

Applying Principle Component Analysis to Explore Climatic Variability in Northern of Thailand

D. Prangchumpol

Abstract—Climate change is a substantial burden for most populations, in particular those in developing countries like Thailand, and specifically north western Thailand. Being largely dependent upon agriculture, this area is hard hit by the climate change. This study is to identify factors according to the climatic variability by creating a resilience index derived from principle component analysis, and then through the application of structural equation modeling utilizing data obtained from the Meteorological Department of Thailand during 1982-2014. In addition, principle component is applied to find the data. Correlation coefficients of up to 0.67 were obtained between the individual principal component loadings. The study determined that the temperature, air volume, and rainfall are highly affected by climatic variability. We further establish a relationship between climatic variables, and forecast the level of climate change in northern Thailand.

Index Terms—Climatic variability, principal component analysis, explore climatic.

I. INTRODUCTION

Recently, changing climate affects have affected farmer livelihoods due to flooding. One of the factors is due to greenhouse effect phenomenon causing global temperature to rise, which leads to the melt of polar ice, a cycle that accelerates over time. Those situations impact on the sea temperatures, as well as cold and warm currents causing abnormal flows. As a result, it causes heavy storms and other forms of natural disasters. All of which are caused by human actions, trying to make progress on the materials, but lack of consideration of the impact on the environment that follows.

In recent years, changes in the nature and the environment create an important phenomenon, called “El Nino and La Nina”. Both phenomena are caused by the abnormal flow of warm and cold currents. Moreover, the changing in seasons, which was the main factor in the dynamics between the atmosphere and ocean surface, influences the variation in seasons in Northeast Asia, otherwise known as monsoon season. Such a phenomenon is expected to affect the variability of the climate more intensely. In Thailand, especially in the North, is geographically characterized by several mountain ranges. Since the Northern climate is a tropical monsoon climate, rain is common throughout the year. These areas are now flooded every year and are likely to occur well into the future. Preparations to protect people from these

natural disasters becomes paramount.

Here, we investigate climate change and its effects in Northern Thailand. This is achieved by applying PCA to be used as a variable in the study and to cope with the effects of natural disasters quickly and more efficiently.

II. RELATED WORK

The impacts of climate change are heterogeneous across a diverse range of geo- political scales as it is perceived that the risks are generally believed to be more acute in developing countries because they rely heavily on climate-sensitive sectors, such as agriculture and fisheries, and have low gross domestic products, high level of poverty, low level of education, and limited human, institutional, economic, technical, and financial capacities [1]-[4]. The vulnerability of countries and societies to the effects of climate change depends not only on the magnitude of climatic stress but also on the sensitivity and capacity of affected societies to adapt to or cope with such stress. Therefore, the vulnerability is the degree to which a system is susceptible or unable to cope with the adverse effects of climate change, including climate variability and extremes. In this regard, vulnerability is a function of the character, magnitude, and rate of climate variation to which a system is exposed, its sensitivity, and its adaptive capacity [5].

A number of climate change impact studies have been conducted in many countries in specific sectors such as water resources, agriculture, health, coastal zones, and forestry by using impact models and to a lesser extent socioeconomic analyses [6]-[8]. Global recommendations for Africa calls for an integrated assessment approach for a vulnerability study, at the national scale and local level [9]. From the perspective of rural farm households, an analysis of vulnerability to climate change is needed at the level that would specifically address specific geographic location so that the smallholders will get a lesson to tackle climate change problems [10]. In Thailand, climate change has led to the increase in frequency and severity of natural disasters related to high temperatures and heavy rains. From 1951 until 2007, records show the average highest and lowest temperatures to be wider than in the past. When considering in 2007, it showed that the average highest temperature was lower, and the area, compared from 1978 to 2007, was also smaller.

When reviewing the related studies regarding to the weather forecast in Northern Thailand, Thachanat Bhatrastapongul [11] studies the relationship of IOD Phenomenon that affected the variation of rainfall in Thailand.

Manuscript received May 4, 2017; revised July 12, 2017.

D. Prangchumpol is with the Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand (e-mail: Dulyawit.pr@ssru.ac.th, dulyawit@gmail.com).

