EnviroInfo 2004 (Geneva)
Sh@ring – EnviroInfo 2004

Assessing Sustainability of EU Regions: The Case of the 'EPSILON' Tool

Marc Bonazountas[1], Yiannis Smirlis[2] and Despina Kallidromitou[3]

The EPSILON Project

The EPSILON project introduces a model and a tool to assess sustainability at regional level for the EU countries. The sustainability model uses a significant number of indicators organized in a four-level hierarchy. It is based on readily available data found in EU and international databases. A software tool integrates the EPSILON model into a Geographic Information System (GIS), operational via the Web. This paper presents different models and techniques which have been developed within the framework of the EPSILON project: a model for assessing regional sustainability, a model for measuring data quality, a space model including a new clustering algorithm and data generation through GIS techniques.

1. Introduction

Sustainable development is a global objective offering a positive long-term vision of a society that delivers a better quality of life for its people. The European Union (EU) has established an Internal Strategy on Sustainable Europe (EC-COM 2002) which must strike a balance between the environmental, economic, social and institutional objectives of a society. This balance optimises well-being at present without compromising the ability of future generations to meet their needs. In the framework of this strategy, the assessment, the reporting and the categorisation of sustainability of the EU regions via standardised indicators are of paramount importance.

In this paper we present the EPSILON[4] conceptual sustainability model and its software tool to assess sustainability at regional level (NUTS 2, 3 areas) for the EU countries, focusing on the technical and scientific innovations achieved. The pro-

[1] National Technical University of Athens, 5, Iroon Polytechniou, GR-15780 Zografos,Greece
[2] University of Piraeus, 80 Karaoli-Dimitriou Str, GR-18341 Piraeus, Greece
[3] Epsilon ltd, Monemvasias 27, Polydroso, GR-15125 Marousi, Greece
[4] EPSILON (Environmental Policy via Sustainability Indicators On a European-wide NUTSIII level) is an EU RTD-IST project - EC Contract No IST-2001-32389

posed sustainability model follows the well established DPSRI - Driving forces, Pressure, State, Response, Impact framework and employs a set of parameters, properly selected to cover the most important fields related to sustainability. It is structured as a 4x4x4 hierarchy consisting of four pillars (environmental, economic, social, and institutional), four themes, four sub-themes and a number of indicators and sub-indicators. The EPSILON model is based only on readily available data - EU (New Cronos, Corine, EMEP, UNFCC etc.) and international databases. Whenever parameters are not readily available, they are generated via mathematical, statistical or GIS computational procedures. Aggregation models are employed to estimate the values for the composite indicators, corresponding to sub-themes, themes and pillars of sustainability. The EPSILON model is integrated into a Geographic Information System (GIS), operational via the web. This GIS system allows the users to benchmark selected regions according to their performance on sustainability indicators by clustering their values into homogeneous groups and to reveal causality links to explain environmental and social and to aid the decision making on policies regarding the sustainability.

Different types of users can benefit from EPSILON: National Statistical Institutes (NSI's), Regional Statistical Institutes (RSI's), Eurostat, Environment Agencies, Ministries, local authorities, private and public organisations, decision makers, academia, and others engaged in regional sustainability assessments and reporting.

The structure of this paper is as follows: Section 2 presents the basic EPSILON model innovations such as the 4x4x4 structure of the sustainability model, the data quality model, an example of data generation procedures and the Space Model clustering model used in the GIS system. Section 3 presents some of the potential uses of EPSILON.

2. EPSILON Model Innovations.

During the development of EPSILON, a number of challenges have been encountered that led the partners to develop techniques and methodologies in scientific areas such as sustainability modeling, data quality, data generation through GIS techniques and statistical clustering. This section presents a summary of these techniques.

2.1 The Sustainability Model

The main objective of the EPSILON sustainability model is to benchmark NUTS 2 and eventually NUTS 3, regions for the EU 15 member countries by enabling assessing of regional evolution over time in different sustainability domains. It employs a comprehensive set of core indicators tracking social, economic, environmental and institutional dimensions to measure sustainability at State, Impact, Driving forces and Response levels. This set of indicators follows a hierarchical structure, which on the top defines the four pillars of sustainability indices: Environmental, Social, Economic and Institutional. Each one of the four pillars is divided in four themes. Each theme is further divided into four sub-themes and each sub-theme into a sufficient number of indicators and sub-indicators, thus comprising a 4x4x4 hierarchical structure. The total number of simple and composite indicators arranged in this structure exceeds the 250, which is the maximum used in the sustainability model literature so far. The structure of the EPSILON model at sub-theme level is presented in Table 1.

