Bulletin of Geography. Socio-economic Series, No. 51 (2021): 109–122
http://doi.org/10.2478/bog-2021-0009

ISSN 1732–4254 quarterly

BULLETIN OF GEOGRAPHY. SOCIO–ECONOMIC SERIES

journal homepages:
https://content.sciendo.com/view/journals/bog/bog-overview.xml
http://apcz.umk.pl/czasopisma/index.php/BGSS/index

# Multivariate classification of provinces of Vietnam according to the level of sustainable development

**Van Canh Truong**

*University of Da Nang,* University of Science and Education, 459 Ton Duc Thang – Da Nang – Vietnam, tel: (+48) 729 451 495, e-mail: tvcanh@ued.udn.vn, https://orcid.org/0000-0001-6024-6692

How to cite:
Truong, V.C. (2021). Multivariate classification of provinces of Vietnam according to the level of sustainable development. *Bulletin of Geography. Socio-economic Series*, 51(51): 109-122. DOI: http://doi.org/10.2478/bog-2021-0009

**Abstract.** The research aims to classify the level of sustainability of 63 provinces in Vietnam upon 24 indicators reflecting three main dimensions of sustainable development by using multivariate classification method for the year 2014-2016. First of all, the principal component analysis (PCA) was applied to group quantitative variables that reflect important aspects of each component of sustainable development of localities in Vietnam into a number of limited dimensions (factors). The results of PCA illustrate 8 principal components in which 3 main components of economic and social pillar, and 2 main components for environmental pillar. After that, the second method was applied by using the hierarchical methods of cluster analysis for the set of 8 principal components conducted by PCA. With the candidate solution divided all of the provinces into 6 clusters, the analysis shows that the higher level of sustainable development belongs to South East, Red River Delta, and North Central, in which 4 provinces of South East (Ho Chi Minh City, Binh Duong, Dong Nai, and Ba Ria – Vung Tau) form a group with the highest level of sustainability. The regions such as South Central Coastal, Highland area, and Mekong River Delta have the medium level; the North midland and mountain areas are regions at the low level of sustainability.

**Article details:**
Received: 27 March 2020
Revised: 29 September 2020
Accepted: 20 October 2020

**Key words:**
principal component analysis,
cluster analysis,
level of sustainability,
indicators

**Contents:**

## 1. Introduction

One of the most striking characteristics of sustainable development is that it means so many different things to so many different people and organizations. Sustainable development involves not just one but three main complex interacting systems (economic, social and environmental subsystem). In each subsystem, it also contains many things and a great deal of relationships. Sustainable economy means diversity, efficiency, creativity and its balanced relations with society and environment. Sustainable society is revealed by justice, diversity, security and its balanced relations with economy and environment. Finally, sustainable environment means closeness of ecosystem, biodiversity, trophic structure, resource renewal and its balanced relations with economy and society (Cohen et al., 1998). Adam Szirmai (2015), Professor of Development Economics at Maastricht University considers, that "sustainable development' is rather fuzzy concept which is sometimes broadened to include a variety of desirable goals such as equitable development, social inclusion or poverty reduction. In addition to these three major aspects, many people also mention other aspects of sustainable development such as governance, culture, spirit, ethnicity, law ... and require calculation and balance them in planning strategies and policies for socio-economic development for each specific country and locality. Other proposed aspect of spatial sustainability which provides for a balance at different territorial levels and results into local sustainable communities. This emprical research only take into account the three main components of sustainability (economic, social, environmental) which are accepted widely.

We desire a tool that includes all relevant dimensions of sustainability in the evaluation process to assess their progress, level toward sustainable development and understand the interlinkages within each component and between components of sustainability (Kaivo-Oja et al., 2014; Fredericks, 2014). Indicators, as variables, are indispensable tool to make the concept of sustainable development operational. They help us to construct a comprehensive picture of state of a complex system, to translate the concepts and objectives of sustainable development into practical terms (Hardi et al., 1997; Bossel, 1999; Li-Yin Shen, 2011). Based on 24 indicators which cover a wide range of three main dimensions of sustainable development, the research aims to classify the level of sustainability of 63 provinces in Vietnam by using multivariate classification method. The multivariate classification aims to bring diverse features into certain groups, with similar characteristics, expressed in certain sets of criteria. First of all, the principal component analysis (PCA) was utilized to group quantitative variables that reflect important aspects of each component of sustainable development of localities in Vietnam into a number of limited dimensions (factors, principal components). After that, the second method was applied by using the hierarchical methods of cluster analysis for the set of principal components conducted by PCA. The analysis use the complete raw existing data set provided for 24 indicators for the years 2014 – 2016 of 63 administrative units of Vietnam.

## 2. Indicators and data source

Primary and secondary data from government agencies and academic institutes have been collected. Primary data are comprised of information gathered directly by technological monitoring, including satellite-derived estimates of air quality. Data for annual mean concentration of Particulate Matter 2.5 (μg/m3) by province was synthesized from the satellite data of air quality for Environmental Performance Index of Yale University of United States and retrieved from remote sensing data provided by the Department of Physics and Atmospheric Science, Dalhousie University in Canada. Secondary data include official statistical data formally reported by General Statistics Office (GSO) of Vietnamese government. The set of 24 sustainable development indicators has been constructed. It covers a wide range of issues with 8 indicators for each component of sustainability, as following:

- Economic component (8 indicators): GDP per capita (PPP current USD), GDP density (million USD PPP per km$^2$), Proportion of employment in agriculture (%), Incremental capital-output ratio (ICOR), Unemployment rate (% labor force), Percentage of trained employed workers (%), Competitiveness Index, and Budget surplus as percentage of GDP (%).
- Social component (8 indicators): Adult literacy rate (%), Proportion of household own permanent house (%), Poverty rate (%), Gini index, Female labor force participation rate (% male), Prevalence of underweight children, weight for age (% of children under 5), Average life expectancy at birth (year), and Proportion of death due to traffic accident (per 100.000 people).
- Environmental component (8 indicators): Forest cover (% total land area), Agricultural land per person (ha), Proportion of household with access to improved sanitation (%), Percentage of household access potable water (%), Proportion of rural households using solid fuels for cooking (%), Annual median concentration of Particulate Matter 2.5 (μg/m3), Total of collected solid waste per capita (kg/person/day), and Proportion of collected solid waste per day that are treated according to national standards (%) (Truong, 2019).

