

Using the Self-Organizing Map to Visualize and Explore Socio-Economic Development

Mikael Collan, Tomas Eklund and Barbro Back

Åbo Akademi University, Turku, Finland

Abstract

The socio-economic situation of a country can be measured in a number of ways, often by looking at indicator values describing different aspects of the social and economic reality in the country in question. Socio-economic development can be measured and observed by comparing yearly indicator values. As a large number of indicators are often required to accurately assess the socio-economic development of a country, the dimensions of the analysis quickly become difficult. Composite variables, such as the Gini-coefficient and the Human Development Index, are often used, but drill-down analysis and visualization of the results reached with these is demanding and may be very laborious. The realm of data mining offers a number of tools for dealing with these problems.

This paper utilizes the self-organizing map (SOM), a two layer unsupervised neural network, to observe, compare, and visualize development in 25 transition economies. Using longitudinal (1998-2002), multidimensional (14 indicators) socio-economic data, the paper presents how

transition economies can be positioned, and their development tracked, with the SOM. The data is from the World Bank Group's World Development Indicators (WDI) on-line database. The paper discusses the advantages achieved by using the SOM over simply comparing yearly indicator values, in observing and visualizing socio-economic development.

The purpose of this paper is to show how the SOM can be used to observe, compare, and visualize development in transition economies. The paper does not seek to propose a new "correct" model for measuring socio-economic development, but to show, with examples, that the SOM could be a helpful tool in intuitively visualizing and comparing multidimensional socio-economic development.

Keywords: Multidimensional comparison; Socio-economic development; Self-organizing map; Transition economies

Introduction

In the years following the break-up of the former Soviet Union, much interest has been focused on the development of the newly independent countries. Recent history has seen precious few similar opportunities for scholars to study the simultaneous development of a large number of newly independent countries, all emerging from a similar economic and political system. Therefore, this has received

Dr. **Mikael Collan** (D.Sc. econ) is a Senior Management Consultant at Gutta Oy in Helsinki, and a Research Fellow at the Institute for Advanced Management Systems Research (IAMSR) at Åbo Akademi University in Turku, Finland. He received his doctoral degree from Åbo Akademi University in 2004 with the thesis entitled "Giga-Investments: Modelling the Valuation of Very Large Industrial Real Investments". His primary research interests include profitability analysis and valuation of assets, application of fuzzy logic in finance and e- and mBusiness models. His publications have appeared, e.g., in *Industrial Management and Data Systems* and *Journal of Decision Systems*.
E-mail: mikael.collan@abo.fi

Dr. **Tomas Eklund** (D.Sc. econ) is a post-doctoral researcher at the Institute for Advanced Management Systems Research (IAMSR) at Åbo Akademi University in Turku, Finland. He received his doctoral degree from Åbo Akademi University in 2004 with the thesis entitled "The Self-organizing Map in Financial Benchmarking". His primary research interests include data mining, self-organizing maps, financial benchmarking and performance analysis, text mining and risk analysis. His publications have appeared in *Information & Management*, *Information Visualization*, *International Journal of Intelligent Systems in Accounting, Finance and Management*, and *Benchmarking – An International Journal*.
E-mail: tomas eklund@abo.fi

¹ Corresponding author.

Prof. **Barbro Back** (D.Sc. BA) is Professor of Accounting Information Systems at the Department of Information Technologies at Åbo Akademi University in Turku, Finland. She received doctoral degree from Åbo Akademi University in 1991, and was appointed full professor in 1998. Her research interests are in the areas of data mining, neural networks, text mining, XBRL, financial benchmarking and enterprise resource planning systems. She has published in journals such as *Journal of Management Information Systems*, *Accounting, Management and Information Technologies*, *International Journal of Accounting Information Systems* and *European Journal of Operational Research*.
E-mail: barbro.back@abo.fi

great attention in the form of numerous research papers (including Fischer, Sahay et al., 1996; Lavigne, 2000; Murrell, 1996; Raiser, Schaffer et al., 2004). Initially, very little data was available to researchers, but the situation has now reversed and currently data concerning the former east block countries is abundant (Filer and Hanousek, 2002; Grosh and Glewwe, 1998). However, there is much ongoing discussion concerning the reliability and accuracy of the data.

