial eastern Pacific Ocean at lead times 8-12 months. For ASO rainfall, five influence predictors consist of u, v, AT RH, and SH were found over the Gulf of Thailand, the South China Sea, Jawa (Indonesia), and the West Pacific Ocean. The significant relationships are observed at 3-6 lead times. For example, a significant relationship between ASO rainfall and v at the level 10 mb has long lead 11-15 months.

The significant relationship between rainfall and predictors in the dry season (i.e. NDJ and FMA) has long lead time more than monsoon season. Generally, the significant relationships appeared in different regions such as the Gulf of Thailand, the South China Sea, the equatorial Indian Ocean, and the East Pacific Ocean. The longest lead time, appeared in the region of the South China Sea with a significant relationship between NDJ rainfall and SH at the level 400 mb, was 5-15 months. For other significant relationships have 4-7 long lead times to accept the significant relationship between NDJ rainfall and O has no long lead time. The predictors for rainfall during FMA were found with the significant relationships between FMA rainfall and O at level 925 mb, which appeared in the east of Pacific Ocean, with the longest lead times 4-12 months compared with other predictors.

The predictors from Table 1 were used to develop the 3-month rainfall forecast model. Instead of using overall predictors which is very complicated to use various predictors to construct the model. Reducing the predictors with minimizing the criteria of generalizes cross-validation (GCV) can solve this complex problem. The advantage of reducing the redundancy of these predictors with GCV results in 1-4 optimal predictors in each lag time. The summary of the optimal predictors for 12 different lag times from GCV was shown in Table 2.

Table 2: Summary of the optimal predictors for 12 different lag times.

Seasonal Rainfall	Lead time (month)											
	1	2	3	4	5	6	7	8	9	10	11	12
MJJ	O	GH	AT1	AT1	AT2	AT2	RH1	AT2	AT2	O	RH1	SH1
			SH2	SH2	PW	RH1	SH1	RH1	RH1	RH1	SH2	
						SH1	SH2	SH1	SH1	SH2		
ASO	U	u	u	u	u	u	u	u	u	v	v	v
	V	AT				v	v	v	v			SH1
	SH2						SH2	SH2				
NDJ	AT	v	GH	AT	OLR	SLP	v	v	SH	v	SH	SH
	SST	SLP	SLP			OLR		PR				
	PR	PR										
	OLR	OLR										
FMA	GH	v1	GH	GH	GH	RH	AT	AT	AT	P	SH1	O
	SH1	AT	SH1	v1	v1	PR	O	O	P	O	P	SLP
	O		O						SLP	SLP	O	PW
											SLP	

4.2 EVALUATION OF MODIFIED K-NEAREST NEIGHBOR MODEL

(a) Goodness-of-fit technique

The goodness-of-fit of a statistical model describes how well it preserves to observations. The box plot of 300 simulated rainfall means, median, SD, interquartile range, and skewness was developed and fitted to the observed rainfall for four periods (MJJ, ASO, NDJ, and FMA) at 12 different lag times. For monsoon season (i.e. MJJ and ASO), the observed mean is 390 and 541 mm. respectively.