## 3. Results and discussion

### 3.1. Principal component analysis (PCA) for sustainable development

By using principal component analysis (PCA), this part of paper aims to group quantitative variables that reflect important aspects of each component of sustainable development of localities in Vietnam into a number of limited dimensions (factors, principal components). In each dimension, variables are more closely correlated with other variables in the same dimension, rather than with variables of other dimensions. It is then possible to interpret each dimension according to the meaning of the loadings of principal component assigned to variables. The loadings (correlation coefficients) mean how many percent of the variance of the i-th variable is explained by the j-th principal component. The reduction of the number of variables allows to better identify the characteristics of the research object (in this study is the sustainable development of provinces and cities). Principal components describe the significant predictors of sustainability in each subsystem and factor scores were used to assess the level of sustainability taking into consideration the positive or negative effect of these factors. One of the rules used in the analysis assumes that cumulatively retained principal components should explain at least 60-70% of total variance.

The first step for PCA of each component is testing the correlation matrix. The research used KMO and Bartlett's Test to test the assumption that variables are not correlated with each other in the whole (the null hypothesis - H0). The results from the tables below show that KMO and the Chi-Square test quantity from the determinant of the correlation matrix have a large value, with higher value for economic and social component (0.788 and 0.713 respectively) and lower value for environmental subsystem (0.649). The initial assumption, the null hypothesis, is rejected, the factor analysis method for each component of sustainability is appropriate. The next step is calculating score of each factor and determining the number of factors. The rotation method (Varimax with Kaiser Normalization) was applied only for economic subsystem to minimize the number of variables which have high coefficient in one factor, avoid one variable with high coefficient in other factor, so it can help to improve ability to explain the meaning of factors. The number of factor was determined based on Initial Eigenvalues >1.

**Economic component**

The results for factor analysis of economic component are illustrated in the table 1. These tables show that there are 3 main factors that explain 77.705% of the total variation of economic component. Before doing procedures for rotation, factor 1 explains

49.465%, while factor 2 explains 15.189% and factor 3 explains 13.051%. However, from the factors matrix table, we can see that the factor loading of some indicators such as GDP density, budget surplus as percentage of GDP, Competitiveness Index, percentage of trained employed workers, and Incremental capital-output ratio is quite high in the two other factors. Therefore, the procedure to rotate the factors has been carried out. Rotated Component Matrix was shown in the table 3. After rotation, factor 1 explains 28.595%, while factor 2 explains 26.508% and factor 3 explains 22.601% of the total variation.

Table 1 illustrates that factor 1 represents for *Prosperity of economy* with the main weights belong to GDP density, GDP per capita, percentage of trained employed workers, and proportion of employment in agriculture. Factor 2 can be used to show the *Competitiveness of economy* with the main weights belong to Competitiveness Index and unemployment rate. With high values for Incremental capital-output ratio and budget surplus as percentage of GDP, factor 3 is the factor representing for *Production efficiency*. The results confirm that extracted components describe better the structure of economic subsystem of sustainability and show the intra-system relationships. The more proportional share of each component in common variance than in other components was affected by the rotation procedure of original dimensions.

Factor scores of each factor for each province were calculated and saved in the original file in SPSS. The results then were illustrated on maps by MapInfo software (see figure 1). In these maps, the legends were divided into 5 ranges by natural break method in MapInfo, in order to make the difference between the values of the data and the average value of the data in a series is the smallest. This is considered a way to reduce errors and show more realistic data (Nguyen Viet Thinh and Do Thi Minh Duc, 2005). If one province (one case) has high factor score in a certain factor, the province must have high value for variables with positive number and low value for variables with negative number. By contrast, if one province has low factor score in a factor, the province must have low value for variables with positive number and high value for variables with negative number. For example, in factor 1, Hanoi, Bac Ninh, Hai Phong, Da Nang, Ho Chi Minh City, Binh Duong, and Ba Ria – Vung Tau have high factor score due to they have high value for positive variables such as GDP density, GDP per capita, and percentage of trained employed workers. On the other hand, most of provinces of Mekong River Delta, and Central Highlands have low score for factor 1 because of high proportion of employment in agriculture, meanwhile, low values of GDP density, GDP per capita, and percentage of trained employed workers.

Regarding to factor 2, most of provinces in the delta and coastal regions have higher factor score for factor 2 than one in the mountain regions (Northern midlands and mountain areas, and Central Highlands), because these provinces have higher values for Competitiveness Index, unemployment rate, and lower value for proportion of employment in agriculture. With respect to factor 3, the provinces in South East, Mekong River Delta region have higher score than others due to the fact that they have higher values for budget surplus as percentage of GDP, meanwhile, lower score for Incremental capital-output ratio.

**Social component**

The same procedures of factor analysis were applied for social component. The results are illustrated in table 4.8. These tables show that there are 3 main factors that explain 83.084% of the total variation of social component. Due to the fact that no indicators have high correlation coefficient with other components, procedures for rotation is not necessary, factor 1 explains 46.568%, while factor 2 explains 23.730% and factor 3 explains 12.786%.

Factor 1 represents for *Quality of life* with the main weights belong to poverty rate, average life expectancy at birth, prevalence of underweight children, and adult literacy rate. Factor 2 can be

