*2007 International Nuclear Atlantic Conference - INAC 2007*
*Santos, SP, Brazil, September 30 to October 5, 2007*
*ASSOCIAÇÃO BRASILEIRA DE ENERGIA NUCLEAR - ABEN*
*ISBN:* 978-85-99141-02-1

# CORRELATION ANALYISIS OF ENERGY INDICATORS FOR SUSTAINABLE DEVELOPMENT USING MULTIVARIATE STATISTICAL TECHNIQUES

**Alvaro Luiz Guimarães Carneiro[1] and Francisco Carlos Barbosa dos Santos[2]**
[1] Instituto de Pesquisas Energéticas e Nucleares (IPEN / CNEN - SP)
Av. Professor Lineu Prestes 2242
05508-000 São Paulo, SP
carneiro@ipen.br

[2] Fundação Instituto de Pesquisas Econômicas – FIPE/USP
Av. Prof. Luciano Gualberto, 908
05508-900 São Paulo, SP
fcarlos@fipe.org.br

## ABSTRACT

Energy is an essential input for social development and economic growth. The production and use of energy cause environmental degradation at all levels, being local, regional and global such as, combustion of fossil fuels causing air pollution; hydropower often causes environmental damage due to the submergence of large areas of land; and global climate change associated with the increasing concentration of greenhouse gases in the atmosphere. As mentioned in chapter 9 of Agenda 21[1], the Energy is essential to economic and social development and improved quality of life. Much of the world´s energy, however, is currently produced and consumed in ways that could not be sustained if technologies were remain constant and if overall quantities were to increase substantially. All energy sources will need to be used in ways that respect the atmosphere, human health, and the environment as a whole. The energy in the context of sustainable development needs a set of quantifiable parameters, called indicators, to measure and monitor important changes and significant progress towards the achievement of the objectives of sustainable development policies. The indicators are divided into four dimensions: social, economic, environmental and institutional. This paper shows a methodology of analysis using Multivariate Statistical Technique that provide the ability to analyse complex sets of data. The main goal of this study is to explore the correlation analysis among the indicators. The data used on this research work, is an excerpt of IBGE[2] *(Instituo Brasileiro de Geografia e estatística)* data census. The core indicators used in this study follows The IAEA *(International Atomic Energy Agency)* framework: Energy Indicators for Sustainable Development[3].

## 1. INTRODUCTION

When choosing energy fuels and associated technologies for the production, delivery and use of energy services, it is essential to take into account economic, social and environmental consequences. The fundamental point is the need of methods for measuring and assessing the current and future effects of energy use, determining whether energy use is sustainable and , if not, how to change it so that is. This is the main purpose of energy indicators, which address important issues within three of the major dimensions of sustainable development: economic, social and environmental. Changes in the indicator values over time mark progress or lack of progress towards sustainable development. Sustainable development has been defined by the Brundtland Commission[4] as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs”.

## 2. OBJECTIVE

A core set of energy indicators proposed by IAEA[5] with contributions from UNDESA[6] *(United Nations Department of Economic and Social Affairs)* and IEA[7] *(International Energy Agency)* for monitoring progress towards sustainable development through the implementation of Agenda 21 is somewhat complex. In the core set, the indicators are grouped in categories covering the economic, social and environmental aspects, with institutional considerations. The set of indicators follows the framework devised by the United Nations´ ongoing Work Programme on Indicators of Sustainable Development[8] *(WPISD)* called DSR, Driving Force/State/Response, that is widely applied for indicators. The framework incorporates three pillars of sustainable development as mentioned before. There are some interrelations between sustainability dimensions in the context of the energy system.

The Figure 1 shows the interrelations illustrating in a simplified way, the environmental state associated with the energy system results from the impact of driving forces originating from the economic and social dimensions of the energy system. The social state of the energy system is influenced by driving forces originating from the economic dimension of the energy, and the institutional dimension affecting the three dimensions through corrective response policy actions affecting the sustainability of the whole energy system.

