Annual Rainfall Model by Using Machine Learning Techniques for Agricultural Adjustment

Dulyawit Prangchumpol and Pijitra Jomsri
Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand
Email: {Dulyawit.pr, Pijitra.jo}@ssru.ac.th

***Abstract*—The change of weather conditions is considered as the major problem which is delicate for populations particularly the developing countries such as Thailand, e.g., the area in the southern region of Thailand, a case study of Andaman seaside which is the area most affected by the change of weather conditions compared with others. This research purposes to develop the model for rainfall forecasting for agricultural adjustment in the areas located at Andaman seaside using Machine Learning Technique to monitor the impact of the change of weather conditions such as the volume of rainfall in each year affecting to agricultural sector by studying the 30-year regressive data in order to estimate the coefficient of the model. The analysis results indicated that the rainfall volume of the areas located at the Andaman seaside of southern region tended to decrease which was resulted from the change of weather conditions where the model was able to provide the awareness of the impact caused by the change of weather to the agriculturists to prepare the proper supporting agricultural plans.**

***Index Terms*—annual rainfall, machine learning, agricultural**

I. Introduction

Currently, the irregular variation of weather conditions affects agricultural activities and causes natural disasters in several areas. In this regard, a factor affecting to the change of weather conditions is the greenhouse effect increasing the world's temperature, the increasingly and rapidly melting of the ice at the North Pole, resulting in the change of the seawater's temperature, irregular warm and cold currents, causing the storms and other types of natural disasters. For Thailand, particularly the Andaman seaside of the southern region consisting of 6 provinces, i.e., Ranong, Phang-nga, Phuket, Krabi, Trang, and Satun, of which their areas are adjacent to the sea, will be affected more than other regions. Due to the weather conditions of the southern region is tropical monsoon climate causing the rain throughout the year. Currently, the areas in the southern region have encountered floods every year and this affects to agricultural products and agricultural production value in the group of provinces adjacent to Andaman seaside in the southern region. The agricultural sector is the primary factor playing important roles in the economy of such a group in part of the agricultural products, i.e., rubber, oil palm. Hence, forecasting the events to occur in the future in order to prepare the protection for the agriculturists or the way to reduce or relieve the severity of impacts from natural disasters is greatly important. This research presents annual rainfall model by using machine learning techniques for agricultural adjustment and predicting the level of climate change in the southern Andaman coast of Thailand.

A literature review is presented in Section II. The research methodology is described in Section III. Section IV deals with the explanation of the IV, evaluating the results while the paper ends with conclusions and suggestions for possible future research specified in Section V.

II. Literature Review

In recent years, the use of machine learning algorithm for rainfall forecasting has considered as [1]-[7]. Some researcher combine model with deep learning algorithms [8] and other use regression model for predict land slide and debris after rainfall [9]. Despite the success of machine learning in a variety of tasks, applications to the problem of weather forecasting has been limited. Most work in weather forecasting to date rely on the use of generative approaches, where then weather systems are simulated via numerical methods [10]-[12], or rely on time-series analysis such as ARIMA models. The research in the literature shows that Autoregressive

In statistical regression analysis, which is used to know how much uncertainty is present in a curve that fits the data observed with random errors. It is apparent that effective time series forecasting relies upon proper model fitting. An appropriate consideration should be given to fitting a satisfactory model to the underlying time series.

In the literature shows that Autoregressive Moving Average (ARMA) models provide analysis of time series as a stationary stochastic process in terms of two polynomials one for Autoregression and second for moving average [13]. Autoregressive Integrated Moving Average (ARIMA) models and the Box-Jenkins methodology became highly popular in the 1970s among academics. The traditional approaches to time series prediction, such as the ARIMA or Box Jenkins [14]-[18] undertake the time series as generated from linear

Manuscript received January 8, 2020; revised June 18, 2020.

© 2020 J. Adv. Inf. Technol.
doi: 10.12720/jait.11.3.161-165

methods. However, they may be inappropriate if the underlying mechanism is nonlinear. In fact, the real world systems are often nonlinear. A pretty successful extension of the ARIMA model is the Seasonal ARIMA (SARIMA) [19]. The Seasonality is considered to understand the structure of time series if there exist repeated patterns over known, fixed periods of time within the data set. The restriction of these models is the pre-assumption of the time series in linear practice which is not suitable in real world scenarios.