Measuring the state of the socio-economic status of a country is interesting from the point of view of predicting social and health related outcomes for that country, such as fertility, health, mortality, and obesity (Higgs, 2002). Tracking the development of the socio-economic status can further enhance our predictive capabilities about the future of a nation. Higgs discusses the measurement taxonomy of socio-economic status, and reviews the measures and variables used for the determination of socio-economic status. For measuring aggregate data between countries, the variables suggested are GDP per capita, the Gini coefficient (a measure of inequality), and Human Development Index (HDI), which incorporate life expectancy, health, income, and education measures. In general, the reviewed variables integrate an aspect of the overall socio-economic situation into one value. Unfortunately, these indicators are sometimes alone used as a measure of the whole socio-economic situation. Our opinion is that using single, even composite, variables does not always illustrate the situation of a country in an unbiased way. Indeed, if possible, the most objective status estimation for countries, when measuring them comparatively against other countries, can probably be constructed by using a (rather large) number of indicator variables simultaneously.

Because of the increasing access to large amounts of data in electronic form, for example, concerning transition economies (Filer and Hanousek, 2002), it is possible to do this, however the importance of good visualization tools suitable for data exploration becomes apparent. Visualization is an important addition to understanding statistical data. Usually two-dimensional graphs are used and, at most, the data visualization is three-dimensional. When visualizing country-to-country comparisons based on a number of variables, 3D visualization does not allow more than three variables to be visualized without using composites of the variables. The problem with

using composite variables is that putting variables together means having implicitly, or explicitly, to assign them weights. Such weight assignments are always subjective, at least to some degree. This issue may become even more complicated if we wish to include the status of a given country for a number of different years, and at the same time compare it to the status of another country for each same year. Such analysis is more complex, however it has some considerable advantages:

‘If the set describes the state of the same system at different times like in the case of time series analysis, then relations of the states will be illustrated. If, on the other hand, the set describes the states of different systems, the systems will become comparable’
(Kaski and Kohonen, 1996: 499).

In this study, a combination of both of the aforementioned is considered. The data set includes longitudinal (multi period) data on different systems (countries), which means that the differences between countries and the development in their own socio-economic situation can be compared.

This study utilizes computer-based methodologies for analysis. IT tools are increasingly being used for the visualization and analysis of multidimensional data. In this paper, we propose the use of the self-organizing map (SOM) for visualization and exploration of socio-economic data. The SOM is an unsupervised neural network that can be used for visual clustering of data. The SOM allows both analysis of the relation of states within a country, and comparison of these with other countries at the same time and according to a number of variables simultaneously. We present the SOM method, and discuss how the SOM has previously been applied in some cases with a macroeconomic focus. Finally, we illustrate using a case study concerning the socio-economic development of transition economies during the period 1998-2002.

The paper does not aim to propose a new “correct” combination of socio-economic variables for the measurement of socio-economic status for a given nation. The number, or the type, of the selected variables has not been exhaustively analyzed to give an optimal picture of the socio-economic country status. The selected variables rather act as a placeholder for such an optimal bouquet of variables, and illustrate how a rather large number of variables can be used simultaneously with the SOM in an easy to understand and tractable way

Methodology

The SOM is a two-layer unsupervised neural network that maps multidimensional data onto a two dimensional topological grid or map (Kohonen, 2001). The data is grouped according to similarities and patterns found in the dataset, using some form of distance measure, usually the Euclidean distance. The result is displayed as a series of nodes or points on the map which can be divided into a number of clusters based upon the distances between the nodes. As the SOM is unsupervised, no target outcomes are provided, and the SOM is allowed to freely organize itself based on the patterns identified, making the SOM an ideal tool for exploratory data analysis. “Exploratory data analysis methods, like SOM, are like general-purpose instruments that illustrate the essential features of a data set, like its clustering structure and the relations between its data items” (Kaski and Kohonen, 1996). Thus, the SOM can be said to perform visual clustering of data.

A simple illustration of a SOM is provided in Figure 1.

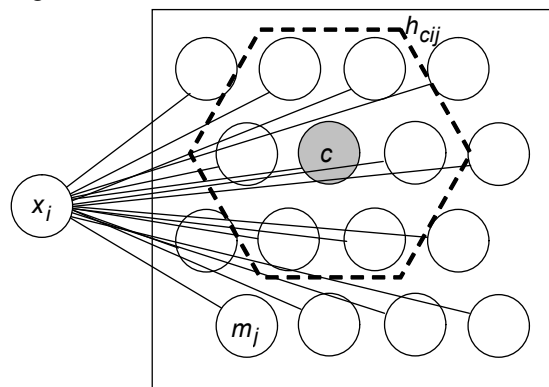


Figure 1 The SOM grid (adapted from Haykin, 1999: 445)

The SOM training process can be described as follows (Haykin, 1999; Kaski and Kohonen, 1996; Kohonen, 2001; Vesanto, 2002). Essentially, the SOM consists of a grid of neurons, j , each containing a parametric reference vector, m_j . The reference vector is of the same length as the data vectors, i.e. it contains as many attributes as the data being compared. After the network size is defined, the network is initialized, i.e., the reference vectors are assigned to the neurons. This is done either randomly, or by using some kind of data feature extraction, such as linear initialization, which extracts the two principal components

based upon their variances. After initialization, the algorithm proceeds by repeating two steps.