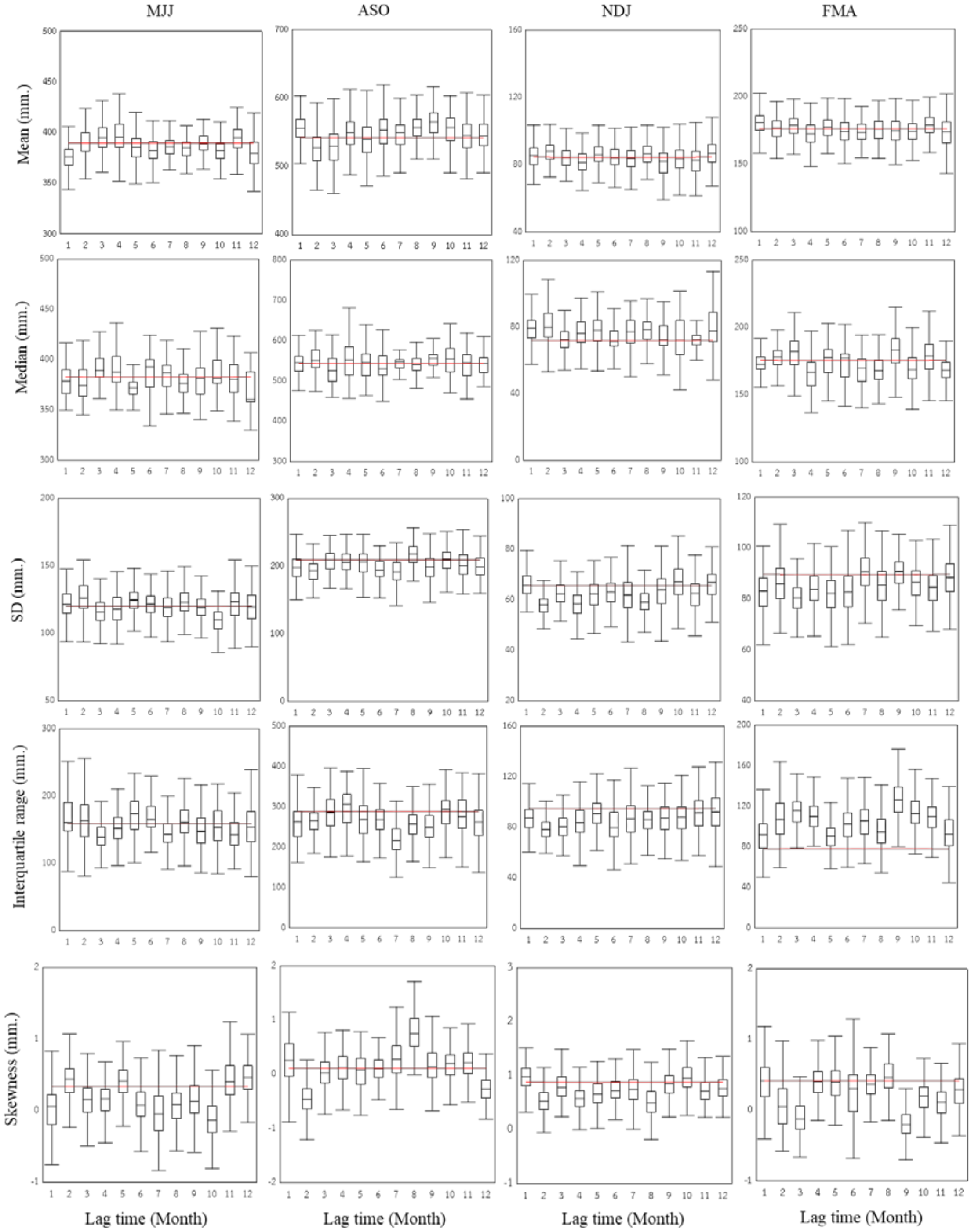

Figure 4: The box plots of 300 simulated rainfall from the modified k-nn model for MJJ, ASO, NDJ, and FMA in 1977-2017 with 12 different lag times. (The red lines represent the statistics of observations.)

The goodness-of-fit technique was used to evaluate the Modified k-nn model by fitting the historical data to the box plots of 300 forecasting ensembles of rainfall using the basic statistics of measurement data for four periods (MJJ, ASO, NDJ, and FMA) ranging from 1977-2017. The basic statistics of measurement data consisting of mean, median, standard deviation, interquartile range, and skewness are represented in Table 3. From Figure 4, the results indicated that the forecasted data could be preserved to the observed data in every period except for the box plots of the interquartile range during FMA at 3, 4, and 5 months lag time which was found to be overestimated. Also, an underestimation of the skewness box plot was found in the same duration (FMA) at 9 months lag time.

(b) Probability density function technique

The model evaluation used the probability of capturing PDF of the 3-month rainfall data by comparing the probability of historical rainfall data and computed data with 5 categories. The results of efficiencies for a 3-month rainfall forecast model in the Klong Yai river basin during 1977-2017 were obtained in Table 3. According to these results, the maximum efficiency to capture the historical 3-month rainfall data was found in pre-monsoon (MJJ) equal to 42.5% at 6- and 9-month lag time. The minimum efficiency of the same period is equal to 15% at a 12-month lag. The efficiency of the model during the monsoon season was lowest. The maximum and minimum efficiency in this period is 31.7% at 9-month lag and 7.3% at a 3-month lag. For the dry season (NDJ and FMA), the maximum efficiency is 35.0% at a 6-month lag and 41.5% at a 7-month lag. The minimum efficiency in NDJ is 15% at 9 and 11-month lag. For FMA, the minimum efficiency is 19.5% at a 12-month lag. Therefore, the efficiency of the Modified k-nn model to forecast 3-month rainfall depends on lag time and the probability in preserve the PDF of the observed data in pre-monsoon and monsoon season which are 7.3-42.5% and 15.0-41.5% respectively.