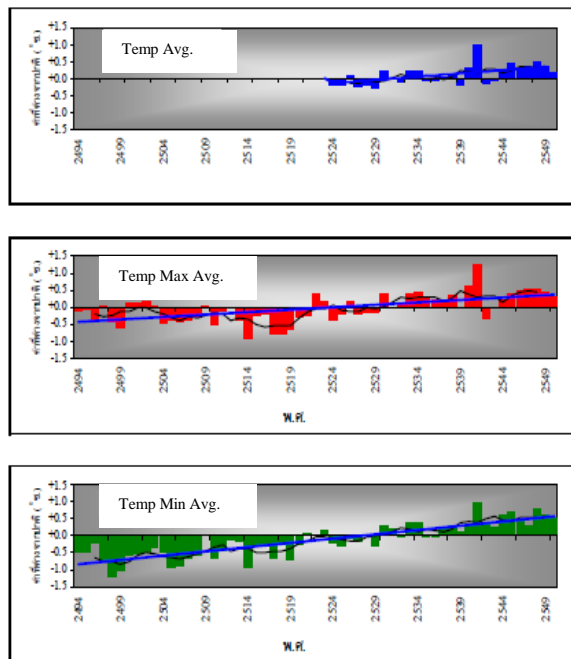


Fig. 1. Average temperature, average highest temperature and average lowest temperature seemed to wider from 1951 to 2007.

When using Principal Components Analysis to study the relationship of annual rainfall and seasonal rainfall with DMI by using the information of rainfall from 80 weather stations of the Thai meteorological Department from 1976 to 2008, it showed that IOD caused an increase in annual rainfall. However, rainfall in the dry season appeared to decrease. Zahoor Jan [12] applied Sea Surface Temperature (SST) to study seasonal climate changes. The statistic and mathematic

models were used to predict the weather. Such experiments have developed a system that used weather information in each region including rainfall, wind speed, dew point, temperature, and data mining.

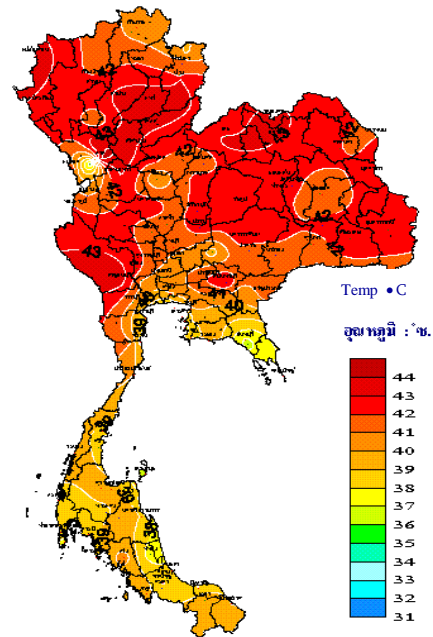


Fig. 2. Average highest temperature in Thailand.

stn_name	year	month	meantemp	A.meantemp	MEANMAXTEMP	A.MEANMAXTEMP	MaxRain24hrs	A.MaxRain24hrs	SUM	Eino
Phetchaburi*	1981	1	24.7	1	30.1	4	0	0	5	1
Phetchaburi*	1981	2	26.8	2	31.3	4	0	0	6	1
Phetchaburi*	1981	3	28.3	3	32.3	5	1.3	1	7	2
Phetchaburi*	1981	4	29.2	3	33.4	1	10.5	2	2	0
Phetchaburi*	1981	5	29.3	3	33.1	2	24.4	2	3	0
Phetchaburi*	1981	6	29.1	3	32.5	2	24.4	2	3	0
Phetchaburi*	1981	7	29.1	3	33	5	8.3	2	6	0
Phetchaburi*	1981	8	28.5	3	31.9	4	23.7	4	3	0
Phetchaburi*	1981	9	28.5	2	32.2	2	31.5	0	4	0

Fig. 3. Example of data sets.

This study applied K-Nearest Neighboring (KNN) to classify the weather information at a specific time, and to forecast the weather in advance for one month. The effect of ENSO Phenomenon on climate changes in Thailand showed that the average highest temperature appeared to increase due to the El Nino phenomenon [13]. It also found that the amount of annual rainfall appeared to decrease during El Nino [14]. The effects by La Nina on temperature and rainfall were the opposite, with lower temperatures [13] and higher rainfall [14]. In addition to these effects on temperature and rainfall, ENSO phenomenon also influenced the formation and movement of tropical cyclones in the western north Pacific and South China Sea. The El Nino phenomenon did not facilitate the formation and development of a tropical cyclone in the area. Wang and Chan [15] and Michael and Steinbach

studied these weather patterns and found that through Eigenvalue data mining [16], that they could forecast the weather worldwide. With the use of PCA and SVD data mining, they categorized ocean data into “Clusters”. The information obtained from the study was used for analysis or prediction including the results of previous analysis until today.