The algorithm works as follows. In the first step, a random data vector x_i (i.e., a row of data) is compared to the reference vectors m_j and the best matching neuron c according to a specified distance measure, in the case the *Euclidean distance* $\|x_i - m_j\|$, is identified using the formula in Eq. 1:

$$\|x_i - m_c\| = \min_j \{ \|x_i - m_j\| \} \quad (\text{Eq.1})$$

The neuron c is termed the winning neuron. The SOM algorithm is often called competitive learning, as each neuron competes to attract as much data as possible (Haykin, 1999).

In the second step, the winning neuron c and the neurons within a certain geometric distance of it, h_{cji} , “learn” from the attracted row of data x_i according to a specified learning rate $\alpha(t)$.

$$m_j(t+1) = m_j(t) + \alpha(t) h_{c,j}(t) [x_i(t) - m_j(t)] \quad (\text{Eq. 2})$$

where t is the training step, $\alpha(t)$ is the learning rate, and $h_{cji}(t)$ is the neighbourhood centred around the winning neuron c . The learning rate $\alpha(t)$ decreases from the centre of $h_{cji}(t)$, for example, using a Gaussian function.

These two steps are repeated a specified amount of times, or until a cut-off threshold is reached, in terms of the accuracy of the mapping. In simple terms, the algorithm first finds the neuron most similar to the input data, and then the winning neuron and its surrounding neurons learn from the data becoming more similar and further attracting similar data. This will result in clusters of data most similar to each other on the map.

The SOM differs from statistical clustering methods in a number of ways however it is similar to k-means clustering. Firstly, when using the SOM the targeted number of clusters does not have to be defined. Secondly, the SOM is more tolerant towards data that does not follow a normal distribution. Thirdly, the SOM is quite efficient and is

faster than most top-down hierarchical clustering methods (Vesanto and Alhoniemi, 2000). Finally, the SOM is a very visual method, as opposed to many statistical methods.

The primary application area of the SOM has been in engineering (Oja, Kaski et al., 2003), in applications such as process monitoring (e.g., Alhoniemi, Hollmén et al., 1999; Simula, Vasara et al., 1999), but it has also been applied in various other fields, including medicine (e.g. breast cancer diagnosis, Chen, Chang et al., 2000) and text retrieval (Kohonen, Kaski et al., 2000; Visa, Toivonen et al., 2000). The SOM has also been used for a variety of tasks relating to financial analysis, for example, credit analysis (Back, Oosterom et al., 1995; Kiviluoto, 1998; Martín-del-Brío and Serrano-Cinca, 1993; Serrano-Cinca, 1996; Tan, van den Berg et al., 2002), financial benchmarking (Back, Irjala et al., 1998; Eklund, Back et al., 2003), macroeconomic analysis (Costea, Kloptchenko et al., 2001; Kasabov, Erzegovezi et al., 2000), and macro environment analysis (Lämsiluoto, Eklund et al., 2004).

A related study was presented in Kaski and Kohonen (1996). In the study, the authors grouped a number of countries according to 39 welfare indicators, such as life expectancy at birth, adult illiteracy, and infant mortality rate. The result of the study was a map on which welfare decreased from the left side of the map (OECD countries) to the right (the poorest African countries). Eastern Europe and South America formed their own clusters somewhere between these. Kasabov, Erzegovezi et al. (2000) used the SOM to analyze the annual development of 15 countries according to five economic indicators (DBT/GDP, DEF/GDP, inflation rates, interest rates, and unemployment). Costea, Kloptchenko et al. (2001) used the SOM to cluster the performance of a number of eastern European countries, also according to primarily economic indicators.

The map in Kaski and Kohonen (1996) constituted a general model of international welfare. Although it is similar to this study, there are a number of important differences. We are interested in the economic development of a selection of transition economies, and have placed an emphasis on the development over a number of years. The focus on the development of specified countries and the temporal dimension are differ-

ent from the Kaski and Kohonen study. Kasabov, Erzegovezi et al. (2000) and Costea, Kloptchenko et al. (2001) are more similar to the application in this study. However, we focus on a mixture of social and economic indicators, and specifically on the socio-economic development of transition economies.