**Table 1.** Results of PCA for economic component of sustainability in Vietnam

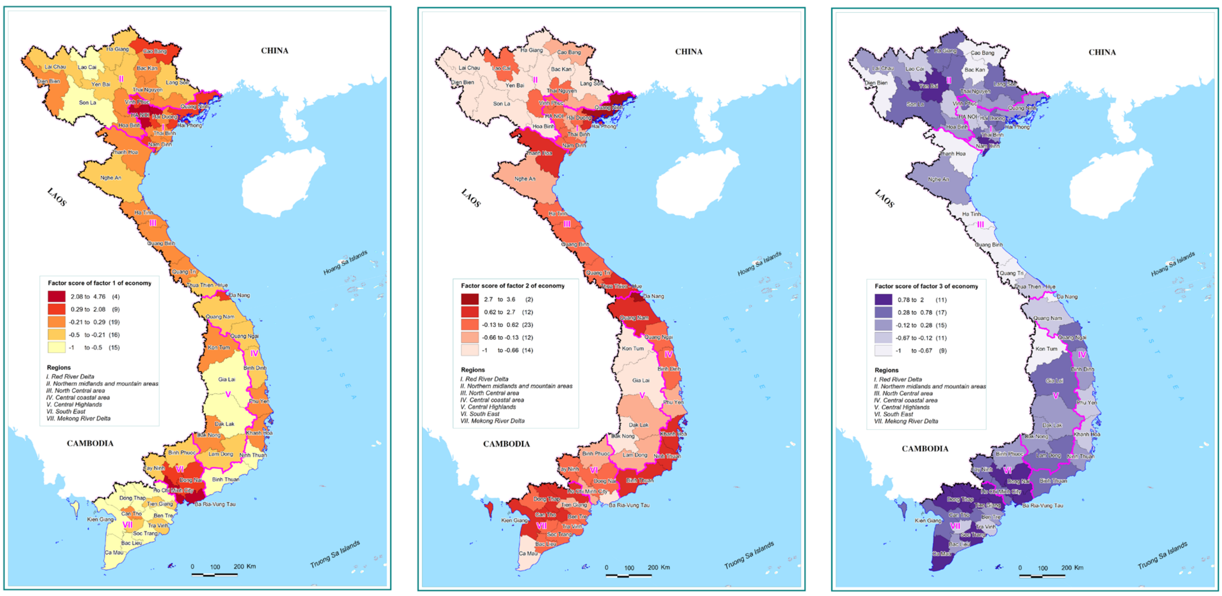
a) Proportion of the total variance explained by three main factors

| Component | Extraction sums of squared loadings | | | Rotation sums of squared loadings | | |
|---|---|---|---|---|---|---|
| | **Eigenvalues** | **% of variance** | **Cumulative %** | **Eigenvalues** | **% of variance** | **Cumulative %** |
| 1 | 3.957 | 49.465 | 49.465 | 2.288 | 28.595 | 28.595 |
| 2 | 1.215 | 15.189 | 64.654 | 2.121 | 26.508 | 55.104 |
| 3 | 1.044 | 13.051 | 77.705 | 1.808 | 22.601 | 77.705 |

b) Rotated component matrix

| | Component | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| GDP density | 0.889 | 0.057 | 0.192 |
| GDP per capita | 0.727 | 0.129 | 0.362 |
| Percentage of trained employed workers | 0.700 | 0.480 | -0.171 |
| Competitiveness Index | 0.134 | 0.837 | 0.160 |
| Unemployment rate | 0.130 | 0.836 | 0.213 |
| Proportion of employment in agriculture | -0.604 | -0.625 | -0.271 |
| Incremental capital-output ratio | -0.083 | -0.146 | -0.910 |
| Budget surplus as percentage of GDP | 0.266 | 0.243 | 0.799 |

Source: Author's calculation




*a) Factor score of factor 1 of economic subsystem (prosperity) by province*
*b) Factor score of factor 2 of economic subsystem (competitiveness) by province*
*c) Factor score of factor 3 of economic subsystem (efficiency) by province*

**Fig.1.** Three main factors of economic component of sustainability
Source: Author's elaboration

used to show the *Equality* with the main weights belong to Gini index, proportion of household own permanent house, and female labor force participation rate. With only one indicator for proportion of death due to traffic accident, factor 3 is the factor representing for *Insecurity of traffic* (see table 2). The results confirm that the extracted components adequately describe the structure of social subsystem of sustainability. Quality of life and social equality are common indicators of social component of sustainability. Last principal component has a very high loading in one original variable and it is difficult to accept that it can be treated as universal indicator of security in the society.

Examining the factor scores shows us that Hanoi and other provinces of Red River Delta, Da Nang, and provinces of South East region have high factor score for factor 1 of quality of life due to they have high value for positive variables (average life expectancy at birth, and adult literacy rate) and low value for negative variables such as poverty rate and prevalence of underweight children. Meanwhile, due to low factor score for positive and high factor score for negative variables, most of provinces of Northern midlands and mountain areas, and Central Highlands have low score for quality of life. Regarding to factor 2, with high value of proportion of household own permanent house, female labor force participation rate and low value of Gini index, most of provinces of Red River Delta and North Central area have higher score for factor 2 than others. Factor scores of factor 3 illustrate that the proportion of death due to traffic accident is higher in the regions in the South of Vietnam, such as Central Highlands, North Central area, Central Coastal area, South East (see figure 2).

**Table 2.** Results of PCA for social component of sustainability in Vietnam

a) Proportion of the total variance explained by three main factors

| Component | Initial eigenvalues | | | Extraction sums of squared loadings | | |
|---|---|---|---|---|---|---|
| | Eigenvalues | % of variance | Cumulative % | Eigenvalues | % of variance | cumulative % |
| 1 | 3.725 | 46.568 | 46.568 | 3.725 | 46.568 | 46.568 |
| 2 | 1.898 | 23.730 | 70.298 | 1.898 | 23.730 | 70.298 |
| 3 | 1.023 | 12.786 | 83.084 | 1.023 | 12.786 | 83.084 |

