Application of Principal Component Analysis Technique to the Socio-economic Factors in Quality of the Lower Songkhram Wetland, Thailand

Somphinith Muangthong[1*] and Nirat Yamoat[2]

Abstract

Data used for analysis are a variety such as query information in order to identify the main factors influencing the quality of the Lower Songkhram Wetland (LSW), Thailand. There are a total of 8,360 data from questionnaire which was based on 20 socio-economic factors. Questionnaires of 418 sets were administered at the LSW in 40 sub-districts. It used survey from December 2012 to January 2013. Data were analyzed by using the technique of Principal Component Analysis (PCA). The results show seven groups of factors that influence the wetland quality as follows (1) the communities' economy (2) changes in the economic system of the country, (3) the participation of the people (4) the awareness of people (5) the conservation of water (6) planning and coordination of community, and (7) the policy of the community. The principal component analysis found the important factors influencing the wetland quality, which could be guideline for the wetlands management.

Keywords : Lower Songkhram Wetland, Socio-economic factors, Principal Component Analysis, PCA

[1] Faculty of Engineering and Architecture, Rajamangala University of Technology Isan, Nakhonratchasima.

[2] Department of Civil Engineering and Environmental Technology, College of Industrial Technology, King Mongkut' s University of Technology North Bangkok.

[*] Corresponding author, E-mail: somphinith.mu@rmuti.ac.th Received 27 April 2017, Accepted 31 July 2017

1. Introduction

During the last Century, wetlands throughout the world have dramatically decreased, largely owing to human induced changes in original native lowland forest cover. This loss of lowland forest also has far reaching impact on the quality of wetlands. The Lower Songkhram Wetland (LSW) is a complex web of ecosystem habitat types [1]. This wetland possess high biological diversity and are known as the most productive area of the lower Mekong River Basin [2]. Nevertheless, during the last decade, significant degradation has occurred in the wetland critically threatening the habitat of some precious wildlife. Hence, the issue of wetland management and restoration receive more focus than before which poses significant potential for the analysis of wetlands quality and related factors influencing the wetlands in order to reasonably manage and restore the wetlands.

This paper aims to identify the factors influencing the quality of the LSW based on socio-economic factors. Thus, 20 variables were considered to determine the quality of the wetland. These variables were identified in several literature e.g. Deka et al. [3], Esty et al. [4], Brazner et al. [5], Donnelly et al. [6], He et al. [7], Wang et al. [8] and Hasyim [9]. A multivariate statistical technique called Principal Component Analysis (PCA) was used method in analysis of data matrix for drawing meaningful information [10]. PCA is normally used to understand the correlation structure of collected data and identify the most important factors that can contribute to the data analysis [11]. The relationship among a number of observed quantitative variables in PCA are represented in terms of a few underlying, independent variables called factors, which may not be directly measured or even measurable [10, 12].

2. Methods

2.1 Site Description

The LSW is located in the lower part of Mekong Basin in Northeast of Thailand, between 17º 20′ – 18º 10′N and 103º 30′ – 104º 30′E covering an area of 3,049 km^2 (Figure 1).

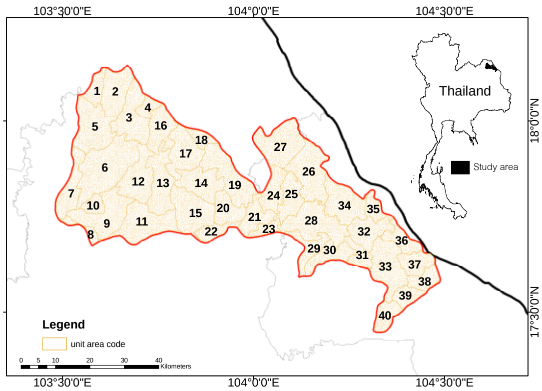

Fig. 1. The study area in Lower Songkhram River Basin, Northeast of Thailand

The wetland's local topography is dominated by flat marshland whose altitude lies in between 145–160 m.s.l. and a gentle gradient of about 1:30,000 [2]. The area has a tropical semi-humid climate, with three distinct seasons (rainy, winter, and summer). More than 90% of the total rainfall occurs within the six month (May-October).