Socio-Economic Development in Transition Economies

In this section we demonstrate, with an illustrative case study, how the development of countries can be compared and visualized with the SOM. A “development map” is built by using a set of statistical indicators. The map presents some of the differences between the state of development between and within countries at different times, according to the selected indicators. The relations between different states (yearly state) of each country are visible as geometrical relationships on the map. The longer the distance between two states, the more different the states generally are. Because longitudinal data is used on each country, the development of countries can be measured and visualized for the period of analysis. Development can be seen as changes in the state of a country on the map.

The socio-economic development of a country, especially in rapidly changing transition economies, can be of interest to many different parties, all of whom have different purposes for their analyses. For example, potential investors are primarily interested in economic risk/return scenarios, but may also be interested in social factors, such as income distribution and availability of schooled labour. International institutions, such as the World Bank, are interested in the socio-economic development of countries, both for identifying necessary support initiatives, as well as for following up current ones. Political leaders are interested in the development of their own country relative to others in a similar position, a form of country benchmarking. The intent of the analysis, of course, determines the variables used. In this case, a general model of socio-economic development will be used to illustrate the use of the SOM in this area. The variables selected do not constitute an ideal or optimal set for studying the socio-economic status of a nation, rather they have been selected on basis of data completeness and so that different types of indicators relevant to socio-economic development are present.

The Data Set

The data set was retrieved from the World Bank Group's World Development Indicators (WDI) on-line database. Data was selected for 27 countries, including 25 transition economies and two European benchmark economies (Germany and Finland). In addition, two classification indicators (European Monetary Union average (EMU) and Heavily Indebted Poor Countries (HIPC)), were included. The countries included are summarized in Table IV. Fourteen (14) variables covering social, technological, financial, and general economic situation were selected to model the state of the socio-economic system in each of the selected countries for the period of five years (1998-2002). The selection of the variables was not based on a thorough analysis on which variable combination would yield the best possible descriptive power on the state of the country, but rather on an ad hoc selection of social, technological, financial, and economic indicators. Higgs (2002) points out that very often the choice of variables is limited by what is available in a dataset. The indicators for which the data were most incomplete were discarded, a choice that narrowed down the number of usable indicators, but enhanced the consistency of the results obtained with the selected indicators. The emphasis in the selection of variables is slightly on financial and economic data. The selected indicators are listed in Table 1.

Table 1 Variables used in the creation of the SOM

<i>Social and Technological Indicators</i>	<i>Economic and Financial Indicators</i>
Life expectancy (years)	GDP growth (% / year)
Population growth (% / year)	Inflation (% / year)
Fixed phones & mobile subscribers (/ 1000 ppl)	Overall budget balance (% / GDP)
Personal computers (/ 1000 ppl)	Gross capital formation (% / GDP)
Aid per capita (/ 1000 ppl)	Total debt service (% / goods & services)
High technology exports (% of manuf. exports)	Import of goods & services (% / GDP)
	Exports of goods & services (% / GDP)
	Trade in goods (% / GDP)

In total, the set of data used consisted of 145 rows, with 14 variables for each. Some variables in the original WDI database, such as infant mortality rates and poverty headcount, had to be discarded due to the substantial amount of missing data. The SOM is capable of handling small amounts of missing data, however, if too much data is missing the variables may become unnecessarily strong classifiers, which may result in the formation of separate clusters containing only cases for which these particular data are available. In some cases (such as with binary attributes) such an effect is preferable, but in our case it was not desired. Comparability of variables demands the data to be in the form of ratios, hence absolute values such as total population or GDP cannot be used without pre-processing.

The data does not follow a normal distribution, as is typical for financial indicators (Salmi and Martikainen, 1994). In addition, the variance between countries can be high, thus scaling of the data was important. In this case the data has been normalized according to the variance. In addition, sigmoid transformation has been used to handle outliers, which are common in financial indicators. Sigmoid transformation creates a more balanced distribution by stretching the centre of the histogram and squeezing the ends (Deboeck, 1998a). Transformation should, however, be used with caution since it may change the relationships of the data.

The software used for creating the SOM was Viscovery SOMine (Ver. 4.0 build 532). There are several different SOM software packages available for academic use, and some examples are shown in Table 2. Readers are referred to Deboeck (1998b) for a detailed discussion concerning some of the different available SOM software packages.