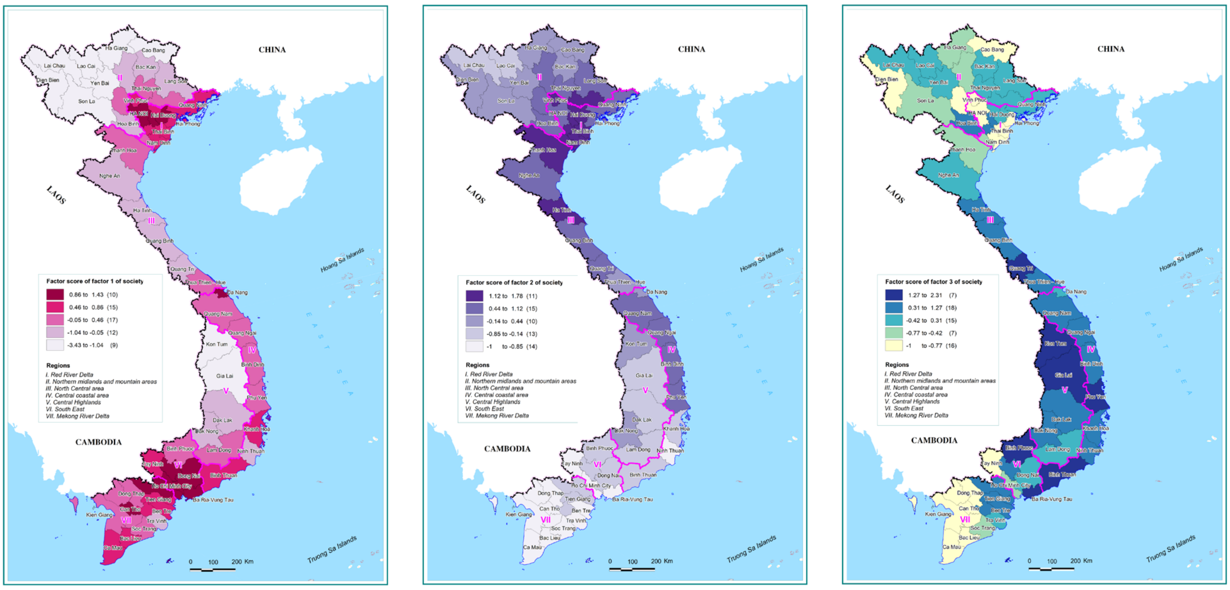
b) Component matrix

| | Component | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Poverty rate | -0.949 | 0.057 | -0.038 |
| Average life expectancy at birth | 0.931 | -0.137 | -0.134 |
| Prevalence of underweight children, weight for age | -0.864 | 0.137 | 0.131 |
| Adult literacy rate | 0.846 | 0.287 | 0.069 |
| Gini index | -0.539 | -0.407 | -0.333 |
| Proportion of household own permanent house | 0.192 | 0.924 | -0.011 |
| Female labor force participation rate | -0.411 | 0.821 | 0.102 |
| Proportion of death due to traffic accident | 0.009 | -0.284 | 0.927 |

Source: Author's calculation

**Environmental component**

The same procedures of factor analysis will be applied for environmental component. The results are illustrated in table 4.9. These tables show that there are 2 main factors that explain 64.086% of the total variation of environmental component. As social component, procedure for rotation is not necessary with environmental component because no indicators have high value with other components. Factor 1 explains 44.823%, while factor 2 explains 19.263%. Based on the results from the table 3, we suggest that factor 1 represents for *Urban-rural living environment* with the main weights belong to proportion of rural households using solid fuels for cooking, proportion of household with access to improved sanitation, percentage of household access potable water, total of collected solid waste per capita, agricultural land per person, forest cover. Factor 2 can be used to represent for *Pollu-*




*a) Factor score of factor 1 of social subsystem (Quality of life) by province* *b) Factor score of factor 2 of social subsystem (Equality) by province* *c) Factor score of factor 2 of social subsystem (Insecurity of traffic) by province*

**Fig. 2.** Three main factors of social component of sustainability
Source: Author's elaboration

*tion* with the main weights belong to Annual median concentration of Particulate Matter 2.5, and proportion of collected solid waste per day that are treated according to national standards.

With respect to factor 1 of urban-rural living environment, most of provinces of Red River Delta, South East, and Central Coastal area with higher score than other regions due to they have higher values for two positive variables which represent for urban living environment: percentage of household

**Table 3.** Results of PCA for environmental component of sustainability in Vietnam
a) Proportion of the total variance explained by three main factors

| Component | Initial eigenvalues | | | Extraction sums of squared loadings | | |
|---|---|---|---|---|---|---|
| | Eigenvalues | % of variance | Cumulative % | Eigenvalues | % of variance | Cumulative % |
| 1 | 3.586 | 44.823 | 44.823 | 3.586 | 44.823 | 44.823 |
| 2 | 1.541 | 19.263 | 64.086 | 1.541 | 19.263 | 64.086 |

**b) Component matrix**

| 1<br>2 | Component | |
|---|---|---|
| Proportion of rural households using solid fuels for cooking | -0.903 | 0.175 |
| Proportion of household with access to improved sanitation | 0.879 | -0.176 |
| Percentage of household access potable water | 0.814 | -0.349 |
| Total of collected solid waste per capita | 0.632 | 0.367 |
| Agricultural land per person | -0.615 | -0.424 |
| Forest cover | -0.558 | 0.354 |
| Annual median concentration of Particulate Matter 2.5 | 0.035 | 0.781 |
| Proportion of collected solid waste per day that are treated according to national standards | 0.494 | 0.555 |