Table 3: The efficiency of 3-month rainfall forecast model in Klong Yai river basin during 1977-2017

Periods	Model efficiency (%)											
	Lag time (month)											
	1	2	3	4	5	6	7	8	9	10	11	12
MJJ	30.0	27.5	37.5	20.0	35.5	42.5	27.5	35.0	42.5	25.0	25.0	15.0
AOS	29.3	24.4	7.3	24.4	9.8	12.2	24.4	26.8	31.7	26.8	26.8	22.0
NDJ	32.5	32.5	25.0	22.5	22.5	35.0	20.0	32.5	15.0	12.5	15.0	17.5
FMA	24.4	22.0	34.1	26.8	29.3	22.0	41.5	34.1	26.8	26.8	34.1	19.5

Legend: Maximum efficiency Minimum efficiency

5. CONCLUSION

High variation of rainfall due to climate change requires a stochastic hydrologic model, considering the uncertainty of rainfall, to forecast rainfall more accurately. A modified k-nn model, based on a stochastic nonparametric approach, was developed in this study to forecast seasonal rainfall over the Klong Yai River basin. According to rainfall characteristics in the area, rainfall was grouped into four seasons which are two-premonsoon seasons (November-December-January: NDJ and February-March-April: FMA) and two monsoon seasons (May-June-July: MJJ and August-September-October: ASO). Monthly rainfall data during 1976-2017 were used for analysis in the study. A variety of Large Atmospheric Variables at different levels were analyzed their correlations with seasonal rainfall over the study area. It has been found that different LAV at different levels influenced seasonal rainfall over the river basin with different lagging time. Generalized Cross-Validation was used to select the optimal predictors for the modified k-nn model. Finally, only one to four LAV were used as predictors to forecast seasonal rainfall for 1 to 12 months lead time with different LAV for each season and for each lead time. Goodness-of-fit and probability density function techniques were used to evaluate the efficiency of the model. The results of the model simulation reveal that the forecasted rainfall was well preserved to the observed data for the goodness-of-fit technique. For the probability density function technique, the modified k-nn model performed well for forecasting of rainfall in pre-monsoon season with 43% efficiency, but the performance reduced to 32% efficiency for forecasting of monsoon rainfall. However, the model developed in the study can be used to forecast seasonal rainfall for one year ahead providing useful information as a guideline for water resources management in Klong Yai River Basin.

6. AVAILABILITY OF DATA AND MATERIAL

Data can be made available by contacting the corresponding author.

7. ACKNOWLEDGEMENTS

The authors are thankful to the Thai Meteorological Department for providing rainfall data and PTT Public Company Limited for research fund and kind cooperation in the testing of the model.

8. REFERENCES

Ansari, H. (2013). Forecasting seasonal and annual rainfall based on nonlinear modeling with Gamma test in North of Iran. *Int. J. Eng. Pract. Res*, 2(1), 16-29.

Barcikowska, M. J., Kapnick, S. B., & Feser, F. (2018). Impact of large-scale circulation changes in the North Atlantic sector on the current and future Mediterranean winter hydroclimate. *Climate dynamics*, 50(5-6), 2039-2059.

Bridhikitti, A. (2019). Multi-decadal trends and oscillations of Southeast Asian monsoon rainfall in northern Thailand. *Songklanakarin Journal of Science & Technology*, 41(1). 74-80.

Djibo, A.G. (2015b). Development and assessment of non-linear and non-stationary seasonal rainfall forecast models for the Serba watershed, West Africa. *Journal of Hydrology*: Regional Studies, 4, 134-152.

Djibo, A.G., et al. (2015a). Linear and Non-Linear Approaches for Statistical Seasonal Rainfall Forecast in the Sirba Watershed Region (SAHEL). *Climate 2015*, 3, 727-752; DOI:10.3390/cli3030727.

Friend, R., & Thinphanga, P. (2018). Urban water crises under future uncertainties: the case of institutional and infrastructure complexity in Khon Kaen, *Thailand. Sustainability*, 10 (11): 3921.

Rasel, H.M. (2016). Application of Artificial Neural Network for Seasonal Rainfall Forecasting: A Case Study for South Australia. *Proceeding of the World Congress on Engineering* 2016, Vol I.

Hamada, S., Badr, Zaitchik, B.F. (2013). Application of Statistical Models to the Prediction of Seasonal Rainfall Anomalies over the Sahel. *Journal of Applied Meteorological and Climatology*, 53, DOI: 10.1175/JAMC-D-13-0181.1.

Hossain, I., Esha, R., & Alam Imteaz, M. (2018). An attempt to use non-linear regression modeling technique in long-term seasonal rainfall forecasting for Australian capital territory. *Geosciences*, 8(8), 282.