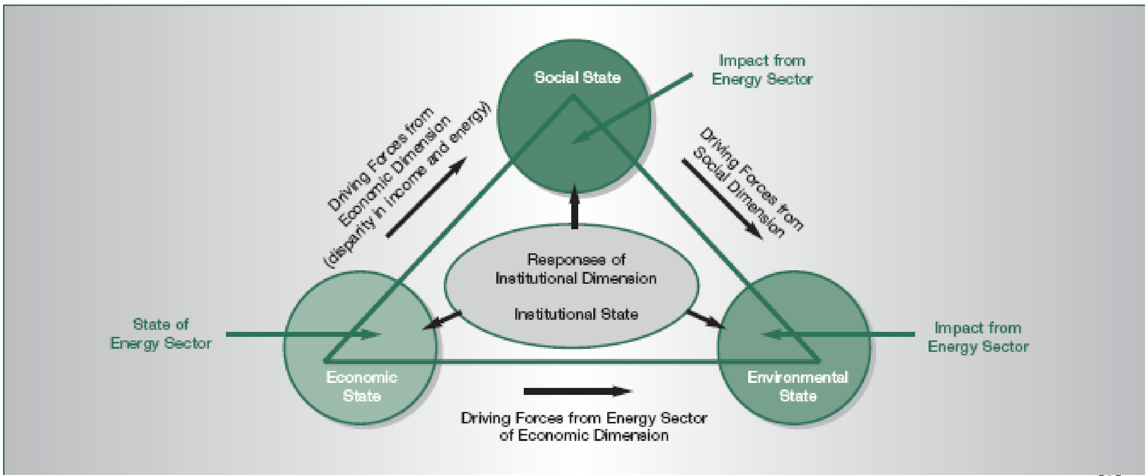


**Figure 1: Interrelations between sustainability dimensions of the energy system[9].**

There is an evidence of the complexity of the data indicators analysis. The main aim of this research work is to explore the analysis of the correlation among the indicators, concentrating information collection, using Multivariate Statistical Technique through the cluster analysis, that is a powerful tool for analysis of data that exhibit “natural” groupings. Cluster analysis is thus a tool of discovery. It may reveal associations and structure in data which, though not previously evident, nevertheless are sensible and useful once found.

## 3. DATA

The indicators in the Energy Indicators for Sustainable Development[3] *(EISD)* core set are presented in Table 1, according to dimensions, themes and sub-themes following the same conceptual framework used by the United Nations Commission on Sustainable Development *(CSD)*. There are 30 indicators, classified into three dimensions (social, economic and environmental), subdivided into 7 themes and 19 sub themes. The data indicators cover the years from 1995 to 2005. There are some indicators that could not be included in the analysis for lack of data information.

**Table 1. Energy Indicators for Sustainable Development *(EISD)*.**

| THEME | SUB-THEME | ENERGY INDICATOR | |
|---|---|---|---|
| **SOCIAL** | | | |
| Equity | Accessibility | SOC1 | *Share of households without electricity* |
| | Affordability | SOC2 | *Share of household income spent on fuel and electricity* |
| | Disparities | SOC3 | *Household energy use for each income group and corresponding fuel mix* |
| Health | Safety | SOC4 | *Accident fatalities per energy produced by fuel chain* |
| **ECONOMIC** | | | |
| Use and Production Patterns | Overall Use | ECO1 | *Energy use per capita* |
| | Overall Productivity | ECO2 | *Energy use per unit of GDP* |
| | Supply Efficiency | ECO3 | *Efficiency of energy conversion and distribution* |
| | Production | ECO4 | *Reserves-to-production ratio* |
| | | ECO5 | *Resources-to-production ratio* |
| | End Use | ECO6 | *Industrial energy intensities* |
| | | ECO7 | *Agricultural energy intensities* |
| | | ECO8 | *Service / commercial energy intensities* |
| | | ECO9 | *Household energy intensities* |
| | | ECO10 | *Transport energy intensities* |
| | Diversification (Fuel Mix) | ECO11 | *Fuel shares in energy and electricity* |
| | | ECO12 | *Non-carbon energy share in energy and electricity* |
| | | ECO13 | *Renewable energy share in energy and electricity* |
| | Prices | ECO14 | *End-use energy prices by fuel and by sector* |
| Security | Imports | ECO15 | *Net energy import dependency* |
| | Strategic Fuel Stocks | ECO16 | *Stocks of critical fuels per corresponding fuel consumption* |
| **ENVIRONMENTAL** | | | |
| Atmosphere | Climate Change | ENV1 | *GHG emissions from energy production and use per capita per unity of GDP* |
| | Air Quality | ENV2 | *Ambient concentrations of air pollutants in urban areas* |
| | | ENV3 | *Air pollutant emissions from energy systems* |
| Water | Water Quality | ENV4 | *Contaminant discharges in liquid effluents from energy systems including oil discharges* |
| Land | Soil Quality | ENV5 | *Soil area where acidification exceeds critical load* |
| | Forest | ENV6 | *Rate of deforestation attributed to energy use* |
| | Solid Waste Generation and Management | ENV7 | *Ratio of solid waste generation to units of energy produced* |
| | | ENV8 | *Ratio of solid waste properly disposed of to total generated solid waste* |
| | | ENV9 | *Ratio of solid radioactive waste to units of energy produced* |
| | | ENV10 | *Ratio of solid radioactive waste awaiting disposal to total generated radioactive waste* |

# 4. METHODOLOGY

## 4.1 The Cluster Analysis

The whole analysis developed in this work has been done using Statistica 6.0, Stat Soft Inc. The term cluster analysis[10] encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. A general question facing researchers in many areas of inquiry is how to organize observed data into meaningful structures, that is, to develop taxonomies. In other words, cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximum if they belong to the same group and minimum otherwise. The cluster analysis can be used to discover structures in data without providing an explanation/interpretation. The basic procedures on the clustering technique application involves: formulate the problem selecting the variables, select the algorithm of similarity measurement and the objects structure organized.