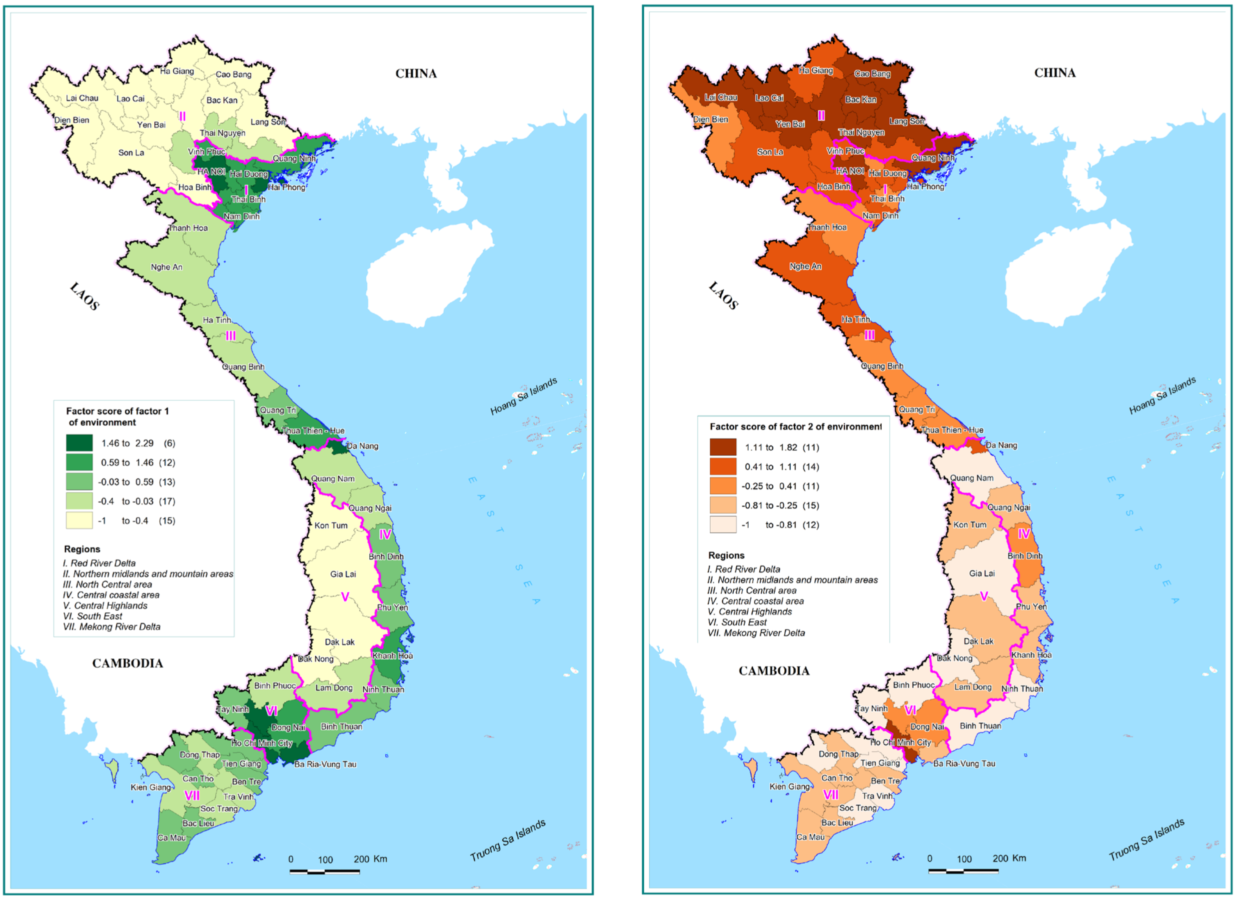
Source: Author's elaboration

access potable water, and proportion of household with access to improved sanitation. Meanwhile, these provinces have lower values for negative variables which represent for rural living environment such as proportion of rural households using solid fuels for cooking, and forest cover. By contrast, provinces of Central Highlands, Northern midlands and mountain areas, and North Central area have higher score for variables of rural living environment and lower values for variables of urban living environment (see figure 3a).

Regarding to factor 2 of pollution, the results show that provinces which distribute from Da Nang to the north of Vietnam have higher level of pollution than others in the south due to they have higher score for positive variables such as proportion of collected solid waste per day that are treated according to national standards, annual median concentration of Particulate Matter 2.5, and total of collected solid waste per capita. One of the main reasons is that most of coal-fired thermal power plants, cement factories, mining industries… are concentrated in the north part of Vietnam. The map also shows that the provinces of Red River Delta and South East where concentrate the big cities of Vietnam have higher level of pollution than others. In fact, this is a clear indicator of environmental degradation in the country due to it concerns also rural areas (especially in northern part of the country), and it is not associated with previous dimension of environmental subsystem (see figure 3b).

### 3.2. Classification of level of sustainability by using cluster analysis

Some kind of synthetic approach can be obtained by using appropriate classification procedures. The simplest way is multivariate classification, which is referred to as cluster analysis (Everitt *et al.*, 1993; Hill et al., 1998; Mierzejewska, 2009). Cluster analy-




*a) Factor score of factor 1 of environmental subsystem (urban-rural living environment) by province*

*b) Factor score of factor 2 of environmental subsystem (pollution) by province*

**Fig. 3.** Two main factors of environmental component of sustainability
Source: Author's elaboration

sis or clustering is a methodology to identify groups of a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). Cluster analysis maximizes the similarity of cases within each cluster while maximizing the dissimilarity between groups that are initially unknown (Roy, Kar and Das, 2015). In cluster analysis, a large number of methods are available for classifying objects on the basis of their (dis)similarities. Major types of cluster analysis are hierarchical methods (agglomerative or divisive), non-hierarchical method (K-Means clustering). Within each type of methods a variety of specific methods and algorithms exist. Perhaps the most common form of analysis in geography is the agglomerative hierarchical cluster analysis (Smelser and Baltes, 2001). It starts with each case as a separate cluster (i.e., there are as many clusters as cases), and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. Groups are constructed by minimizing the variance of squared Euclidean distances for each variable between cases (provinces). The number of stages in the process is one less than the number of cases. A hierarchical tree diagram or dendrogram can be generated to show the linkage points: the clusters are linked at increasing levels of dissimilarity (Roy, Kar and Das, 2015).

The research used hierarchical methods according to the squared Euclidean distance and Ward method for selection of sequential clusters. Each province was characterized by eight standardized variable scores (z-scores) extracted as factor-scores of principal components in PCA procedure for each set of original variables which characterized each subsystem of sustainable development: three factors for economic component, three factors for social component ad two factors for environmental one (see section 3.1). The results of cluster analysis are shown in tree diagram (dendrogram) (Figure 4).

The basic problem to be solved in the classifying procedure is to choose the best number of classes (clusters). So far, there are no clear and certain rules for determining the number of clusters. In other words, the number of necessary and logical clusters is not a completely technical issue, but it depends on many other factors (Trong Hoang and Mong Ngoc Chu Nguyen, 2008). Most often, two different approaches are used. In the first one, certain number of classes is assumed taking into account the separation of the number of classes. In the second approach, the relatively best number of classes is sought with the use of appropriate statistical methods. In the study the 'optimal' cluster solution was prepared according to the suggestion in paper by Everitt *et al* (1993) and Hill *et al.* (1998). The base for identification of 'correct' number of clusters is agglomeration schedule for cluster analysis. The agglomeration coefficient is the sum of the within group variance of the two clusters combined at each successive stage. The marked increase in the value of the agglomeration coefficient between two stages indicates that heterogenous clusters are being combined. Greater increase in total variance means higher heterogeneity of cluster at next stage.

Table 4 contains the partial agglomeration schedule for the cluster analysis, include 5 columns: number of stage agglomeration, clusters in solution, agglomeration coefficient, the first quotient (the percentage change in the value of the agglomeration coefficient from the previous stage), the second quotient (the percentage change, of the percent change, found in column four). The quotient means the rate change in agglomeration coefficient from one stage to the next. The largest increase in the quotient in column fourth and fifth indicates that next set of clusters is characterized by higher level of heterogeneity than previous one, which should be accept as correct solution of clusters' number.

There are three candidate solutions were detected, at 3 clusters, 6 clusters and 10 clusters, based on the first and second quotients of the agglomeration schedule with the largest quotients by the following moves: 3 clusters to 2, 6 clusters to 5, and 10 clusters to 9. Hence the 6 clusters have been selected as medium solution to present results to distinguish the groups of provinces. The statistical characteristics of each cluster are illustrated by tables. The map of figure 5 illustrates the results of classification of Vietnamese provinces according to the level of sustainable development.