LSW is very humid, with average humidity of 65% – 78%, depending on the season and time of the day. Based on data during 1981–2010 by Thailand Meteorological Department, the minimum and maximum temperatures were 19°C and 34°C respectively with a mean annual temperature of 27°C. The average annual evaporation was about 1,205 – 1,898 mm/yr.

2.2 Data Analysis

2.2.1 Sample Size

The factors influencing the quality of wetlands were collected directly from local people (100%), household survey using questionnaires. Households for questionnaire survey were selected on the basis of sampling size. This calculation of sample size is illustrated by the following equation.

$$N = \frac{n}{1+n(e)^2} \qquad (1)$$

Where N = sample size

n = household number, this study area equal to 70,801 households

e = confidence interval, expressed as decimal e.g. 0.05 for 5 percent

Thus, the sample size in the LSW is equal to:

$$N = \frac{70{,}801}{1+70{,}801(0.05)^2} = 397.75 \approx 398$$

The obtained sample size from above equation is then divided amongst the sub-unit areas (40 sub-districts) which was calculated using equation (2), thus, if decimal value of sub unit area more than 0.5 will be adjust to 1 household value. Therefore, the total numbers of samples in this study were 418 which covered 40 sub-districts in LSW.

$$\text{Number of sub-unit area} = \frac{m*100}{398} \qquad (2)$$

Where m = household number in each subdistrict

2.2.2 Questionnaire Design

A structured questionnaire was used to obtain the general information of household structure and factors influencing the quality of LSW. A simple questionnaire was developed, which asked respondents to report on the factors influencing the quality of the LSW, which covered approximately 3,050 sq.km in 40 sub-districts of 3 provinces, Northeastern of Thailand. The questionnaire was designed to be as short and simple as possible, it was translated into Thai language and pretested in the study area for adjustments. The questionnaire survey was conducted of about 418 households linked with the wetlands in the LSW during December 2012 to January 2013.

Figure 1 shows unit areas for questionnaire survey in the study area. The questionnaire survey data of some parts were adapted from Stak et al. [13], which have already been considered for global use as part of a suite of ecological indicators of the effectiveness of the convention. According to Lemly [14], to formulate the effective solutions, wetland problems must be considered at watershed, landscape, and ecosystem scales. Therefore, Table 1 show twenty variables from the question which are used for identifying the factors influencing the quality of the LSW.

Table 1 List of variables that influence the quality of the LSW

Variable	Effect	Symbol	Independent Variables
1	-	CE	The communities economy
2	-	PG	Population growth
3	-	IG	Intrusion of wetlands for agriculture
4	+	EP	Employment
5	-	UD	Urban development
6	-	ID	Industrial development
7	-	ES	Changes in the economic system of the country
8	+	PP	The participation of the people
9	+	ED	Education
10	+	AP	The awareness of people
11	-	NW	Use of natural wetland resources useless
12	+	CW	The conservation of water
13	-	SW	Settlement of the surrounding wetlands
14	-	FW	Fishing in wetlands
15	-	AG	Agricultural growth
16	+	PO	Planning and coordination of community
17	-	CA	Capturing animals in the wetlands
18	-	LC	Local community encroachment into wetlands
19	+	LP	Local policy
20	-	HO	The household occupation

2.2.3 Determining Factors Influencing the Health of Wetlands

This study was conducted in the wetlands at the LSW. Twenty variables were considered to explain the quality of wetland based on questionnaire survey. Principal Component Analysis (PCA) was used to identify factors influencing the quality of the LSW. For the purpose of describing the underlying factor structure, the Kaiser-Meyer-Olkin (KMO) value >0.5 was considered as first criteria to be satisfied for conducting factor analysis [11-12]. Moreover, communalities value >0.5 [15], and eigenvalues >1 were used to determine the number of components to be extracted for further analysis [11-12, 16]. Therefore, factor rotation using Varimax method with Kaiser Normalization was also used to facilitate interpretation by providing simpler factor structure with high loadings on each factor [11-12].

