

## Sustainability Index (SI) Model by Using PCA

Haodong Liu <sup>a</sup>, Xi Wang <sup>b</sup> and Demin Liu <sup>c</sup>

School of North China Electric Power University, Baoding 071000, China

<sup>a</sup>229259729@qq.com, <sup>b</sup>453239080x@qq.com, <sup>c</sup>15383220223@189.com

**Keywords:** Ecology Health Index (EHI) Model; Principal Component Analysis; Sustainable development.

**Abstract.** The problems caused by the unbalanced and fragmented development has become more and more obvious. In order to help these government and organizations to make a better decision to give support to the country's most in need, we need to develop a model to assess the sustainability of a country. On the basis of the Ecology Health Index (EHI) Model, the Sustainability Index (SI) Model is established to describe and assess the sustainability of a country. To modify SI Model into a quantized model. There are three main steps: Selection, Standardization and Weight Determination of the Indicators. We use the Principal Component Analysis (PCA) method to determine the weighting factors of each indicator.

### 1. Introduction

Sustainable development is defined in the 1987 Brundtland Report as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs”.

People realize that the fragmented development of economy, society and environment can only be a devastating disaster to the earth and the human society. In fact, quite some developing countries have difficulties to achieve that goal due to the poor economy and conservative society, especially the Least Developed Countries. So the United Nations has been promoting other countries and international organizations to give support to these countries all the time. However, the support is limited. So we need to develop a method to evaluate the sustainability and predict the development trend of a country, which can help us make the most of the support and investment and help the country's most in need.

### 2. Ecosystem Health Index (EHI) model

EHI Model is a quantitative model to assess the state of the ecosystem health. The scope of EHI is from zero to one hundred. It was assumed that when EHI is zero, the healthy state is “worst”; when EHI is 100, the healthy state is “best”.

Similar to EHI, a Sustainability Index (SI) is defined to assess the sustainability quantitatively. In order to facilitate the description on sustainability, SI is equally divided into five segments or ranges: 0-20, 20-40, 40-60, 60-80, 80-100, which correspond with the five sustainable states, “worst”, “bad”, “middle”, “good”, and “best” respectively.

The SI Value can be calculated by using the following equation based on the calculation method of EHI,

$$SI = \sum_{j=1}^m (P_j \cdot E_j) \quad (1)$$

where  $SI$  is a synthetic index,  $E_i$  is the  $j$ th indicator which reflects the state of one specific domain,  $\varepsilon_j$  is the weighting factor for the  $j$ th indicator.

When assessing the sustainability of country, lots of statistics will be used to calculate the value of SI. All the data shall be collected in the same period of time. Namely, the Sustainability Index is also determined by the data collection time  $T$ .

Define

$$SI[T] = f[(\varepsilon, E), T] \tag{2}$$

Where  $SI [T]$  is the Sustainability Index of T time,  $T$  refers the data collection time.

When it is calculated based on the data in a certain year, the value of SI can just describe the sustainability in the corresponding year. The data of the adjacent year has no effect on the result calculated in the specific year.

### 3. Selection of the Indicators

To quantify the model and help us to do assessment, the study in this section focuses on the indicators and their weighting factors. Three main parts are included in this section: Selection, Standardization and Weight Determination of the Indicators.

The overarching purpose of the CSD indicators has been to inform policy at the national level. So it is appropriate and reasonable to base on the CSD indicators to build our own indicator system.

The CSD indicator system includes more than fifty core indicators of sustainable development. These indicators cover three domains and fourteen themes in total.

Table 1: The themes of CSD indicator

CSD indicator themes		
Social dimension	Ecologic dimension	Economic dimension
Poverty   Governance   Health Education   Demographics	Natural hazards   Atmosphere Land   Freshwater   Biodiversity Oceans, seas and coasts	Economic development Global economic partnership Consumption and production patterns

### 4. Standardization of the Indicators

The dimensions of the indicators differ in thousands of ways. So it is indispensable to standardize the data to enhance the comparability. Nondimensionalization is the most important part.

At present, there are mainly two ways to nondimensionalize the data.

One way is to use the ideal value or reference value as the basis to deal with the data, and then get the relative value of each index. This method can accurately quantify the difference between the indexes, but cannot directly reflect the significance.

In the other method, people firstly define the evaluation standard of different level, and the actual value is replaced by the average of the range it belongs to. This method can intuitively reflect the significance of the evaluation index, but as a result of using the average of the standard range, it is difficult to accurately quantify the differences.

A new integrated method is proposed which combines advantages of the previous two methods. In this method, the evaluation standard of different level shall be defined at first. Then the indicators can be standardized by using a linear fitting normalized function. When the original value of one indicator is not in the domain of this function, its standardized values is 100 (maximum) or 0 (minimum).

The standardization process of the indicator “GDP Per Capita (2013)” is as follows:

GDP Per Capita of a country reflects its economic sustainability in some degree, and they are positively correlative. Based on the trend of this curve, we define a raw criterion:

Table 2: GDP Standard data

Original Data (X)	3,000	5,000	10,000	45,000	70,000
Standard Value (Y)	20	40	60	80	90

The fitting result in Matlab is as follows

$$Y_{GDPPC} = 0.000865 * X + 34.99 \quad (3)$$

Where  $Y_{GDPPC}$  refers to the standard value of the GDP Per Capita.

$X$  refers to the original data.

By using Equation (3), the original GDP per capita can be transformed to a standard value. The result of Luxembourg is 103 which has been revised to 100.

**5. Every indicator can be handled in the same way. Then all the data will be standardized into dimensionless values.**

Weight determination of the Indicators

In fact, although all the indicators play various roles in the sustainability assessment, but the importance of each indicator is not the same. So in order to assess the sustainability of a country, to determine the weighting factors of each indicator is also a necessity. We use the Principal Component Analysis (PCA) method to calculate the weight of each indicator.

Firstly, calculate the Correlation Coefficient matrix  $\mathbf{R}$ .

$$\mathbf{R} = (r_{ij})_{m \times m} \quad (4)$$

Where  $m$  stands for the amount of the indicators

$$r_{ij} = \frac{\sum_{k=1}^n \tilde{a}_{ki} \cdot \tilde{a}_{kj}}{n-1}, (i, j = 1, 2, \dots, m)$$

$r_{ij}$  stands for the correlation coefficient between the  $i$ th and the  $j$ th indicator, and  $r_{ii} = 1, r_{ij} = r_{ji}$

$\tilde{a}_i = \frac{a_i - \mu_i}{s_i}$  Stands for the standardized value of the indicators, where  $\mu_i$  is the mean and  $s_i$  is

the standard deviation.

The eigenvalues of  $\mathbf{R}$  is  $\lambda_i$  and its corresponding eigenvector is  $\mathbf{u}_i = [u_{1i}, u_{2i}, \dots, u_{mi}]^T$ , in which  $i = 1, 2, \dots, m$ .

As a result, a new set of indicators  $b_i$  is proposed based on these eigenvectors.