Table 2 Some of the available SOM software packages

<i>Software package</i>	<i>Available from</i>
SOM_PAK 3.1	http://www.cis.hut.fi/
Nenet v1.1	http://koti.mbnet.fi/~phodju/nenet/Nenet/General.html
Viscovery SOMine 4.0	http://www.eudaptics.au
eSom	http://www.ellipse.fi
SOM Toolbox for Matlab	http://www.cis.hut.fi/projects/somtoolbox/

The map was created using a network size of 2000 nodes and a tensil value of 0.5 (neighbourhood range indicator, low values produce more locally detailed maps, whereas higher values produce more general, coarser maps). The size of the map is influenced by the purpose of the study. If clustering is desired, a small map is created (Deboeck, 1998a). If visualization and detail are desired, a larger map is created. We, therefore, chose a large map, with a large number of nodes available to be fine-tuned to specific countries, allowing for greater detail and accuracy. Once the map is trained, the software package includes a function for determining a “natural” amount of clusters for the map. This approach, two-level clustering, was proposed by Vesanto and Alhoniemi (2000), and essentially consists of objectively determining the number of clusters on the map using a hierarchical clustering algorithm. In this case, six clusters were determined to provide sufficient detail for our purposes.

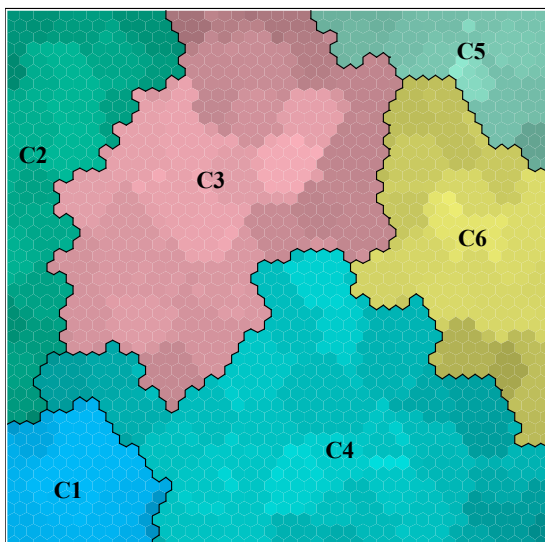


Figure 2 The clustered map

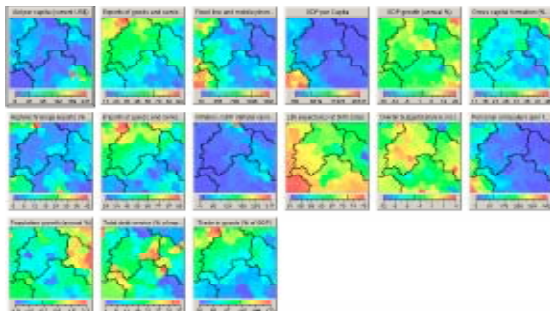


Figure 3 The component planes of the map

Results

In Figure 2, the final map and the identified clusters are displayed. The characteristics of the different clusters are identified using the component planes in Figure 3, which show the values of the different variables as they are distributed across the map. Warm colours on the component planes illustrate high values, whereas cool colours illustrate low values, in black and white warm colours are mostly lighter.

Figure 2 shows that the 145 sets of data, one for each year and each country, have formed clusters according to similarity. Different clusters represent different types of socio-economic states. Table 3 summarizes the characteristics of each cluster based upon the component planes in Figure 3. The table, therefore, illustrates the different types of socio-economic states that are predominant in each cluster, as well as examples of countries that can be found in a particular cluster.

For example, looking at the component planes in Figure 3, we can see that the countries in Cluster 1 display characteristics such as low inflation, very high GDP per capita, average GDP growth, very high life expectancy, and a high number of phones and computers per capita. The countries in this cluster are western European countries. Cluster 6, on the other hand contains countries that display very high population growth, low GDP per capital, low life expectancy, and a low amount of phones and computers per capita.

Table 3 Cluster characteristics

Cluster	Type of socio-economic state <i>(examples of countries)</i>
Cluster 1 (C1)	hi-tech, slow growth, long life (EU countries, Finland, Germany)
Cluster 2 (C2)	hi-tech, rapid growth, long life, high-export (Estonia, Czech Republic, Hungary)
Cluster 3 (C3)	medium-tech, medium growth, medium-life (Bulgaria, Lithuania, Ukraine)
Cluster 4 (C4)	low-tech, low trade, long life, miscellaneous (Georgia, Armenia, Poland)
Cluster 5 (C5)	low-tech, high growth, short life, very poor (Mongolia, Turkmenistan)
Cluster 6 (C6)	low-tech, medium growth, short life, poor (Kyrgyz Republic, Uzbekistan)

It is important to note that clusters are not equally uniform internally. Specialized clusters tend to form on the fringes of the map, such as the highly developed countries in Cluster 1 and the low development-level countries in Cluster 6, whereas the countries in the two general clusters (3 and 4) tend to display more variation.

In the following section, we will present two different types of countries; stable and rapidly changing. We will compare these to the benchmarks selected for the study.