The analysis of the structure of the dendrogram and the 6-element system and the average values of factor scores for these six clusters (table 5 and 6) showed that the dendrogram can be divided into

**Table 4.** Partial agglomeration schedule for cluster analysis of 8 principal components

| Stage | Clusters solution | Agglomeration coefficient | First quotient | Second quotient |
|---|---|---|---|---|
| 51 | 12 | 102.46 | 9.73 | -2.70 |
| 52 | 11 | 112.23 | 9.54 | -1.92 |
| 53 | 10 | 122.43 | 9.08 | -4.78 |
| 54 | 9 | 136.84 | 11.77 | 29.60 |
| 55 | 8 | 153.93 | 12.49 | 6.09 |
| 56 | 7 | 173.62 | 12.80 | 2.47 |
| 57 | 6 | 194.56 | 12.06 | -5.76 |
| 58 | 5 | 223.14 | 14.69 | 21.84 |
| 59 | 4 | 260.87 | 16.91 | 15.07 |
| 60 | 3 | 306.23 | 17.39 | 2.85 |
| 61 | 2 | 396.14 | 29.36 | 68.86 |
| 62 | 1 | 496.00 | 25.21 | -14.15 |

Source: Author's calculation

two basic groups of provinces differing in the degree of advancement of sustainability. On the one hand, these are clusters of more urbanized areas in the North and South of the country (class I and II), with higher values of dimensions describing sustainability in southern part of the country (class I), and medium sustainability according to national average (class III and class IV). On the other hand, the second branch of dendrogram includes clusters of provinces with the lowest advancement of sustainability (class V and class VI). It was assumed that values of these average factor-scores between -0.5 and +0.5 describe average national characteristics, while higher than +1.0 mean the highest sustainability, and by analogy lower than -1.0 the lowest sustainability. Of course the interpretation should take into account the stimulants and destimulants of sustainability.

Class I contains 4 provinces of South East (Ho Chi Minh City, Binh Duong, Dong Nai, Ba Ria – Vung Tau). They are the most urbanized areas, ranked at the top of national arrangement with highest level for prosperity of economy, production efficiency, and quality of life. However, unlike the class II, this class has a high level of social inequality, and average insecurity of traffic

Class II embraces 19 provinces in which all of 11 provinces of the Red River Delta belong to this group, 1 province of Central region (Da Nang), 4 provinces of the North Central (Thanh Hoa, Nghe An, Ha Tinh, and Quang Binh), and 3 provinces of North midland and mountain areas which have proximity with Red River Delta (Phu Tho, Thai Nguyen, and Bac Giang). The characteristics of this cluster are represented by the average level for prosperity and competitiveness of economy, quality of life, but the highest social equality in the country. However, it has higher scope of urban environment and higher level of pollution as well.

Class III includes 19 provinces in which most of them come from Central Coastal area and Central Highlands. It also combines some provinces of Mekong River Delta and North Central area. The main characteristic of this class is that it still has average level of dimensions, but it has the highest insecurity of traffic in the country and relatively lower pollution.

Class IV combines 10 provinces. It contains almost administrative units (9 provinces) of the Mekong River Delta and one province of South East (Tay Ninh). This cluster has the same characteristics with cluster III. However, unlike cluster III, it has also the lowest social equality and low insecurity of traffic.

Class V includes 9 provinces and all of them come from North midland and mountain areas (Hoa Binh, Son La, Lai Chau, Lao Cai, Ha Giang, Bac Kan, Yen Bai, Tuyen Quang, and Lang Son). All of factor scores of 8 principal components for this cluster are very low, except prosperity and efficiency of economy, social equality and traffic insurance, which are average, but pollution is the highest in the country.

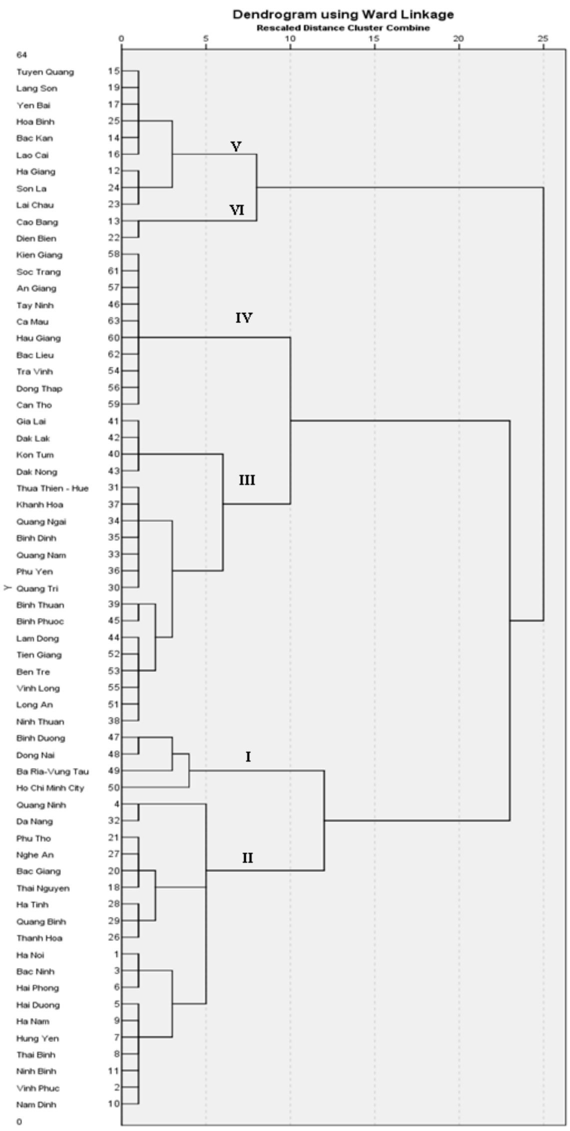
Class VI has lowest factor scores for all of 8 principal components. It includes only two provinces (Dien Bien and Cao Bang) belong to the North midland and mountain areas. The level of sustainability is the lowest in the country due to the lowest efficiency of economy and quality of life, and lower competitiveness of economy and higher pollution.