The results showed that this method was accurate, and more complete in the weather forecasts. In addition, Michael Steinbach analyzed climate/ocean data, data mining temperature, humidity, pressure, rainfall and time difference parameters or “Clusters” also called Shared Nearest Neighbor (SNN). The model can be applied to all the oceans on Earth due to the powerful of the correlation analysis.

III. DATA AND METHODOLOGY

This study was conducted in Northern Thailand. This area is affected by the climate change more than other areas as it is affected by monsoons and rain throughout the year. In this research, we focus on three climatic variables which we have studied as explanatory variables. These are Average highest temperature (Aht), Air volume (AV) and Highest rainfall (Hr). Data was obtained from The Meteorological Department of Thailand during 1982-2013. Examples of data sets are shown below in Fig. 3.

In this study we explored the climatic variables and/or the climatic factors (structure) which may influence for climate change in Northern Thailand. To find the climatic structure we used the Principle Component Analysis (PCA) method. Will et.al discussed that the PCA technique is recognized as a reduction and extraction for dimensionality of the data, and rated the variation presented in the original data sets [17]. This is one of the ways of identifying patterns in the data. It is difficult to find patterns in the data so for this purpose we use PCA which highlights the similarities and differences in the data. The pattern in the data can be found by “squeezing” the data. In other words by suppressing the dimensionality of the data by avoiding missing values. Edwards discusses that the purpose of applying the PCA is to reduce the manifest variables, in this way reducing the set of components [18]. The new components are called PC1, PC2, PC3... and so on, (for the first, second and third principal components), which are independent and decrease the amount of variance from the original data set. PC1 (the first component) captures most of the variance, PC2 captures the second most of the variance and so on until all the variance is accounted for, in this way leaving very little data for further study.

The components have been treated as climatic factors or climatic structures. The Principle component method transforms what has been observed correlated variables into uncorrelated variables which are linear combinations of observed data. The required condition in which to apply this technique is correlation/covariance which can then be defined. In this way, we can apply the PCA without issue. For the purpose of statistical analysis, PCA and other statistical models have been utilized to find the relationship between DF and climatic parameters.

IV. PRINCIPAL COMPONENT ANALYSIS/FACTOR ANALYSIS

Use of the Principal Component Analysis (PCA) is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables. PC provides information on the most meaningful parameters, which describes a whole data set affording data reduction with minimum loss of original information [19]. The principal component (PC) can be expressed as:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \dots + a_{im}x_{mj} \quad (1)$$

where z is the component score, a is the component loading, x is the measured value of variable, i is the component number, j is the sample number and m is the total number of variables.

Factor analysis (FA) is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. PC is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables [19]. PCA of the normalized variables was performed to extract significant PCs and to further reduce the contribution of variables with minor significance; these PCs were subjected to varimax rotation (raw) generating varifactors [20], [21]. As a result, a small number of factors will usually account for approximately the same amount of information as do the much larger set of original observations. The FA can be expressed as:

$$Z_{ji} = a_{j1}f_{1i} + a_{j2}f_{2i} + a_{j3}f_{3i} + \dots + a_{jm}f_{mi} + e_{ji} \quad (2)$$

where z is the measured variable, a is the factor loading, f is the factor score, e is the residual term accounting for errors or other source of variation, i is the sample number and m is the total number of factors.

V. RESULTS AND DISCUSSION

The following is the table constructed to determine the correlation between monthly climatic change and climatic variables for the period 1982-2014. In this section correlation between all pairs of variables were studied. Correlation coefficients are computed for monthly data and shown in Table I. It appears that the highest correlation obtained was found between Air volume (AV) and average highest temperature (Aht) with 0.67. This means that for average highest temperature monthly, the environment was dependent upon Air volume.

TABLE I: CORRELATION AMONG MONTHLY CLIMATE

Variable	Aht	Av	Hr
Aht	1	0.67**	0.46**
Av	0.67**	1	0.416**
Hr	0.46**	0.416**	1

**: significance at $\alpha = 0.01$

VI. CONCLUSION

The Correlation coefficients computed for monthly data revealed that the highest correlation is between Average highest temperature and Highest Rainfall. Average highest temperature (Aht), Air volume (AV) and Highest Rainfall (Hr). Although the principle component analysis/factor analysis did not result in a significant data reduction, it allowed for the extraction and identification of the factors/sources responsible for variations in climate change. This shows that for the average highest temperature per month, the environment is dependent on Air volume. In this study we may conclude that relative Air volume and average highest temperature have some impact on climate. If the Air volume recorded increases from previous day, the temperature was observed to increase in the following days. Further research will investigate regional climate model (RCM) and data mining.