III. Research Methodology

In general, climate models can be divided into 2 types: Global Circulation/Climate Model (GCM) and Regional Climate Model (RCM). The GCM model is a complex model that links the processes and relationships of the atmosphere, ground, ocean, and human activities. In general, the cell grid size is rough (about 100 km or more), which is not detailed enough to be used in regional studies. For this reason, it is necessary to use the RCM model to help with the results from GCM as the initial conditions. RCM has the capability to support simulations with grid cell resolution of up to 60 km or less (such as 10-20 km) and has developed many types of this model. RCM techniques is a high-resolution climate model in a limited area with downscaling which is divided into 3 types, statistic, dynamic, and statistic and dynamic downscaling. Because it is able to determine the area details that affect the local climate forecasts in greater detail, the dynamic downscaling is widely used. Using dynamic downscaling to create future climate data sets according to established climate scenarios, it must be used Global Climate Models (GCMs). Therefore, the accuracy of the RCMs model depends on the accuracy of the GCMs model. Therefore, the results from GCMs will be created as initial conditions and boundary conditions in RCMs. There are more details that must be considered such as border characteristics, topographical features, ground utilization, and types of ground cover, etc. With the same equations and conditions in dynamics and physics as those used in the GCMs model, the disadvantage of dynamic downscale is the waste of computer resources that will be used in calculations. The higher the calculation, the longer the time used for each calculation. Errors that result from calculation results in the GCMs model with large calculation grids prevent RCMs from displaying the characteristics of the calculation space sufficiently and there is a lot of data from GCMs when calculating RCMs each time. In addition, the Providing Regional Climate for Impact Studies model (PRECIS) is also applied. The PRECIS aims to create a model that can be used in all areas of the world. It can be calculated on a high-performance personal computer to meet the needs of countries that wish to study the climate change in their region, with emphasis on the development of convenience and impressions to users.

In the past, when we considered predictions the use of time series, a popular statistical technique, will be used. However, today the technology has improved and data science has made the forecasting to be more detailed by using machine learning to predict together with the use of regression algorithms. For regression it is to find the relation of various variables. Whether it's the past itself or other variables which do not limit the number of variables or time series format. There are many regression methods to choose from, such as Multiple Linear Regression (MLR). This method is the most well-known that finding the linear relationship of many variables. However, the limitation of the MLR is that it is a linear relationship only. When the problem is not linear or more complex, MLR cannot be used to solve problems. For this reason, we use more sophisticated algorithms such as Artificial Neural Network, Multivariate Adaptive Regression Splines, Gradient Boosting, Extra Tree, Random Forests etc. Each algorithm has a different description, source, and internal adjustment method. Therefore, is the origin of Machine Learning for Forecasting.

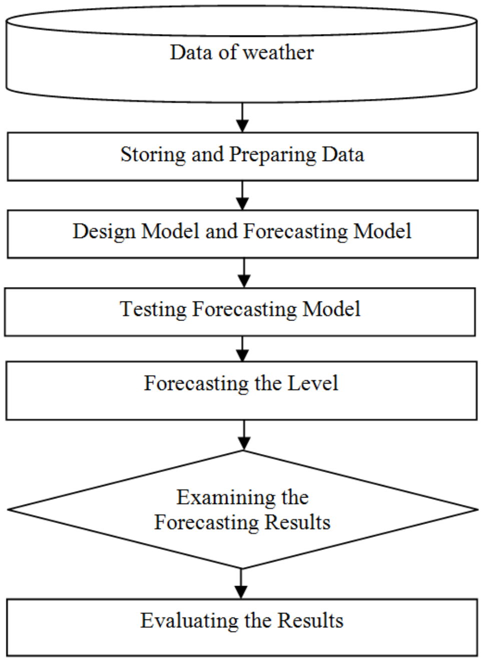

Figure 1. Framework of annual rainfall model by using machine learning techniques for agricultural adjustment

This research consisted of these 6 major processes. The first is data collection and preparation. For the next step we design the models, and the third process is testing the model. Forecasting climate change in southern Thailand used to apply after that. When the forecasting model is finished we've to check the results. The last process is evaluating results from finding relationships and checking predictions and conclusions as show in Fig. 1.

A. *Storing, Collecting, and Preparing the Information of Weather Conditions*

The 30-year consecutive data in relation to the weather conditions of the areas adjacent to the Andaman seaside of the southern region consisting of 6 provinces, i.e., Ranong, Phang-nga, Phuket, Krabi, Trang, and Satun was collected in order to analyze the relationship and forecast the level of the change of weather conditions. The data collected consisted of the name of the province, year, month, average temperature, maximal and minimal temperature, maximal atmospheric pressure, and total rainfall volume and maximal rainfall volume. The map area was show in Fig. 2.