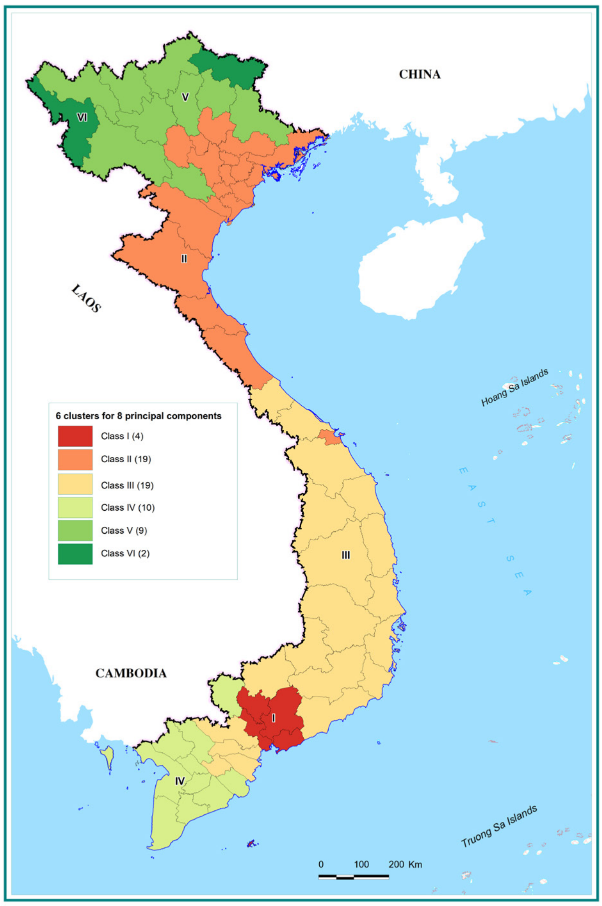
**Table 5.** The classification of Vietnamese provinces by principal components of sustainability

| Classes of provinces | PC1* | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 |
|---|---|---|---|---|---|---|---|---|
| Class I | 2.35** | -0.10 | 1.22 | 1.16 | -1.15 | 0.57 | 1.72 | 0.38 |
| Class II | 0.41 | 0.40 | -0.18 | 0.49 | 1.13 | -0.32 | 0.62 | 0.67 |
| Class III | -0.41 | 0.08 | 0.06 | -0.04 | -0.28 | 1.01 | -0.13 | -0.85 |
| Class IV | -0.67 | 0.29 | 0.40 | 0.39 | -1.31 | -1.05 | 0.02 | -0.97 |
| Class V | -0.36 | -1.13 | 0.10 | -1.43 | 0.16 | -0.29 | -1.39 | 1.14 |
| Class VI | 0.28 | -0.72 | -3.81 | -2.09 | -0.02 | -1.10 | -1.92 | 0.75 |

*PC1: Prosperity of economy; PC2: Competitiveness of economy; PC3: Production efficiency; PC4: Quality of life; PC5: Equality; PC6: Insecurity of traffic; PC7: Urban-rural living environment; PC8: Pollution. ** Mean values of principal components factor-scores from Tab. 5.3

Source: Author's calculation




▲ **Fig. 5.** Clustering provinces into 6 groups according to 8 principal components
Source: Author's elaboration

◄ **Fig. 4.** Dendrogram for cluster analysis of provinces according to 8 principal components

**Table 6.** Descriptive statistics for 6 groups of provinces according to 8 principal components

| Cluster | Principal component | Count | Mean | Maximum | Minimum |
|---|---|---|---|---|---|
| I | Prosperity of economy | 4 | 2.35 | 4.76 | 0.34 |
| | Competitiveness of economy | 4 | -0.10 | 0.29 | -0.72 |
| | Production efficiency | 4 | 1.22 | 1.99 | 0.38 |
| | Quality of life | 4 | 1.16 | 1.42 | 0.99 |
| | Equality | 4 | -1.15 | -0.70 | -2.17 |
| | Insecurity of traffic | 4 | 0.57 | 2.03 | -0.77 |
| | Urban-rural living environment | 4 | 1.72 | 2.28 | 1.22 |
| | Pollution | 4 | 0.38 | 1.82 | -0.35 |
| II | Prosperity of economy | 19 | 0.41 | 2.70 | -0.34 |
| | Competitiveness of economy | 19 | 0.40 | 3.59 | -0.63 |
| | Production efficiency | 19 | -0.18 | 0.83 | -2.07 |
| | Quality of life | 19 | 0.49 | 1.33 | -0.30 |
| | Equality | 19 | 1.13 | 1.77 | 0.09 |
| | Insecurity of traffic | 19 | -0.32 | 1.01 | -1.64 |
| | Urban-rural living environment | 19 | 0.62 | 1.97 | -0.40 |
| | Pollution | 19 | 0.67 | 1.62 | -0.20 |
| III | Prosperity of economy | 19 | -0.41 | 0.11 | -0.79 |
| | Competitiveness of economy | 19 | 0.08 | 1.42 | -2.53 |
| | Production efficiency | 19 | 0.06 | 0.79 | -0.98 |
| | Quality of life | 19 | -0.04 | 0.94 | -1.74 |
| | Equality | 19 | -0.28 | 0.81 | -1.05 |
| | Insecurity of traffic | 19 | 1.01 | 2.31 | -0.18 |
| | Urban-rural living environment | 19 | -0.13 | 0.75 | -1.20 |
| | Pollution | 19 | -0.85 | 0.22 | -1.91 |
| IV | Prosperity of economy | 10 | -0.67 | -0.05 | -1.23 |
| | Competitiveness of economy | 10 | 0.29 | 1.47 | -0.81 |
| | Production efficiency | 10 | 0.40 | 0.97 | -0.67 |
| | Quality of life | 10 | 0.39 | 0.93 | -0.15 |
| | Equality | 10 | -1.31 | -0.97 | -1.83 |
| | Insecurity of traffic | 10 | -1.05 | -0.30 | -2.24 |
| | Urban-rural living environment | 10 | 0.02 | 0.39 | -0.36 |
| | Pollution | 10 | -0.97 | -0.47 | -1.89 |
| V | Prosperity of economy | 9 | -0.36 | 0.06 | -0.72 |
| | Competitiveness of economy | 9 | -1.13 | 0.15 | -2.34 |
| | Production efficiency | 9 | 0.10 | 0.79 | -1.41 |
| | Quality of life | 9 | -1.43 | -0.35 | -3.42 |
| | Equality | 9 | 0.16 | 0.81 | -0.47 |
| | Insecurity of traffic | 9 | -0.29 | 0.42 | -0.72 |
| | Urban-rural living environment | 9 | -1.39 | -1.02 | -1.77 |
| | Pollution | 9 | 1.14 | 1.47 | 0.62 |
| VI | Prosperity of economy | 2 | 0.28 | 0.50 | 0.06 |
| | Competitiveness of economy | 2 | -0.72 | -0.66 | -0.78 |
| | Production efficiency | 2 | -3.81 | -3.05 | -4.57 |
| | Quality of life | 2 | -2.09 | -1.70 | -2.48 |
| | Equality | 2 | -0.02 | -0.01 | -0.03 |
| | Insecurity of traffic | 2 | -1.10 | -0.83 | -1.36 |
| | Urban-rural living environment | 2 | -1.92 | -1.66 | -2.19 |
| | Pollution | 2 | 0.75 | 1.23 | 0.27 |