Highly weighted variables from each principal component (PC) were defined as the variables within the highest factor loading and retained in the data set [17]. The weighted variables were significantly correlated, and later one variable was considered redundant and was eliminated from the PC [15]. Among the significantly correlated variables remaining in the PC, the variable with the highest sum of correlation coefficients was chosen to be part of the principal components (PCs). Consequently, the retained variables having absolutely high factor loading were identified as factors that influence the quality of wetlands.

3. Result and Discussion

3.1 Factors Influencing the Wetland Quality

In order to reveal the full depth of the data, PCA was applied to analyze the efficient values. As a result, the adequacy of the data was checked on the basis of Kaiser-Meyer-Olkin (KMO) sampling adequacy test which suggests that the KMO value should be more than 0.5 to fulfill the adequacy of the data set [18]. Furthermore, the Bartlett's test of sphericity is accepted when it is significant at $p<0.05$. KMO value for this study is measured as 0.552 and the Bartlett's test of sphericity was highly significant ($p<0.001$), indicating that the present data has fulfilled the criteria for conducting PCA.

The PCA extracted seven PCs based on eigenvalue criteria >1. Table 2 shows the variables that were considered in PCA as well as their factor loadings and communalities. The communality of a variable is the proportion of the variance that is explained by the common factors. Communalities are inserted in the diagonal of the correlation matrix. The extraction was terminated as soon as the communality of a variable exceeded 1.0 [11]. Thus, components are the dimensions identified using PCA. All the extracted communalities are reasonably high and acceptable, as is shown in Table 2. Therefore, the PCA is considered to be acceptable after fulfilling the criteria of communalities value >0.5 as suggested by Halim et al. [15]. Only the variable that had the absolute value of highest factor loading in the PC was selected as the representative of the group (component) [15, 17].

The first seven PCs in the initial eigenvalues have an eigenvalue >1 with an overall cumulative variance of about 81% and the rest neglected, as shown in Table 3. According to Kaiser Criterion, only the first seven components are considered because of subsequent all eigenvalues would be >1.

Component loadings were used to measure correlation between variables and the components. A loading close to ±1 indicates a strong correlation between a variable and the component while a loading close to zero indicates a weak correlation [12]. Unrotated solutions of component loadings were not suitable for interpretation purposes since the variables generally tend to load on multiple components [12]. The components are rotated with the use of Varimax rotation which is a standard rotation method [11].

Population of the LSW has significantly increased during last decade. Thus, increasing populations led to degradation of wetlands health. There are many factors expected to influence the quality of wetland in the LSW. PCA is able to identify significant sources of threat and sources of the good quality of wetland inputs to LSW. The seven principal components (Table 3) are selected by examining the components which have relationship to the original variables but are not interrelated.

Table 2 Results of PCA of the 20 dominant variables comprising seven PCs

Description	Components							Communality
	1	2	3	4	5	6	7	
The communities economy	-0.89							0.81
Population growth	-0.82							0.85
Intrusion of wetlands for agriculture	-0.63							0.81
Employment	-0.56							0.71
Changes in the economic system of the country		-0.96						0.95
Industrial development		-0.85						0.95
Urban development		-0.56						0.88
The participation of the people			0.86					0.87
Education			0.58					0.70
The awareness of people				0.84				0.92
Use of natural wetland resources useless				-0.67				0.81
Fishing in wetlands				0.57				0.74
The conservation of water					0.86			0.86
Settlement of the surrounding wetlands					-0.74			0.74
Agricultural growth					-0.58			0.71
Local community encroachment into wetlands					-0.51			0.69
Planning and coordination of community						0.92		0.89
Capturing animals in the wetlands						-0.56		0.93
The household occupation						-0.56		0.81
Local policy							0.78	0.76