Figure 2. The area of Andaman seaside of the southern for rainfall forecasting

B. Design Model and Determination of Forecasting Model

The researcher employed the Machine Learning Technique which is suitable for the weather conditions constantly changed and affected by the influence of the related surrounding variables. Therefore, the current weather forecasting employs Regressive Machine Learning to create the forecasting model. The analysis is conducted by using the Time-Series Model and meteorological data of Thailand by emphasizing the average rainfall volume changed in each year which its nature is the quantitative data able to be explained by polynomial form in (1).

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_p x^p + e \qquad (1)$$

when β is Regression coefficient
e is error
x is Factor for rainfall
y is Rainfall value testing forecasting model

The results calculated from the model were compared with the measured data. If the results are ranged in the acceptable scope, this model would be applied to forecast the events in the future. The current data would be input to the model and the statistics used in revising the values would be employed for the forecasting.

C. Forecasting the Level of the Change of Weather Conditions in the Areas of Southern Thailand

Forecasting the level of the change of weather conditions in the areas of Southern Thailand employed the results from data analysis to forecast the situations.

D. Examining the Forecasting Results

The forecasting results were compared with the actual events. If the data used to determine the relationships are identical to the forecasts, then the forecasting results are deemed accurate. The accuracy is measured in the percentage of accuracy of the forecasts. In addition, the data was divided into two sets, learning data and testing data, and the mean square error of the model was examined.

E. Evaluating the Results from the Relationships and Examination of Forecasting Results

The results obtained from the determination of relationships and forecasting examination were evaluated by measuring from the specified efficiency and reliability including testing the significance of the rules by the statistical theories.

IV. Evaluating the Results

According to the results of the study in the model for average rainfall volume forecasting for the agricultural adjustment in the areas adjacent to Andaman seaside of the southern region using Regressive Machine Learning and comparing the relationships of the two models, linear and quadratic model, the coefficients of multiple correlations appeared in Table I indicated that the quadratic model yield such coefficient greater than the linear model, then the quadratic model was subsequently studied. Table II indicated the regression analysis in order to forecast the average rainfall volume using T-Test which is the basic technique used to test the statistical hypotheses in order to compare the mean from the two groups of independent samples and the P-Value (probability value) was used for the hypothesis test.

TABLE I. Correlation Coefficient

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
Linear	.656	.431	.427	.224
Quadratic	.789	.623	.618	44.793

TABLE II. Type Sizes for Camera-Ready Papers

Value	B	Std. Error	β	t	Sig
YEAR	-54.50	6.25	-37.86	-8.72	0.00
YEAR ** 2	.013	.002	37.18	8.56	0.00
(Con stant)	57802.71	6486.64		8.911	0.00

The analysis of average rainfall each year using the Quadratic model, which has found a relationship. With the correlation coefficient of 78.9 percent and able to predict the average rainfall each year 62.3 percent

With statistical significance at the level of .001 and forecasting average rainfall for each year. Values of β_1=54.5079 and β_2=0.0129 follow in (2).

$$y = 57802.71 + -54.5079 * x + 0.0129 * x^2 \qquad (2)$$

© 2020 J. Adv. Inf. Technol.

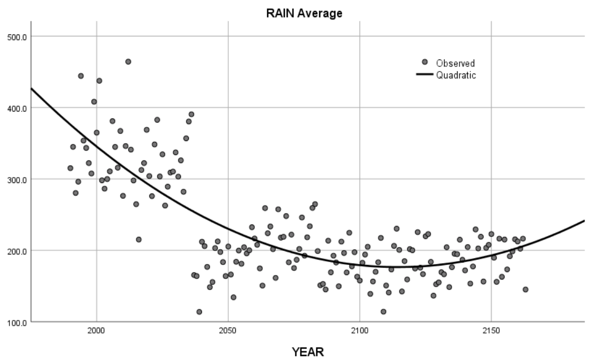

Figure 3. Forecasting average rainfall / ml

Fig. 3 shows the forecasting trend of average rainfall during the period 1984-2016 from the forecasting trend using mechanical learning. Found that the average amount of rainfall each year tends to decrease This may affect agriculture because some plants need to have enough rainfall, such as rubber, requiring rainfall not less than 1,255 millimeters per year and must have 120-150 days of rain per year.

V. Conclusion

This study in the model for average rainfall volume forecasting for the agricultural adjustment in the areas adjacent to the Andaman seaside of the southern region employed the Machine Learning Technique and collecting the data in relation to the 30-year regressive rainfall volume in order to study the tendency of the future rainfall volume. These factors affected the agricultural variations and the change of weather conditions in the seaside area of Southern Thailand.

