APPLICATION OF FACTOR ANALYSIS TO IDENTIFY RESOURCE CHARACTERISTICS IN RURAL AREA OF INDONESIA*

by Muhammad RIDHO**, Jawoto Sih SETYONO***, Qiang LI**** and Tomonori SUMI****

1. Introduction

It is necessary for every country to grasp the resource characteristics when development policies are making out and there is no exception to the developing countries. In Indonesia, an agricultural country, it is significant to identify the specific resource characteristics in rural area so as to fulfill adaptable development well. The rural area is under constant pressure from urban developments. Multi-purpose resources characteristics of rural area are becoming more and more important[1)]. Spatial pattern of land-use changes because complex interactions between physiographic and socio-economic factors. Human activity is a major factor in shaping the rural landscape, whereas the physical structure of a landscape often constrains land use[2)].

The objective of this study was to explain a methodology to identify resource characteristic in rural area. This methodology used data from many resources variable, including land-use, human resources and number of public facility. Statistical resulted from this study will get significant information that must provided by government and private sector. In this paper, we take Demak regency as a study area and apply the factor analysis method.

This article presents application of factor analysis to identify resource characteristics in rural area. We were trying to reduce many variables to detect structure in the relationships between one variable and others. Section 2 presents statistical method and data explanation in this study. Section 3 presents the results of the study about identifying resource characteristics. The conclusions and discussion of the study are presented in Section 4.

2. Method and Data

Many statistical methods are used to study the relation between independent and dependent variables. Factor analysis is different; it is used to study the patterns of relationship among many dependent variables, with the goal of discovering something about the nature of the independent variables that affect them, even though those independent variables were not measured directly. The inferred independent variables are called as factors[3)]. The main applications of factor analysis techniques are: (1) to reduce the number of variables and (2) to detect structure in the relationships between variables, that is to classify variables[4)]. In this paper factor analysis consists of few main steps. The first step is standardizes variable to synchronize the measurement of each variables. Second step involves calculating an initial correlation matrix to examine that the initial data are suitable for factor analysis calculation. Step third involves extracting factor. This step yields a matrix contained factor loading. Step fourth, factor loading is transformed, with the intention of making the matrix interpretable. Final step, involves calculation factor scores. These are the scores of each data set on each of extended factors.

We use data of Demak regency in Indonesia based on following reason: (1) Demak regency as a part of Indonesia, which included as developing countries. (2) Demak regency is an agricultural area, which has similar character between one rural area and others. These areas have 247 rural places and 34 characters that will be observe. (3) These areas are the buffer of two big cities, but have 3rd lowest local GDP in province scale. Thereby this area must prepare to develop later.

(1) Standardizes variables

Analysis was conducted from data collected in Demak Regency at 1995. Every variable has the different measurement based on characteristic. Standardizes variables needs to get same measurement between variables. We use standardizes data using equation 1. Basic matrix was made from 247 sample cases and 34 variables, where the rural area names are the case of analysis for this study, and values of rural area characteristic as a variable. Comrey & Lee describe a sample size (N) of 50 as very poor; N = 100 as poor; N = 200 as fair; N = 300 as good; N = 500 as very good and N = 1000 as excellent[5)]. Based on that statement, sample size in this paper included between fair and good category.

$$V_{ij}` = (V_{ij} / V_{i\,max}) * 1000 \qquad (1)$$

Where: $V_{ij}`$ is value of variable i at rural area j in standardizes measurement
V_{ij} is value of variable i at rural area j in original measurement
$V_{i\,max}$ is value of variable i at the highest value of rural area in original measurement

* Keywords: Rural area, Factor analysis, Resource characters
** Graduate School of Engineering, Kyushu University
(Hakozaki 6-10-1, Fukuoka, Japan, TEL 092-642-3275, FAX 092-642-3306)
*** Department of Regional and City Planning Engineering. Diponegoro University, Semarang, Indonesia
**** Member of JSCE, Graduate School of Engineering, Kyushu University

(2) Correlation and factor extraction

Calculating an initial correlation matrix needed to examine that the initial data are suitable for factor analysis calculation. We are aware of potential errors associated with this analytic strategy. Factor analysis has been criticized as unreliable method where the optimal statistical solution may be overly sensitive to the presence of certain variables. We anticipated this problem by using the principal components option in Statistica®. This method standardizes all variables to a mean of zero and standard deviation of one prior to analysis thereby minimizing problems associated with scaling. One rule of thumb to determine the number factors at principal components, called by Kaiser Criterion, explains that only factors with eigenvalues greater than 1.00 able to explain the observed variance. Based on the rules, five factors were generated.