Table 3 Eigenvalues, variance and cumulative percentages for the first seven PCs

Description	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	2.79	2.62	2.48	1.52	1.48	1.33	1.28
Percentage of variance (%)	14.28	13.56	13.12	12.59	10.18	9.11	8.35
Cumulative percent (%)	14.28	27.84	40.96	53.55	63.73	72.84	81.19

The first component explains about 14% of the total variance of the 20 observed variables with strong negative loading of the communities economy, population growth, intrusion of wetlands for agriculture and employment. The second component incorporates changes in the economic system of the country factor with 13% of the total variance having strong negative loading of changes in the economic system of the country, industrial development and urban development. The third component consists of about 13% of the total variance with strong positive loading of the participation of the people and education variables.

The fourth component consists of about 12% of the total variance with strong positive loading on the awareness of people and fishing in wetlands, and the total variance with strong negative loading on use of natural wetland resources useless. The fifth component consists of about 10% of the total variance with strong positive loading of the conservation of water, and the total variance with strong negative loading on settlement of the surrounding wetlands, agricultural growth, and local community encroachment into wetlands. The sixth component consists of about 9% of the total variance with strong positive loading on planning and coordination of community, and the total variance with strong negative loading on capturing animals in the wetlands and the household occupation. The seventh component consists of about 8% of the total variance with strong positive loading of local policy variable. Lemly [14] described wetland problems must be considered at watershed scales.

Therefore, the factors are becoming more effective to protecting the quality of wetlands i.e. participation of the local community, local community awareness, and local policy. Hence factors from the participation of the local community group were dominated by the people within wetland areas as their profession which gives a better view of the wetland ecosystem, so they will have better knowledge about the changes and its impacts.

Increasing of the population trend is being affected to wetland quality in the LSW. The natural population trend has been interrupted within wetlands health by human activity such as the pollution by agriculture without knowledgebase that can lead to the degradation and destruction of wetlands. In addition, the policy makers and organization working should be aware of the immediate gains in education, income, and employment arising from wetland conversion.

4. Conclusion

This study focused on the identification of the factors influencing the quality of the LSW. Analysis indicated that the data collected from the site has fulfilled the criteria for conducting PCA. Based on PCA, seven factors comprised the principal component that have high factor loading (eigenvectors). Thus, seven factors are found to predominantly influence the quality of the LSW, the communities economy, changes in the economic system of the country, the participation of the people, the awareness of people, the conservation of water, planning and coordination of community, and the policy of the community.

The results of this study can provide a useful tool in assessing the wetland condition in the LSW based on questionnaire survey. However, the factors obtained in this study may only be applicable in the wetlands of the LSW. Extension of this methodological approach to other regions of the floodplain wetlands may be possible if reference wetlands are available to adjust the factors influencing local natural conditions. The knowledge of the causal factors influencing the wetland quality is a vital prerequisite in planning and executing an effective and holistic approach to address the issue of wetland degradation.

Furthermore, the findings of this study will also be useful for policy makers and organizations working on the conservation of wetland since they can focus their efforts more effectively.

5. Acknowledgement

The authors wish to thank the Faculty of Engineering, Rajamangala University of Technology Isan for providing financial assistance.

6. Reference

[1] Thai Baan Research. “Ecology and local history of the seasonally flooded forest in the Lower Songkhram Basin”, The World Conservation Union (IUCN), Available: www.mekongwet-lands.org/common/download/IUCNT_ThaiBaan Research.pdf, 12 December 2005.

[2] D. Blake and R. Pitakthepsombut, "Situation analysis: Lower Songkhram River Basin, Thailand", A publication of the Mekong Wetlands Biodiversity Conservation and Sustainable Use Programme, IUCN-2006-017, 2006.