It was found from the results that the future rainfall volume forecasted by the model tended to be reduced the annual rainfall volume would be less than the past. In this regard, for applying the results in the climate, the user should consider the contexts of weather conditions by studying the collective data of weather conditions of each period of time of data collection for a long time to build the reliability of the model including considering the changes by comparing with the conditions found in the base years resulted from the simulation of the model, e.g., considering the weather conditions in the future, collectively in the range of 5-10 years, or considering the change of weather conditions in each year, or even each month, or considering the number of the years having the levels of weather conditions lower or higher than the mean. In addition, the result of this research may be useful for water allocation in order to prepare the data for adjustment towards the world's climate change and agricultural planning for the best interest.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

D. Prangchumpol and P. Jomsri conceived of the presented idea. D. Prangchumpol developed the theory and performed the computations, and verified the analytical methods. P. Jomsri encouraged D. Prangchumpol to investigate theoretical calculations. Both authors discussed the results and contributed to the final version of the manuscript.

Acknowledgment

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

References

[1] R. C. Deo and M. Şahin, "Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in Eastern Australia," *Atmospheric Research*, vol. 153, pp. 512-525, 2015.

[2] R. Hashim, *et al.*, "Selection of meteorological parameters afecting rainfall estimation using neuro-fuzzy computing methodology," *Atmospheric Research*, vol. 171, pp. 21-30, 2016.

[3] P. Nayak, K. Sudheer, and K. Ramasastri, "Fuzzy computing based rainfall-runof model for real time food forecasting," *Hydrological Processes*, vol. 19, pp. 955-968, 2005.

[4] P. C. Nayak, K. Sudheer, D. Rangan, and K. Ramasastri, "A neuro-fuzzy computing technique for modeling hydrological time series," *Journal of Hydrology*, vol. 291, pp. 52-66, 2004.

[5] W. Sun, S. Shan, C. Zhang, P. Ge, and L. Tao, "In Artificial Intelligence and Computational Intelligence (AICI)," presented at the IEEE International Conference, 2010, pp. 252-256.

[6] R. Teschl, W. L. Randeu, and F. Teschl, "Improving weather radar estimates of rainfall using feed-forward neural networks," *Neural Networks*, vol. 20, pp. 519-527, 2007.

[7] M. N. French, W. F. Krajewski, and R. Cuykendall, "Rainfall forecasting in space and time using a neural network," *Journal of Hydrology*, vol. 137, pp. 1-31, 1992.

[8] M. Yen, *et al.*, "Application of the deep learning for the prediction of rainfall in Southern Taiwan," *Scientific Reports*, vol. 9, p. 12774, 2019.

[9] C. Chen, "Event-based rainfall warning regression model for landslide and debris flow issuing," *Environ. Earth Sci.*, vol. 79, 2020.

[10] L. F. Richardson, *Weather Prediction by Numerical Process*, Cambridge University Press, 2007.

[11] G. Marchuk, *Numerical Methods in Weather Prediction*, Elsevier, 2012.

[12] V. M. Krasnopolsky and M. S. Fox-Rabinovitz, "Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction," *Neural Networks*, vol. 19, no. 2, pp. 122-134, 2006.

[13] B. Choi, *ARMA Model Identification*, Springer Science & Business Media, 2012.

[14] G. E. Box, G. M. Jenkins, G.C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976.

[15] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.

[16] C. Brooks, "Univariate time series modelling and forecasting," in *Introductory Econometrics for Finance*, 2nd ed., Cambridge University Press, 2008.

[17] J. H. Cochrane, "Time series for macroeconomics and finance," manuscript, University of Chicago, 2005.

[18] K. W. Hipel and A. I. McLeod, *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier, 1994, vol. 45.

[19] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664-672, 2003.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

Dulyawit Prangchumpol was born in Thailand in 1980. He received his B.Sc. in statistic from Thamasat University, and M.S. in computer science from Slipakorn University. He received his doctorate degree from Chulalongkorn University in 2016 in the field of computer science and information technology. He works as a lecturer in major information technology in Faculty of Science and Technology, Suan Sunandha Rajabhat University. His researches are in the areas of virtualization system, network system, data analysis, and data mining.

Pijitra Jomsri was born in Thailand in 1981. She received her B.Sc. in statistic degree from Thammasat University, and M.S. in computer science from Slipakorn University. She received her doctorate degree from Chulalongkorn University in 2013 in the field of computer science and information technology. She is a lecturer in major information technology at Suan Sunandha Rajabhat University. Her research work is in the areas of virtual system, information retrieval, web technology and network system.