Source: Author's calculation

## 4. Conclusion

The research has utilized quantitative methods (Principal component analysis – PCA and Cluster analysis) to classify the level of sustainable development of provinces in Vietnam based on a set of 24 indicators which cover main issues of three components (economic, social, and environmental) of sustainability in Vietnam. However, in the context of Vietnam, selecting essential indicators becomes an actual challenge for the research due to lacking available data. Some aspects of sustainable development have no essential indicators, especially for the environmental component such as land quality, environmental conservation, preservation of biodiversity, renewable energy, and so on.

The PCA procedure was utilized at first to group quantitative variables of sustainable development of 63 provinces of Vietnam into a number of limited factors, principal components. The results of PCA illustrate 8 principal components in which 3 main components of economic (prosperity, competitiveness, and production efficiency), 3 components of social pillar (quality of life, equality, and insecurity of traffic), and 2 main components for environmental pillar (urban-rural living environment, and pollution). The factor scores of each factor showed the significant spatial differentiations of sustainable development lelvel between regions of Vietnam. The second method was applied to classify the level of sustainabily for 63 provinces of Vietnam by using cluster analysis for the set of 8 principal components conducted by PCA. The analysis show that the higher level of sustainable development belongs to South East, Red River Delta, and North Central, in which 4 provinces of South East (Ho Chi Minh City, Binh Duong, Dong Nai, and Ba Ria – Vung Tau) form a group with the highest level of sustainability. The regions such as South Central Coastal, Highland area, and Mekong River Delta have the medium level; the North midland and mountain areas are regions at the low level of sustainability.

## Reference

**Bossel, H.** (1999). *Indicators for Sustainable Development: Theory, Method, Applications*. The International Institute for Sustainable Development, Manitoba, Canada.

**Everitt, B. Landau, S. Leese, M. and Stahl, D.** (2011). *Cluster analysis*. John Wiley and Sons, Ltd Registered. DOI: http://doi.org/10.1007/BF00154794.

**Cohen, S. Demeritt, D. Robinson, J. and Rothman, D.** (1998). Climate change and sustainable development: Towards dialogue, *Global Environmental Change*, 8(4): 341–371. DOI: http://doi.org/10.1016/S0959-3780(98)00017-X

**Fredericks, S.E.** (2014). *Measuring and Evaluating Sustainability: Ethics in Sustainability Indexes*. Routledge. DOI: http://doi.org/10.4324/9780203694312

**Hardi, P. Barg, S. Hodge, T. and Pinter, L.** (1997). *Measuring Sustainable Development: Review of Current Practice*.

**Hill, E. W. Brennan, J.F. and Wolman, H.L.** (1998). What is a central city in the United States? Applying a statistical technique for developing taxonomies, *Urban Studies*, 35(11): 1935–1969. DOI: http://doi.org/10.1080/0042098983962

**General Statistics Office (GSO).** (2015, 2017). *Result of the Vietnam household living standards surveys 2014, 2016.* Hanoi: Statistical Publishing House.

**General Statistics Office (GSO).** (2017). *Vietnam Statistical Yearbook 2016*. Hanoi: Statistical Publishing House.

**General Statistics Office (GSO)** (2017). *Vietnam rural, agriculture and fishery census 2016.* Hanoi: Statistical Publishing House.

**General Statistics Office (GSO).** (2010). *Vietnam Population and Housing census 2009.* Hanoi: Statistical Publishing House.

**General Statistics Office (GSO).** (2017). *Vietnam Statistical Yearbook 2016*. Hanoi: Statistical Publishing House.

**General Statistics Office of Vietnam (GSO).** (no date). Statistical data. Available at: https://gso.gov.vn/Default_en.aspx?tabid=766 (Accessed: 27 January 2019).

**Kaivo-Oja, J. Panula-Ontto, J. Vehmas, J. and Luukkanen, J.** (2014). Relationships of the dimensions of sustainability as measured by the sustainable society index framework, *International Journal of Sustainable Development and World Ecology*. Taylor and Francis, 21(1): 39–45. DOI: http://doi.org/10.1080/13504509.2013.860056

**Li-Yin Shen, Jorge Ochoa, J. Mona N. Shah, and Xiaoling Zhang** (2011). The application of urban sustainability indicators - A comparison between various practices, *Habitat International*. Elsevier Ltd, 35(1): 17–29. DOI: http://doi.org/10.1016/j.habitatint.2010.03.006

**Mierzejewska, L.** (2009). *Rozwój zrównoważony miasta. Zagadnienia poznawcze i praktyczne*. Poznań: Adam Mickiewicz University Press. (in Polish)

**Nguyen Viet Thinh and Do Thi Minh Duc.** (2005). *Socio-Economic Typology of Provinces and Districts of Vietnam*. Hanoi: The Ecyclopedia Publishing House. (in Vietnamese).

**Roy, K. Kar, S. and Das, R.N.** (2015). *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. Academic Press, an imprint of Elsevier.

**Smelser, N.J. and Baltes, P.B.** (2001). *International encyclopedia of the social & behavioral sciences*. Elsevier publishing.

**Szirmai, A.** (2015). *Socio-Economic Development*. Cambridge University Press. DOI: http://doi.org/10.1017/CBO9781107054158

**Truong, V.C.** (2019). Measuring sustainable development level in Vietnam by using a comprehensive index', *Prace i Studia Geograficzne*, 64(2): 7–21.

Ministry of Science and Higher Education Republic of Poland

The proofreading of articles, positively reviewed and approved for publishing in the 'Bulletin of Geography. Socio-economic Series', was financed from the funds of the Ministry of Science and Higher Education earmarked for activities popularizing science, in line with Agreement No 695/P-DUN/2018.