Table 1 : Themes and sub-themes of the EPSILON model

ECONOMIC		ENVIRONMENTAL	
Theme	**Sub-theme**	**Theme**	**Sub-theme**
Business efficiency		Air	
	Ease of funding		Pollutants with respiratory effects
	Labour productivity		Pollutants with carcinogenic effects
	IT skills & infrastructure		Noise
	Energy efficiency		Climate Change
Domestic performance		Biodiversity	
	Employment rate		Total area of wetlands
	Unemployment		Ratio of natural land over intensive agriculture
	Living standard		Wilderness area
	GDP Growth		Fragmentation of ecosystems and habitats
Innovation capacity		Soil & raw materials	
	Venture Capital attractiveness		Soil quality
	Labour market attractiveness		Land use
	Research & Development capacity		Mineral resources
	Jobs and companies creation		Fossil Energy use
Structural conditions		Water	
	Economic robustness		Oceans, sea and coasts
	Economic reforms		Surface water quality
	Connection facilities		Groundwater quality
	Business autonomy		Fresh water quantity
INSTITUTIONAL		**SOCIAL**	
Theme	**Sub-theme**	**Theme**	**Sub-theme**
Autonomy		Equity	
	Financial autonomy		Male/Female
	Legal autonomy		Youth / Elderly people
	Dependence on a neighbouring region		Low/high income

Public Debt	Equity - National/Foreigners
Citizens & civil society involvement	Human health & satisfaction
Political parties	Physical illnesses & injuries
Local election turnout	Psychological health
Local medias	Total mortality
NGOs	Physical illnesses & injuries
Government adequacy	Interpersonal relations
Democratic values	Social activities
Local responsiveness	Social exclusion
Civil society consultation	Professional relations
Agenda 21	Household and Facilities
Organisational outcomes	Knowledge
Spatial infrastructures	Secondary and vocational training
Security conditions	Information
Health care facilities	Basic Education
Security conditions	Higher education

A mathematical aggregation model estimates the composite indicators by using rationales for grouping and weighing of the different sustainability indicators taking advantage of the framework developed within the UNEP-United Nations Environmental Programme) - SETAC Life Cycle Initiative (UNEP/ SETAC Life Cycle Initiative, 2001).

The EPSILON sustainability model includes unique features that facilitate studying interactions between different regions. It enables the users to define the role of imports and exports in the impact generated within or outside the region and to study driving forces using both consumers' and producers' based distribution (e.g. steel can be produced in one specific part of a country, but consumed in other parts of Europe). Interrelations between the different dimensions of sustainability, using Input-Output Matrices can also be derived via multivariable algorithms. The sustainability model has been developed by the Swiss Federal Institute of Technology Lausanne - EPFL, Switzerland.

2.2 The Data Model

The EPSILON model is based only on readily available data - EU (New Cronos, Corine, EMEP, UNFCC etc.) and International databases. Whenever parameters are not readily available, they are generated via mathematical, statistical or GIS computational procedures. In addition to the challenges of finding, extracting and generating data, an indicator fact sheet is created into the data base containing information about the data definitions, sources, quality, availability and sustainability model parameters (weights, contribution to sustainability) for each indicator. The data model's main innovative areas are described in the sections below:

2.1.1 Assessing the Quality of the Data

The organizations that maintain the databases from which the data is extracted have already performed data quality management procedures and they guarantee the good quality to a great extent. However, some important EPSILON data quality issues should be considered for many reasons: data may have different definitions since it comes from various sources; data may have been collected by using different standards; has been pre-processed; harmonized, disaggregated etc. by special data processing techniques.

The most established and comprehensive way to define data quality is to analyse and associate it with multi-dimensional elements such as relevance, accuracy, accessibility etc. This approach is very common to standard frameworks for assessing data quality proposed by established working groups in National Statistical Services of EU countries (e.g. UK, Netherlands, Germany etc.) (UK Government Statistical Service, 1997), (Rudddock, 1999) and EUROSTAT. Many technical reports and documents that describe practices and methodologies for assessing data quality in various projects involving data collection, have already been presented by those working groups. The EU Leadership Expert Group on Quality and the Interest Group "Quality in Statistics" produce specific recommendations on how data quality should be defined and implemented within the European Statistical System (Franchet, 1999).

Some suggested quality criteria like Accessibility, Timeliness and Punctuality are irrelevant to the case of EPSILON data model. Therefore EPSILON redefined the EUROSTAT standard data quality framework to fit the EPSILON concept and to specify methods for measuring each particular quality criteria thus obtaining a specific total quality indicator. EPSILON proposes the quality criteria shown in Table 2.