In this work has been used the *Euclidian Distance* as the similarity measurement and the *Dendogram* or *Hierarchical Tree* as the object structure organized.

### 4.1.1 The Euclidian Distance

The Euclidean Distance[10] between two points $P=(x_1,x_2,...,x_n)$ and $Q=(y_1,y_2,...,y_n)$ in Euclidian *n-space*, is defined as:

$$d_{\overline{PQ}}=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+...+(x_n-y_n)^2}=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}=\sqrt{(X-Y)'(X-Y)} \qquad (1)$$

The statistical distance between the same two observations is of the form:

$$d_{\overline{PQ}}=\sqrt{(X-Y)'A(X-Y)} \qquad (2)$$

Ordinarily, $A = S^{-1}$ , where S contains the sample variances and covariances. However, without prior knowledge of the distance groups, these sample quantities cannot be computed. For this reason, Euclidian distance is often preferred for clustering.

In a example the Euclidian distance between two values $P=(x_1,x_2,x_3)$ and $Q=(y_1,y_2,y_3)$ in a three dimensional space is showed in Figure 2:

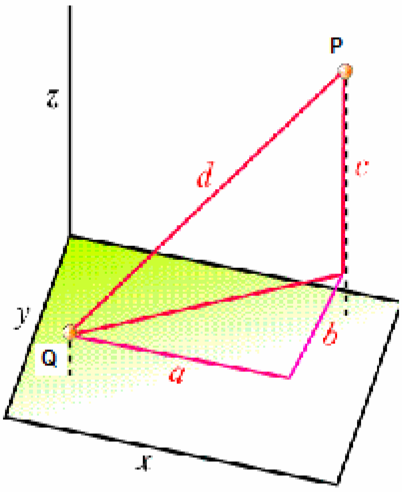


**Figure 2. The Euclidian Distance in three dimensional space**

This method start a initial single group of “objects” is divided into two subgroups such that the objects in one subgroup are “far from” the objects in the other. These subgroups are then further divided into dissimilar subgroups; the process continues until there are many subgroups as “objects” – that is, until each object forms a group.

### 4.1.2. Hierarchical Tree Diagram

The results may be displayed in the form of a two-dimensional diagram knowed as *hierarchical tree diagram*[11] or *dendogram*. As we shall see, the *dendogram*, illustrates the merger or divisions that have been made at successive levels.

The branches come together (merge) at nodes whose positions along a distance (or similarity) axis indicate the level at which the fusions occur.

The *hierarchical tree diagram* picturing the hierarchical clustering concluded in the analysis of the data purposed is shown in Figure 3. The groupings and the distance levels at which they occur are clearly illustrated in the *dendogram*.