Analysis of the Map

In this section, the use of the SOM for comparing the socio-economic development of countries will be illustrated using a few examples. The map containing all of the included data can be found in the Appendix. Studying the map reveals that countries not only show different development, but also that the pace of development in different countries displays substantial differences. Generally speaking, based upon the map countries can be divided into stable and mobile economies. In this study, we will provide examples of both types of development by looking at the development of Estonia (EST), Finland (FIN), Germany (DEU), Lithuania (LTU), Poland (POL), Romania (ROU), the Russian Federation (RUS), Turkmenistan (TKM), and the Ukraine (UKR). These are illustrated in Figure 4.

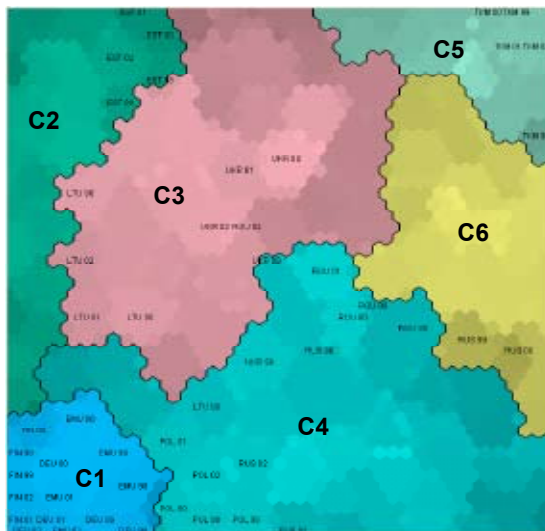


Figure 4 Stable countries on the map, with cluster indicators (C1 - C6)

Estonia (EST), Finland (FIN), Germany (DEU), Poland (POL), and Turkmenistan (TKM) are good examples of stable economies, although their socio-economic statuses differ considerably. Estonia (Cluster C2) is an example of a rapidly developing country that has already reached a considerable level of development compared to the other former Soviet countries. As we can see from the feature planes (Figure 2), exports, especially in high technology, are very high, and life expectancy and infrastructure are quite near the level of EMU countries (Cluster C1). GDP per capita, however, is still considerably lower. Trade in goods is still predominant in 2002, although the service sector was growing considerably.

Finland and Germany (Cluster C1) are examples of highly developed countries in which growth has slowed down considerably. In this case, this cluster is used as a benchmark to measure the development of transition economies against.

Poland (Cluster C4) displays fair social development according to measures such as life expectancy and technological infrastructure, but the economic development is lagging far behind the economies in Cluster C2. Exports and trade in goods is lower, and indebtedness is higher.

Turkmenistan is located in one of the poorest clusters (Cluster C5). Population growth is high, but life expectancy is fairly low. However, GDP growth and investments are very high. Cluster C5 contains several very rapidly growing economies.

Lithuania (LTU), Romania (ROU), the Russian Federation (RUS), and the Ukraine (UKR) are examples of countries displaying more or less rapid changes. Lithuania (Cluster C3) is developing towards Cluster C2, although in 2002 technological infrastructure and trade development is still behind that of the countries in Cluster C2.

Of the illustrated countries, the Russian Federation displays the largest changes in socio-economic status for the period. During 1999 – 2000, the primary drivers for the change from a poor part of Cluster C4 to the fringes of Cluster C6 are slight increases in GDP growth, overall budget balance, and exports. In 2001-2002, development is very positive. High technology exports, increased investments, and increased life expectancy are important drivers for the change.

Romania and the Ukraine display similar development ending up in virtually the same position in 2002. However, during 1998-2001 socio-economic development was much more rapid in the Ukraine than in Romania, especially in economic terms. In particular the budget balance, GDP growth, and exports are better in Ukraine.

Benchmarks	Transition economies	
Countries	Albania (ALB)	Moldova (MDA)
Finland (FIN)	Armenia (ARM)	Mongolia
Germany (DEU)	Azerbaijan (AZE)	Poland (POL)
Indexes	Belarus (BLR)	Romania (ROU)
European Monetary Union (EMU)	Bulgaria (BGR)	Russian Federation (RUS)
Heavily indebted poor countries (HIPC)	Croatia (HRV)	Serbia and Montenegro (SCG)
	Czech Republic (CZE)	Slovak Republic (SVK)
	Estonia (EST)	Slovenia (SVN)
	Georgia (GEO)	Tajikistan (TJK)
	Hungary (HUN)	Turkmenistan (TKM)
	Kazakhstan (KAZ)	Ukraine (UKR)
	Kyrgyz Republic (KGZ)	Uzbekistan (UZB)
	Lithuania (LTU)	

List of countries included in creating the map

Summary

The socio-economic state of a country is an interesting issue to research, due to the business and policy decision-making implications that the results may have. Comparing the status across several years and observing the changes gives us an indicator of the development of the country. Comparing development between countries reveals information about the differences between nations.