Table 2 : The proposed data quality framework for the EPSILON data model

Quality Dimension	Definition and explanation
Relevance	The conceptual model of EPSILON provides a strict definition of each parameter. Relevance measures how close the actual data stored in the EPSILON database is to the indicator definition.

Accuracy	Accuracy gives an estimation of how close the reported and estimated data is to the real data. Assignment of accuracy scores considers both reports on data quality found in the EU databases and accuracy estimations concerning the data treatment procedures (transformations from GIS data, estimations for the missing data, other dissagregation procedures etc).
Comparability	This criterion will measure the difference in definitions and accuracy of the data between different NUTS regions and years. Thus comparability will be very high if data for all NUTS regions and years come from the same source database.
Completeness	Completeness is a pure quantitative measurement. It counts, in terms of percentage of the total number (%), the existing values for the NUTS regions.

Since these criteria are of qualitative nature, four ordinal levels, 1=excellent, 2=good, 3=average and 4=poor, are used to distinguish different quality levels in the EPSILON database. The ordinal scores are assigned to the criteria by the data analysts involved in the data collection and processing procedures. The data quality model also provides an aggregation model to obtain, for each one indicator, subtheme, theme and pillar, a total quality indicator. Data quality information is properly presented into the GIS tool.

The data model has been designed and developed by the National Technical University of Athens - NTUA, Greece.

2.1.1.1 Data Generation Through GIS Techniques

Due to limited availability of data at regional level, important data for the model is not found in databases and has to be generated by mathematical, statistical or GIS data manipulation procedures. A significant number of the environmental indicators were generated using GIS data manipulation procedures. In this section, an example of the estimation of Ozone concentrations at regional level by using a GIS procedure is presented.

Data about ozone are taken from the EMEP database (Schaug et. al. 1992). They are measured in a network of 155 stations across the countries of the whole Europe and in 106 stations located in the 15 EU states shown in Figure 1.

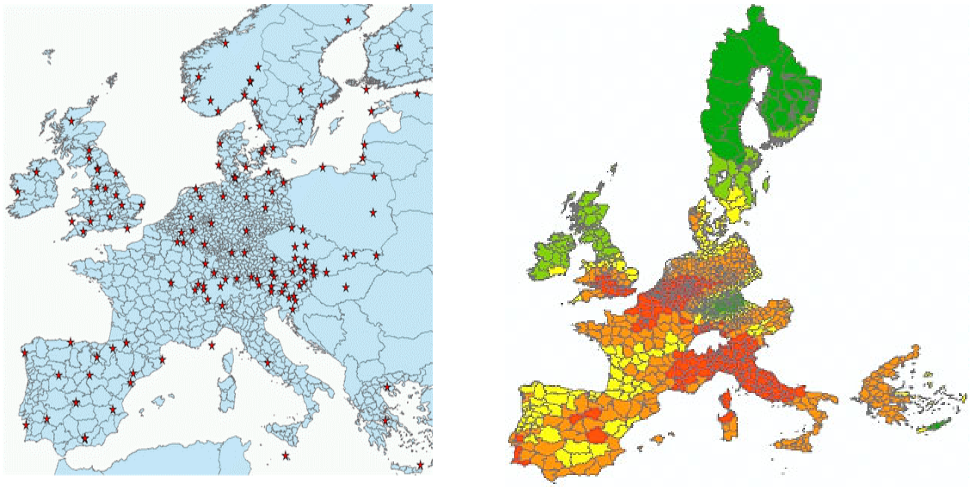

Figure 1: EMEP Ozone stations across Europe and the NUTS 3 regions map of Europe with clustered by the level of Ozone concentration

Since EPSILON requires regional data for ozone, a GIS based estimation has been applied as follows: the locations of the stations have been transformed from the Lat/Long coordinate system into decimal degrees and then imported to ArcGIS as a point theme shape file, which is then overlaid on the NUT3 regions shape file. For each one of the regions, the average annual ozone concentrations have been calculated using the Kriging interpolation method [Krivoruchko and Gotway, 2004] within the Geostatistical Analyst Extension of ArcGIS and then the resulted values were clustered in four distinct categories of ozone concentrations, ranging from ‘very high’ to ‘very low’.

This broad classification of the regions, although not very precise, gives the tendencies: higher O3 concentration are in the south Europe compared to the north and higher O3 concentration are found in big cities and the surrounding area than in rural areas. Figure 3 shows the NUTS 3 regions map of Europe with clustered by the level of Ozone concentration. The data generation procedures by using GIS techniques have been developed by Planungsbüro, PbS, Germany.