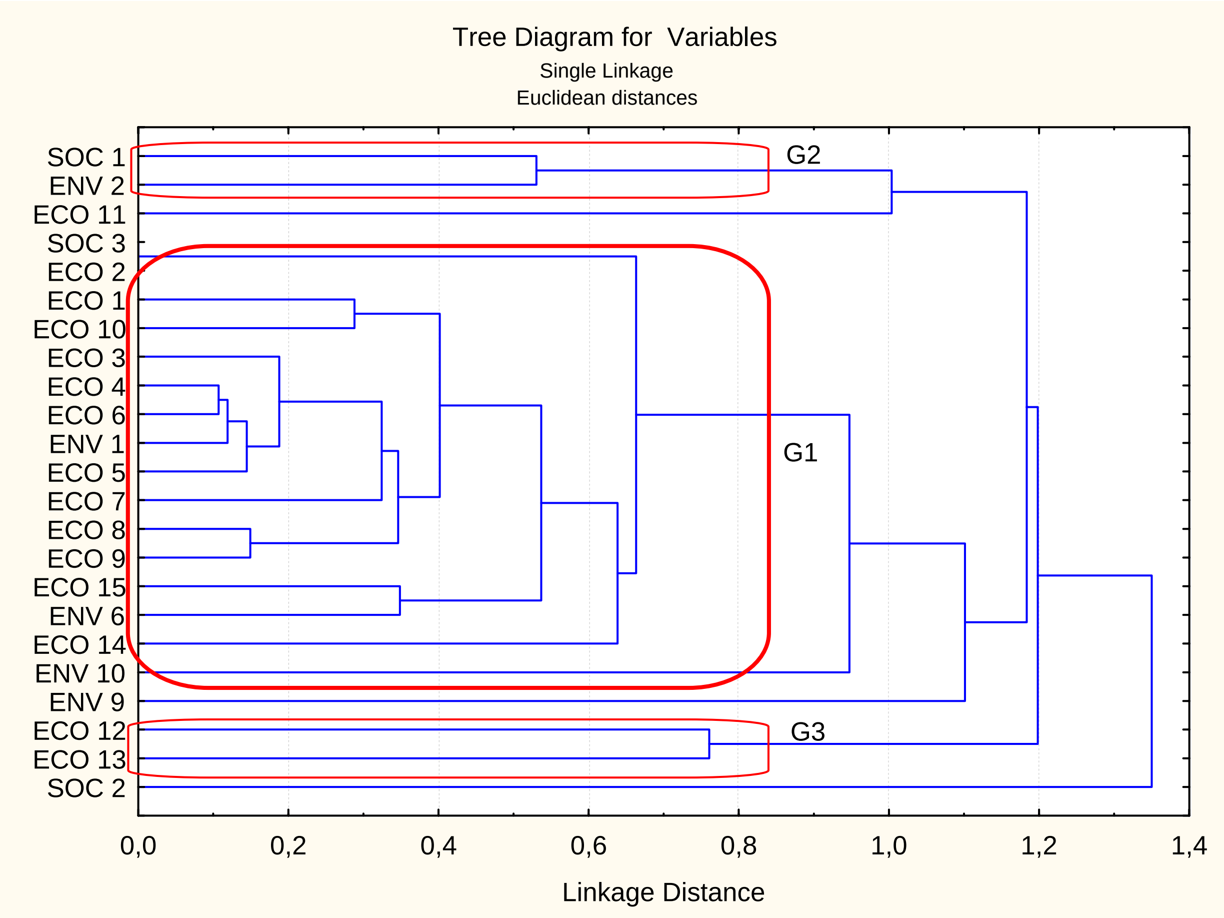


**Figure 3: The dendogram of the indicators analyzed.**

## 5. CONCLUSION

The results obtained for the energy indicators for the sustainable development analyzed using multivariate statistical technique, specifically cluster analysis, have presented three groups, being G1, the big one, grouping 16 indicators and the others smalls groups G2 and G3 grouping 2 indicators respectively. The cluster analysis has showed to be a useful tool for this kind of study evolving variables from different dimensions, social, economic and environmental with apparent similarities.

The aim of this paper does not intend to deep in the analysis of the context of the indicators it self, but explores the methodologies available for analysis of the data complex nature aiming to summarize the information into an appropriate set of indicators. Moreover, this research work ongoing intends in the second stage to purpose the elaboration of synthetic indicators helping in a friendly way the understanding and becoming it easier to decision-makers to see which programmes are necessary for sustainable development changing the current reality.

## 6. REFERENCES


[1] SENADO FEDERAL. Agenda 21 - Conferência das Nações Unidas sobre Meio Ambiente e Desenvolvimento. 3.ed. Brasília: Senado Federal, Subsecretaria de Edições, 2001. 598 p.

[2] IBGE-Instituto Brasileiro de Geografia e Estatística, www.ibge.gov.br.

[3] ENRGY INDICATORS FOR SUSTAINABLE DEVELOPMENT: GUIDELINES AND METHODOLOGIES, International Atomic Energy Agency (IAEA), Vienna/Austria april/2005.

[4] BRUNDTLAND, Gro Harlem. Nosso Futuro Comum. Editorta da FGV, 1987. *"Our Common Future"*, Oxford: Oxford university Press,1987.

[5] http://www.iaea.org/worldatom/Programmes/Energy/pess/ISED.shtml

[6] http://www.un.org/esa/desa/

[7] http://www.iea.org/

[8] http://www.un.org/esa/sustdev/natlinfo/indicators/isd.htm

[9] *"Indicators for Sustainable Energy Development"*, International Atomic Energy Agency Information Series, Division of Public Information, 02-01570 / FS Series 2/02/E.

[10] Hair J. F., Anderson R. E., Tatham R. L., Black W. C., *"Análise Multivariada de Dados"*, Bookman, 5ª Edição, 2005.

[11] Johnson, R. A., Wichern D. W., *"Applied Multiuvariate Statistical Analysis"*, 5th Edition, Prentice Hall, 2002.