(3) Factor rotation

In orthogonal rotation, we used varimax rotation option. This method rotated the axes to minimize the number of variable that have high loading on a factor. Rotated factor matrix showed five factors that available to interpreting. Only variable with loading factor of 0.5 or greater are considered in interpreting each factor[6]. Based on that statement, only 24 variables will have particular emphasis, others variables were dropped.

Significant clusters of variables, or factors are identified through optimally weighted linear combination of observed variables that maximize the amount of explained variance. The first factor explains the highest proportion of observed variance in dataset. The second factor accounts for the majority of variance not explained by factor 1, and so on. Factor loading represent the degree of correlation between an individual variable and given factor. Values range from –1 to +1, with a larger absolute value indicating a stronger contribution of a variable to that factor. Within a factor, a positive loading indicates a direct association with the factor, while a negative loading indicates an inverse association[7].

3. Results Analysis

(1) Factor Interpretation

Factor loadings and percentage of observed variance are presented in Table 1 and Figure 1. Factor 1 has an eigenvalue of 9.91, and explains 28.31%, highest value of the observed variance. We have defined this factor "potential facility service" since it loads positive factor loading for public facility service. Such as the presence regional market, high school, sub district hospital, etc. Factor 2 has two definitions. First definition is "paddy farm resource" since it loads positive factor for area of paddy farm. Second definition is "industrial and wheatland resource" since it loads negative factor loading for industry labor and wheatland area. Factor 2 has an eigenvalue of 3.68 and explains 10.52% of the observed variance. Factor 3 has an eigenvalue of 2.96 and explains 8.46% of the observed variance. We called "human resource for agricultural activity" as it loads positive loading factor for agricultural labor and agricultural labor per 1000 population. Factor 4 is labeled "industrial and urban activity" since it loads positive factor for industrial labor, service labor, population density and public bank. This factor has an eigenvalue of 2.44 and explains 6.97% of the observed variance.

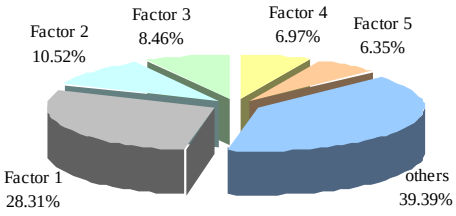

Figure 1: Percentage of observed variance based on eigenvalue

Table 1: Rotated factor loadings

Variable	Factor 1 potential of facility service (+)	Factor 2 paddy farm resource (+) industrial & wheatland res. (-	Factor 3 human resource for agricultural (+)	Factor 4 industrial & urban activity (+)	Factor 5 natural resource for brackish fishpond (-)
Number of population				0.620	
Population density					
Agricultural Labor			0.807		
Agricultural Labor per 1000 population			0.544		
Industrial Labor		-0.590		0.571	
Industrial Labor per 1000 population					
Service Labor				0.880	
Service Labor per 1000 population					
Total of Labor				0.635	
Total of Labor per 1000 population					
Paddy farm area		0.500			
Proportion of paddy farm area		0.772			
Wheatland area		-0.879			
Proportion of wheatland area		-0.771			
Brackish fishpond area					-0.910
Proportion of brackish fishpond area					-0.916
Forest area		-0.546			
Proportion of forest area		-0.553			
Hospital				0.931	
Subdistric hospital	0.745				
Medical clinic					
Family planning clinic	0.505				
Doctor	0.560			0.699	
Nurse				0.823	
Elementary school					
Junior high school	0.749				
High school	0.569				
Village market					
Regional market	0.749				
Village corporation store	0.822				
Village bank	0.836				
Regional bank				0.931	
Post office	0.726				
Bus station	0.553				

Factor 5 has an eigenvalue of 2.22 and explains 6.35%, the lowest of the observed than other factors. We defined this factor “natural resource for brackish fishpond”, since it loads negative loading factor for breadth of brackish fishpond and these proportion.

(2) Combining score factors

Final outputs of factor analysis stages are factor scores. These are scores calculated for each case reflecting each experimental unit's scores on the extracted factors, after any rotation have taken. There are 4 categories at combining score factor are (1) both of score factors have positive values, (2) first score factor has positive values but second score factor has negative values, (3) both of score factor have negative values, and (3) first score factor has negative values but second score factor has positive values.

Score factors combining based on interpreted loading factors to represents the better summary, which able to explain about resource character from each rural area. First, combining score factor 1 and 4 illustrated potentiality of service center activity, see Table 2 and Figure 2. Areas in category 1 have very high resources as service center. These areas have potentiality as service area and supported by urban and industrial activity.