2.3 The Clustering Model

Clustering is needed in the EPSILON tool to provide to the users a more understandable vision of the data by grouping the NUTS 0,1,2 and 3 regions according to their performance on the sustainability indicators. The grouping is arranged so all the regions in one cluster have high similarity with each other but are very dissimilar to the regions in other clusters [Han and Kamber, 2001].

EPSILON developed a new Space Model clustering technique that unifies regions with similar performance on sustainability indicators. This can help to better explain the characteristics of the underlying data distribution and allow the identification of regions that behave as outliers. The core of the Space Model is a new clustering algorithm that is based on the k-nearest neighbour clustering technique and assumes no previous definition of the number of clusters (k value) or other initial parameters; it is more efficient than other data clustering techniques and clearly identifies the outliers. The Space Model is implemented within the GIS tool and allows the user to dynamically change the number of clusters so to have a better understanding of the performance of the regions. Figure 2 shows different views of the NUTS 3 regions in Europe according to the population density indicator.

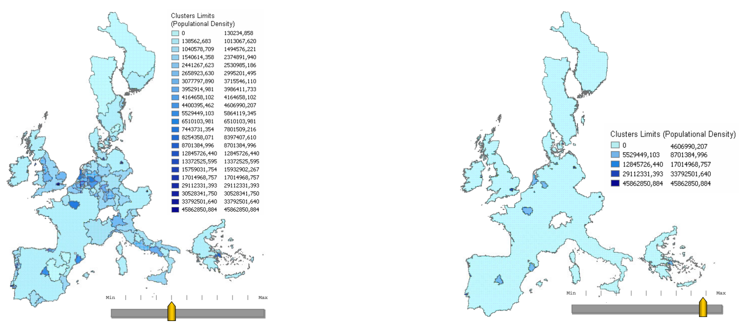

Figure 2 : Two different clustering results of NUTS 3 regions in Europe according to the population density indicator.

The Space Model and the clustering algorithm have been developed by University of Minho, Portugal.

3. EPSILON Potential Uses

EPSILON is a truly unique model and tool utilizing cutting edge technologies, such as clustering techniques, statistical permutations, algorithm creation and GIS techniques. It allows users to view and understand the different parts that all together make sustainability a viewable, comprehensible and workable entity. With the help of advanced technologies, the users could explore important indicators that raise or diminish the very important sustainability index.

The EPSILON tool will be available on the web and could be used to generate a great number of applications. Any potential user may combine the Economic, Environmental, Social and Institutional indicators presented above to benchmark specific EU regions, to examine the interrelation between regions and to identify the causes of an environmental, social, economic or institutional phenomenon. Moreover the EPSILON tool is able to measure various dimensions of the RESPONSE of a local government to problems related to sustainability.

EPSILON consists of a model platform for incorporating different kinds of sustainability statistical and mathematical models. Any interested research group may use the data base to test new models the output of which can directly be projected by the GIS system.

Through an exploitation questionnaire the interest of the potential user groups mentioned above in evaluating and using the tool is measured. So far positive answers have been received especially from the new European Member States who are not covered by the Epsilon Project. This demonstrates a big interest in sustainability, in collecting and processing the related indicators and in using the EPSILON tool to generate applications measuring the sustainability indices of the 4 pillars.

Bibliography

A Sustainable Europe for a Better World: AS European Union Strategy for Sustainable Development COM(2001)264 final, 15.5.2001 Brussels

EC-COM (2001) A Sustainable Europe for a Better World: A European Union Strategy for Sustainable Development COM(2001)264 final, 15.5.2001 Brussels

EC-COM(2002) Towards a global partnership for sustainable development: Communication from the Commission of the European parliament, the Council, the Economic and Social Committee and the Committee of the Regions, COM(2002)82 final, 13.2.2002, Brussels

Franchet Yves (1999) : The Quality of the ESS, Statistical Office of the European Communities

Han Jiawei and Kamber Micheline, 2001: Data Mining: concepts and techniques, Academic Press, 2001.

Krivoruchko K., Gotway A. C, (2004) :Creating Exposure Maps Using Kriging, Public Health GIS News and Information, January,

Rudddock Vera (1999) : Measuring and Improving Data Quality, U.K Government Statistical Service

Schaug J., Pedersen U., Skjelmoen J.E. and Arnesen K. (1992) : Data Report 1992. Part 2: Monthly and seasonal summaries. EMEP/CCC-Report 5/94.