[3] T.K. Deka, M.M. Goswami and M. Kakati, “Causes of fish depletion-A factor analysis approach”, World Fish Center Newsletter 28(1&2), 2005, pp. 37-42.

[4] D.C. Esty, M. Levy, T. Srebotnjak and A.D. Sherbinin, “Environmental sustainability index: Benchmarking national environmental stewardship” Available:www.yale.edu/esi/ESI2005_Main_Report.pdf, 23 October 2008.

[5] J.C. Brazner, N.P. Danz, G.J. Niemi, R.R. Regal, A.S. Trebitz and R.W. Howe, “Evaluation of geographic, geomorphic and human influences on Great Lakes wetland indicators: A multi-assemblage approach”, Ecological Indicators 7, 2007, pp. 610-635.

[6] A. Donnelly, M. Jones, T. O'Mahony and G. Byrne, “Selection environmental indicator for use in strategic environmental assessment”, Environmental Impact Assessment Review 27, 2007, pp.161-175.

[7] R. He, L. Song and J. Zhang, “On the evaluation of regional ecosystem health”, In: L. Ren, Q. Li, D. Zhang and J. Xia (Ed.), “Methodology in Hydrology IAHS Publish (311)”, IAHS Press, Wallingford, 2007.

[8] X. Wang, S. Xu and D. Liu, "Analysis of ecohydrological characteristics and factors affecting the Zhalong riparian wetland", In: L. Ren, Q. Li, D. Zhang and J. Xia (Ed.), "Methodology in hydrology (Vol. IAHS Publish no. 311, pp. 371-378)", IAHS Press, Wallingford, 2007.

[9] A.W. Hasyim, "Urban climate: Poor land use management as a factor on urban heat Island (UHI) a case of Klojen sub district in Malang city", Available: http://awhasyim.wordpress.com/2008/09/17/urban- climate- poor- land- use- management- as- a- factor- on- urban-heat-island- uhi- a- case- of- klojen- sub- district- In- malang- city/, 10 October, 2008.

[10] J.G. Kennen, L.J. Kauffman, M.A. Ayers, D.M. Wolock and S.J. Colarullo, "Use of an integrated flow model to estimate ecologically relevant hydrologic characteristics at stream biomonitoring sites", Ecological Modeling 211, 2008, pp. 57-76.

[11] J.E. Hair, C.B. William, J.B. Barry, E.A. Rolph and R.L. Tatham, "Multivariate data analysis", (6th Eds.), Pearson Printice Hall, New Jersy, 2006.

[12] N.O. Singh, S. Kumar, P.C. Mahanta, M.K. Panda and N.G. Singh, "Evaluation of water quality from Gaula river by factor analysis", Journal of Ecophysiology and Occupational Health 7, 2007, pp. 165-169.

[13] M. Stark, N. Davidson and S. Kouvelis, "A qualitative assessment of the status of Mediterranean wetland" 6th meeting of the Mediterranean wetlands committee: Tipaza, Algeria, 12-14 December 2004.

[14] A.D. Lemly, "Risk assessment as an environmental management tool: considerations for freshwater wetlands", Environmental Management 21(3), 1997, pp. 343-358.

[15] R. Halim, R.S. Clemente, J.K. Routray and R.P. Shrestha, "Integration of biophysical and socio-economic factors to assess soil erosion hazard in the upper Kaligarang watershed, Indonesia" Land degradation & development 18, 2007, pp. 453-469.

[16] J.J. Brejda, T.B. Moorman, D.L. Karlen and T.H. Dao, "Identification of regional soil quality factors and indicators: I. Central and southern high plains" Soil Science of America Journal 64, 2000, pp. 2115-2124.

[17] S.S. Andrews, D.L. Karlen and J.P. Mitchell, "A comparison of soil quality indexing methods for vegetable production systems in Northern California", Agriculture, Ecosystems and Environment 90, 2002, pp. 25-45.

[18] H.F. Kaiser, "An index of factorial simplicity", Psychometrika 39, 1974, pp. 31-36.