Second, combining score factor 3 and 2 illustrated paddy farm and wheatland resource, see Table 3 and Figure 3. The most suitable areas for paddy farm were areas in category 1. These areas have enough agricultural human resources and supported by paddy farm field. These area spread in center and south region at study area. Next, the most suitable areas for wheatland were areas in category 2. These areas have enough agricultural human resources and supported by wheatland field. These areas spread in west and southwest region at study area.

Third, combining score factor 3 and 5 illustrated potentiality of brackish fishpond resource, see Table 4 and Figure 4. Areas in category 2 have high resources for brackish fishpond activity. These areas have human resources in agricultural activity and supported by brackish fishpond area. These areas spread in coastwise land at study area (northwest region).

Fourth, combining score factor 2 and 4 illustrated industrial activity resource, see Table 5 and Figure 5. The most suitable area for industrial activities was area in category 4, because these areas have high labor force, industrial labor and not potentiality for paddy farm. These areas spread alongside boundary between Demak regency and Semarang municipality (province capital), at southwest and west region at study area. Figure 6 explained spatial resources character based on rural area at Demak regency, Indonesia.

Table 2: Combining score factor 1 and 4

Category	Values of score factor 1	Values of score factor 4	Resource characters	Explanation
(1)	+	+	Very High	These areas have potentiality facility service and supported by urban and industrial activity
(2)	+	-	High	These areas have potentiality facility service, but unsupported by urban and industrial activity
(3)	-	-	Very Low	These areas have not potentiality facility service and unsupported by urban and industrial activity
(4)	-	+	Low	These areas have not potentiality facility service, but supported by urban and industrial activity

Table 3: Combining score factor 3 and 2

Category	Values of score factor 3	Values of score factor 2	Resource characters	Explanation
(1)	+	+	Paddy farm (high)	These areas have human resources for agricultural activity and supported by paddy farm field
(2)	+	-	Wheatland (high)	These areas have human resources for agricultural activity and supported by wheat land field
(3)	-	-	Wheatland (middle)	These areas have wheatland field, but unsupported by human resources for agricultural activity
(4)	-	+	Paddy farm (middle)	These areas have paddy farm field, but unsupported by human resources for agricultural activity

Table 4: Combining score factor 3 and 5

Category	Values of score factor 3	Values of score factor 5	Resource characters	Explanation
(1)	+	+	Low	These areas have human resources for agricultural activity but unsupported by brackish fishpond
(2)	+	-	High	These areas have human resources for agricultural activity and supported by brackish fishpond
(3)	-	-	Middle	These areas have brackish fishpond, but unsupported human resources for agricultural activity
(4)	-	+	Low	These areas have not both of potentiality requirement

Table 5: Combining score factor 2 and 4

Category	Values of score factor 2	Values of score factor 4	Resource characters	Explanation
(1)	+	+	Middle	These areas have only one factor score supported industrial activity
(2)	+	-	Low	These areas have not both of supported industrial activity
(3)	-	-	Middle	These areas have only one factor score supported industrial activity
(4)	-	+	High	Both of score factors on these areas supported industrial activity