Li, Y., Strapassonb, A., Rojas, O. (2019). Assessment of ElNiño and LaNiña impacts on China: Enhancing the Early Warning System on Food and Agriculture. Weather and Climate Extremes, doi.org/10.1016/j.wace.2019.100208.

Loo, Y. Y., Billa, L., & Singh, A. (2015). Effect of climate change on seasonal monsoon in Asia and its impact on the variability of monsoon rainfall in Southeast Asia. *Geosci Front* 6: 817–823.

Mamombe, V., Kim, W., Choi, Y.S. (2016). Rainfall variability over Zimbabwe and its relation to the large-scale atmosphere-ocean processes. *Int. J. Climatol*, DOI: 10.1002/joc.4752

Marks, D.. (2011). Climate Change and Thailand: Impact and Response. International Contemporary

Southeast Asia. *Journal of International and Strategic Affairs*, 33(2), 229-258.

Singhrattna N, Babel, M.S., and Perret, S.R. (2012). Hydroclimate Variability and Long-Lead Forecasting of Rainfall over Thailand by Large-Scale Atmospheric Variables. *Hydrological Sciences Journal-Journal des Sciences Hydrologiques*, 57(1), 2012.

Singhrattna N, Rajagopalan B, Kumar KK, and Clark. (2005). Interannual and Interdecadal variability of Thailand summer Monsoon Season. *J Clim*, 18, 1697-1708.

Singhrattna N., Mukand, Perret, S.R. (2013). Changes in Summer Monsoon Rainfall in the Upper Chao Phraya River Basin, Thailand. *Climate Research*, 49: 155-168, 2011, DOI 10.3354/cr01015.

Sipayung, S. B., Nurlatifah, A., Siswanto, B., & Slamet, L. S. (2018). Analysis of climate change impact on rainfall pattern of Sambas district, West Kalimantan. In *IOP Conference Series: Earth and Environmental Science*, 149(1), 1755-1315.

Ueangsawat, K., Nilsamranchit, S., & Jintrawet, A. (2015). The fate of the ENSO phase in upper Northern Thailand, a case study in Chiang Mai. Agriculture and Agricultural Science Procedia, 5, 2-8.

Weesakul, U., Oonta-on, K. (2015). Impact of Climate Change on Annual Rainfall over Eastern River Basin in Thailand: A Warning signal for Future Industrial Water Supply. *The 4th International Symposium on Engineering, Energy, and Environment.*

Weesakul, U., Singhrattna, N., and Luangdilok, N. (2013). Effect of Climate Change on Thailand Rainfall Variability. The 3rd International Symposium on Engineering, Energy and Environments.

Weesakul, U., Singhrattna, N., and Luangdilok, N. (2014). Rainfall Forecast in Northeast of Thailand Using Modified K-Nearest Neighbor. *KKU Engineering Journal*, 41(1): 1-10.

Weesakul, U., Singhrattna, N., Yodpongpiput, P. (2016). Seasonal Rainfall Forecast for Cropping Pattern Planning Using a Modified K-Nearest Neighbor Model. *KKU Engineering Journal*, 43(3): 156-161.

Kwanchanok Oonta-on is a doctoral candidate at the Department of Civil Engineering, Faculty of Engineering, Thammasat University. She graduated with her Bachelor of Civil Engineering from Sirindhorn International Institute of Technology, Thammasat University. She earned her Master of Engineering in Civil Engineering from Thammasat University. Her research interests Concern Rainfall Forecast Model and Water Resources Management and Planning.

Dr.Uruya Weesakut is an Associate Professor at the Department of Civil Engineering, Faculty of Engineering, Thammasat University. She received her Bachelor of Engineering in Civil Engineering (Honors) from Khonkaen University, Thailand. She received her Master of Engineering in Water Resources Engineering from the Asian Institute of Technology, Thailand. She then extended her research on the application of remote sensing and gained another Master's degree in Remote Sensing from GDTA, Toulouse, France. Later, she received her Ph.D. in Mechanical and Civil Engineering from the University of Montpellier II, France. Her research interests concern the rainfall forecast model and water resources management and planning.

Dr. Nkrintra Singhrattna is a Civil Engineer of the Structural and System Engineering Bureau, Department of Public Works and Town & Country Planning, Thailand. She received her Bachelor of Engineering (Civil Engineering) from Chiang Mai University, Thailand. She got her Master of Science degree in Water Engineering and Management from the University of Colorado at Boulder, USA. Later, she received her Ph.D. in Water Engineering and Management from the Asian Institute of Technology.