ACKNOWLEDGMENT

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

REFERENCES

- [1] A. Ignatius, "Spatial vulnerability of rural house-holds to CC in Nigeria: Implications for internal security," Lagos, 2012.
- [2] United Nations Framework Convention on CC, "CC: Impacts, vulnerabilities and adaptation in developing countries," CC Secretariat, Bonn, 2007.
- [3] B. L. Preston, T. Smith, C. Brooke, R. Gorddard, T. Mea-sham, and G. Withycombe, "Mapping CC vul-nerability in the Sydney coastal councils group," Sydney Coastal Councils Group by the CSIRO Climate Adapta-tion Flagship Canberra, Sydney, 2008.
- [4] T. Utarasakul, "Climate change effects on fruit security in amphawa, samut songkram, Thailand," *International Congress of Environmental Research*, Nov 22-24, 2012, Kuala Terengganu, Malaysia.
- [5] Intergovernmental Panel on CC, "CC 2001: Impacts, adaptation and vulnerability," in *Proc. 3rd Assessment Report of the Intergovernmental Panel on CC, Adaptation to CC in the Context of Sustainable Development and Equity, Chapter 18*, Cambridge University Press, Cambridge, 877-912, 2001.
- [6] T. Deressa, R. M. Hassan, and C. Ringler, "Meas-uring ethiopian farmers' vulnerability to climate change across regional states," IFPRI Discussion paper No. 806, Washington, 2008.
- [7] L. Pearson and J. Langridge, "CC vulnerability assessment: Review of agricultural productivity," CSIRO Climate Adaptation Flagship Working Paper No. 1, 2008.
- [8] P. A. Odjugo, "General overview of CC impacts in Nigeria," *Journal of Human Ecology*, vol. 29, pp. 47-55, 2010.
- [9] United Nations Framework Convention on CC, "Background paper on impacts, vulnerability and adapta-tion to CC in Africa," African Workshop on Adaptation Implementation of Decision 1/CP.10 of the UNFCCC Convention, Accra, 21-23 September 2006.
- [10] R. J. T. Klein, "Approaches, methods and tools for cc impact, vulnerability and adaptation assessment," Session Workshop on Impacts of, and Vulnerability and Adaptation to CC, 21st Session of the UNFCCC Subsidiary Body for Scientific and Technical Advice, Buenos Ai-res, 8 December 2004.
- [11] T. Bhattarapongul, "Thailand's climate variability associated with some oceanographic anomalies," Thailand Research Fund (TRF).
- [12] J. Zahoor, "Seasonal to inter-annual climate prediction using data mining KNN technique," Computer and Emerging Sciences/FAST-National University, 2008.
- [13] A. Limsakul, and J. Goes, "Empirical evidence for interannual and longer period variability in Thailand surface air temperatures," *Atmospheric Research*, vol. 87, pp. 89-102, 2008.
- [14] A. Limsakul, J. Goes, "Empirical evidence for international and longer period variability in Thailand surface air temperature," *Journal of Atmospheric Research*, vol. 87, pp. 89-102, 2008.
- [15] B. Wang and J.C.L. Chan, "How strong ENSO events affect tropical storm activity over the western North Pacific," *J. Climate*, 2002.
- [16] M. Steinbach, P.-N. Tan, V. Kumar, C. Potter, and S. Klooster, "Data mining for the discovery of ocean climate indices," in *Proc. the Mining Scientific Datasets Workshop*, 2nd Annual SIAM International Conference on Data Mining, April 2002.
- [17] T. Will, "Introduction to the singular value decomposition," Davidson College, 1999.
- [18] E. Jackson, "A user guide to principal components," *Wiley Series in Probability and Mathematical Statistics*, 1991.
- [19] B. Helena, R. Pardo, M. Vega, E. Barrado, J. M. Fernandez and L. Fernandez, "Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis," *Water Research*, 2000, pp. 807-816.
- [20] K. P. Singh, A. Malik, and S. Sinha, "Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: A case study," *Analytic aChimica Acta*, 2005, pp. 355-374.
- [21] G. Brumelis, L. Lapilla, O. Nikodemus, and G. Tabors, "Use of an artificial model of monitoring data to aid interpretation of principal component analysis," *Environmental Modelling & Software*, 2000, pp. 755-763.



D. Prangchumpol was born in Thailand in 1980. He received his B.Sc. in statistic from Thamassat 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.