We suggest that the analysis or measurement of the socio-economic state of a country should be done with a number of indicators simultaneously to avoid assigning subjective weights to different indicators, and thus having a more objective result. Naturally the selection of indicators itself may be subject to a subjective selection. Visualization of the results should make it easier to understand the socio-economic state and development of a country in relation to other countries. Unfortunately the commonly used 2D and 3D graphs have limitations. The SOM offers a solution that solves many of the problems apparent in the use of 2D and 3D graphs, as well as some of the limitations of using composite variables.

Conclusions

In this study, the self-organizing map has been proposed as a tool for visualizing and exploring socio-economic data available in electronic form. Using a case example we have illustrated the capabilities of the SOM for visualization and exploration of socio-economic data.

The SOM is a tool capable of visualizing multidimensional data and provides a good way to analyze and present the socio-economic development of transition economies on a macro-economic level. We feel that the study shows that the SOM can provide an interesting alternative to the traditional and more frequently used methods, and in particular that visualization capabilities of the SOM are intuitive and easy to understand, making them superior to commonly used graphical representations of multidimensional data. The SOM is particularly suitable tool for visualizing and analyzing countries in phases or rapid development, such as transition economies, because the changes their direction are clearly visible, comparable, and easy to intuitively understand.

The purpose of this paper has not been to propose a new model for the measurement of socio-economic development, but to illustrate how the SOM can be used to extend other models and support new analyses that were, previously difficult to perform. As always, the choice of measures is a delicate problem, due to reliability issues with the measures available. Another problem is free or inexpensive access to data. However, using data sources such as the World Bank overcomes some of these issues.

This research should now be further developed by utilizing different validated models for measuring socio-economic development.

As the SOM seems to be a useful method in visualizing development, it might be a good idea if the World Bank Group would publish an independent SOM map with standard indicators for the world to enhance the intuitive understanding of development and direction of development in different countries.

Acknowledgements

The authors would like to thank the Academy of Finland (Domino, grant no. 104639) for their financial support of this research.



Appendix The final map with all countries labelled

References

Alhoniemi, E., Hollmén, J., Simula, O. and Vesanto, J. 1999. Process Monitoring and Modeling using the Self-Organizing Map. *Integrated Computer Aided Engineering*, 6 (1), 3-14.

Back, B., Irjala, M., Sere, K. and Vanharanta, H. 1998. Competitive Financial Benchmarking Using Self-Organizing Maps. In: Vasarhelyi, M. and Kogan, A. (eds.) *Artificial Intelligence in Accounting and Auditing. Towards New Paradigms*. Princeton, NJ: Marcus Wiener Publishers, 69-81.

Back, B., Oosterom, G., Sere, K. and van Wezel, M. 1995. Intelligent Information Systems within Business: Bankruptcy Predictions Using Neural Networks. In: Doukidis, G., Galliers, R.D., Jelassi, T., Kremer, H. and Land, F.F. (eds.). *Proceedings of The 3rd European Conference on Information Systems (ECIS'95)*. Athens, Greece, June 1-3, 99-111.

Chen, D., Chang, R. and Huang, Y. 2000. Breast Cancer Diagnosis Using Self-Organizing Map for Sonography. *Ultrasound in Medicine and Biology*, 26 (3), 405-411.

Costea, A., Kloptchenko, A. and Back, B. 2001. Analyzing Economical Performance of Central-East-European Countries Using Neural Networks and Cluster Analysis. In: Ivan, I. and Rosca, I. G. (eds.) *Proceedings of The Fifth International Symposium on Economic Informatics*. Bucharest May 10-13. Academy of Economic Studies Press, pp. 1006-1011.

Deboeck, G.J. 1998a. Best Practices in Data Mining Using Self-Organizing Maps. In: Deboeck, G. J. and Kohonen, T. (eds.) *Visual Explorations in Finance with Self-Organizing Maps*. Berlin: Springer-Verlag, 203-229.

Deboeck, G.J. 1998b. Software Tools for Self-Organizing Maps. In: Deboeck, G. J. and Kohonen, T. (eds.) *Visual Explorations in Finance Using Self-Organizing Maps*. Berlin: Springer-Verlag, 179-194.

Eklund, T., Back, B., Vanharanta, H. and Visa, A. 2003. Using the self-organizing map as a visualization tool in financial benchmarking. *Information Visualization*, 2 (3), 171-181.