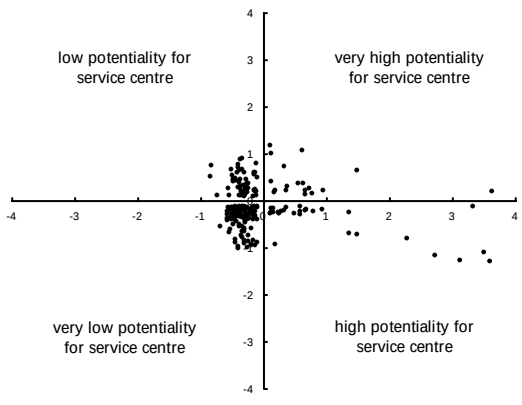

Figure 2: Service center potentiality combining score factor 1 and 4

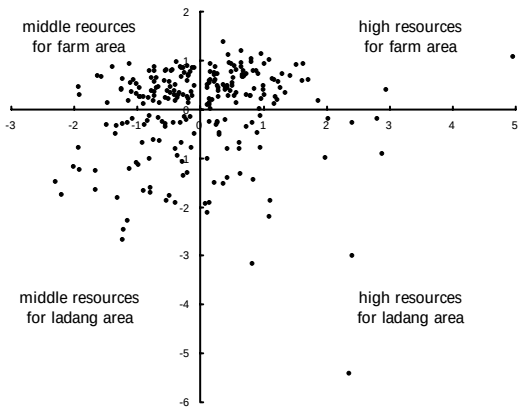

Figure 3: Paddy farm and wheatland resources combining score factor 3 and 2

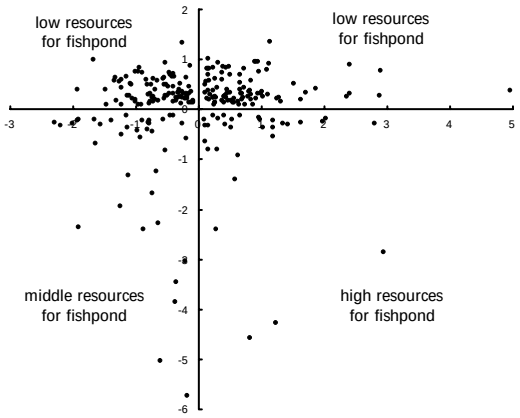

Figure 4: Fishpond resources combining score factor 3 and 5

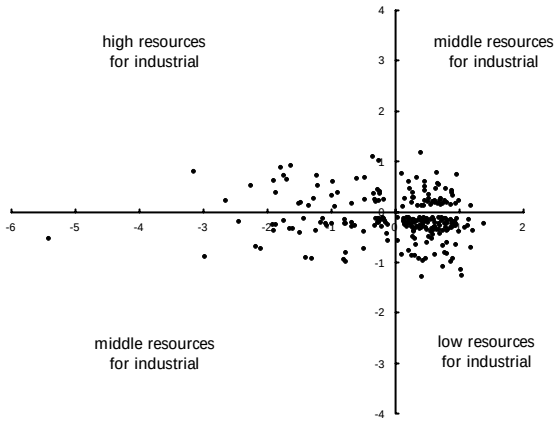

Figure 5: Industrial activity resources combining score factor 2 and 4

4. Conclusions and Discussions

Score factor that resulted in this study will be valuable to provide significant information for development stakeholders, especially government and private sector. First, government will get input to make land-use planning based on the specific resources character. Second, local government able to detect the valuable place as a new service center area based on location and potentiality. Third, assist government to determine infrastructure development and subsidy priorities that necessary for each area. During this time, these area undeveloped because have no specific information about resource characteristics. Fourth, assist private sectors to detect the most suitable area for invests, for example: industrial area and urban area. Fifth, define the relationship variable, for example: industrial area has inverted relation with paddy farm, but has significant relation with wheatland and population.

This study only discussed about general variable of resource characteristics. In this respect, further study should consider about detailed of soil and water characteristics for every rural area to make better result. Moreover, change of resource characteristics will be very significant to make rural area planning based on trend data. This paper only focused one-year data, actually data trend in 5-10 years able to explain about resource characteristics changes. Furthermore, data set collection in wider region and bigger variable will get better comparison than narrow and little. However, this paper able to explained about resource characteristics of rural area that will be valuable for development stakeholders.

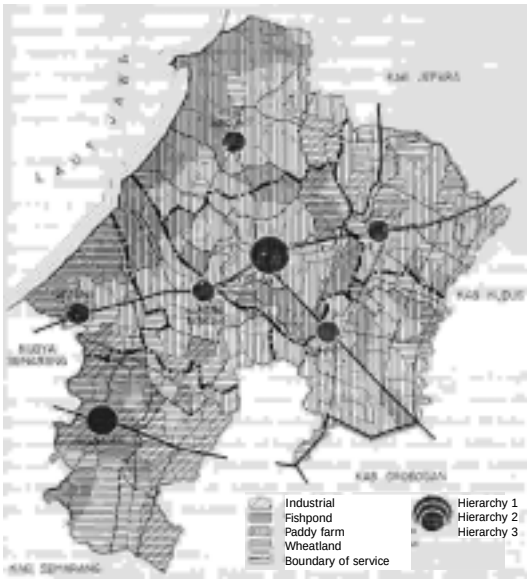

Figure 6: Resource characteristics in Demak Regency, Indonesia

References

1) Carsjens, G. J. and Knaap, W.: Strategic land-use allocation: dealing with spatial relationships and fragmentation of agriculture, Landscape and Urban Planning, Vol. 58, pp. 171-179, 2002.
2) Bouchard, A. and Domon, G.: The transformations of the natural landscapes of the Haut Saint Laurent (Quebec) and their implications on future resource management. Landscape and Urban Planning, Vol. 37, pp. 99-107, 1997.
3) Darlington, R. B.: Factor Analysis, Department of psychology, Cornell University, USA, 2001.
4) Statsoft, Inc.: Electronic statistics textbook, http://www.statsoftinc.com, 2002.
5) Comrey A. L. and Lee H. B.: A First Course in Factor Analysis, Assessment Systems Corporation, 1992.
6) Sugiri, A.: The study of regional policy at Madiun regency, Department of Regional & City Planning, Bandung Institute of Technology, 1990 (In Indonesian).
7) McDade, T. W. and Adair, L. S.: Defining the “urban” in urbanization and health: a factor analysis approach, Social Science & Medicine, Vol. 53, pp. 55-69, 2001.