Filer, R.K. and Hanousek, J. 2002. Data Watch: Research Data from Transition Economies. *The Journal of Economic Perspectives*, 16 (1), 225-240.

Fischer, S., Sahay, R. and Veigh, C.A. 1996. Stabilization and Growth in Transition Economies: The Early Experience. *The Journal of Economic Perspectives*, 10 (2), 45-66.

Grosh, M.E. and Glewwe, P. 1998. Data Watch: The World Bank's Living Standards Measurement Study Household Surveys. *The Journal of Economic Perspectives*, 12 (1), 187-196.

- Haykin, S. 1999. *Neural Networks - A Comprehensive Foundation*. Upper Saddle River, N.J: Prentice Hall International, Inc.
- Higgs, N. 2002. Measuring Socio-Economic Status: A Discussion and Comparison of Methods. Working paper. Johannesburg: University of the Witwatersrand.
- Kasabov, N., Erzegovezi, L., Fedrizzi, M., Beber, A. and Deng, D. 2000. Hybrid Intelligent Decision Support Systems and Applications for Risk Analysis and Prediction of Evolving Economic Clusters in Europe. In: Kasabov, N. (ed.) *Future Directions for Intelligent Information Systems and Information Sciences*. Heidelberg: Springer Verlag, 347-372.
- Kaski, S. and Kohonen, T. 1996. Exploratory Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World. In: Apostolos, P. N., Refenes, Y. A., Moody, J. and Weigend, A. (eds.) *Neural Networks in Financial Engineering*. Singapore: World Scientific, 498-507.
- Kiviluoto, K. 1998. Predicting bankruptcies with the self-organizing map. *Neurocomputing*, 21 (1-3), 191-201.
- Kohonen, T. 2001. *Self-Organizing Maps*. Berlin: Springer-Verlag.
- Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V. and Saarela, A. 2000. Self organization of a massive document collection. *IEEE Transactions on Neural Networks*, 11 (3), 574-585.
- Lämsiluoto, A., Eklund, T., Back, B., Vanharanta, H. and Visa, A. 2004. Industry Specific Cycles and Companies' Financial Performance - Comparison with Self-Organizing Maps. *Benchmarking: An International Journal*, 11 (4), 267-286.
- Lavigne, M. 2000. Ten years of transition: a review article. *Communist and Post-Communist Studies*, 33 (4), 475-483.
- Martín-del-Brío, B. and Serrano-Cinca, C. 1993. Self-organizing Neural Networks for the Analysis and Representation of Data: Some Financial Cases. *Neural Computing and Applications*, 1 (2), 193-206.
- Murrell, P. 1996. How Far Has the Transition Progressed? *The Journal of Economic Perspectives*, 10 (2), 25-44.
- Oja, E., Kaski, S. and Kohonen, T. 2003. Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum. *Neural Computing Surveys*, 3, 1-156.
- Raiser, M., Schaffer, M. and Schuchhardt, J. 2004. Benchmarking structural change in transition. *Structural Change and Economic Dynamics*, 15 (1), 47-81.
- Salmi, T. and Martikainen, T. 1994. A Review of the Theoretical and Empirical Basis of Financial Ratio Analysis. *The Finnish Journal of Business Economics*, 43 (4), 426-448.
- Serrano-Cinca, C. 1996. Self organizing neural networks for financial diagnosis. *Decision Support Systems*, 17 (3), 227-238.
- Simula, O., Vasara, P., Vesanto, J. and Helminen, R. 1999. The Self-Organizing Map in Industry Analysis. In: Lain, L. C. and Vemuri, V. R. (eds.) *Industrial Application of Neural Networks*. London: CRC Press, 87-112.
- Tan, R.G.H., van den Berg, J. and van den Bergh, W. 2002. Credit Rating Classification Using Self-Organizing Maps". In: Smith, K. and Gupta, J. (eds.) *Neural Networks in Business: Techniques and Applications*. Hershey: Idea Group Publishing, 140-153.
- Vesanto, J. 2002. Data Exploration Process Based on the Self-Organizing Map. Acta Polytechnica Scandinavia, Mathematics and Computing Series No. 115. Espoo: Helsinki University of Technology.
- Vesanto, J. and Alhoniemi, E. 2000. Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, 11 (3), 586-600.
- Visa, A., Toivonen, J., Back, B. and Vanharanta, H. 2000. A New Methodology for Knowledge Retrieval from Text Documents. In: Yliniemi, L. and Juuso, E. (eds.) *Proceedings of TOOLMET2000 Symposium - Tool Environments and Development Methods for Intelligent Systems*. Oulu April 13-14. Oulu: University of Oulu, 147-152.