SCARLED
Structural Change in Agriculture and Rural Livelihoods

Priority Area
"Sustainable management of Europe's natural resources"

CONTRACT No. SSPE-CT-2006-0044201 (STREP)
Project start: 01 January 2007
Duration: 36 months

DELIVERABLE 3.3
"Typology of rural areas in the new Member States"

Sabine Baum[1]

WP leader for this deliverable	**IAMO**
Partners involved	
Document status:	**Final Version**
Due date of deliverable:	**31 March 2008**
Date:	**16 April 2008**

Dissemination level (see DoW p. 27-30)		
PU	Public	☑
PP	Restricted to other programme participants (including the Commission Services	☐
RE	Restricted to a group specified by the consortium (including the Commission Services)	☐
CO	Confidential, only for members of the consortium (including the Commission Services)	☐

[1] The author gratefully acknowledges financial participation from the European Community under the Sixth Framework Programme for Research, Technological Development and Demonstration Activities, for the Specific Targeted Research Project "SCARLED" SSPE-CT-2006-044201.

Thanks are given to Erik Siwkowski, who created the map in this Deliverable.

The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

This deliverable was internally reviewed by Dr. Jana Fritzsch of the Leibniz Institute of Agricultural Development in Central and Eastern Europe, Halle (Saale).

Abstract

The Deliverable 3.3 of the SCARLED project develops a typology of the 175 rural NUTS3 regions in the new Member States in order to reveal similarities and differences of demographic, socio-economic and agricultural patterns. Rural areas have been defined according to the OECD definition. The starting point for classification is the comprehensive database completed in Deliverable 3.1 using harmonised Eurostat data, and the conclusions drawn out of Deliverable 3.2. Five variables on NUTS3 level have been chosen out of the database for classification: 1) Change of population 2000-2005; 2) GDP per capita 2004; 3) Change of GDP per capita 2000-2004; 4) Share of employment in industry and services 2004; 5) Share of holdings <2 ESU 2005. The applied hierarchical agglomerative cluster analysis (Ward's method) revealed five different clusters (regional types) as the most plausible result: 1) Backward agrarian regions; 2) Dynamic agrarian regions; 3) Intermediate regions; 4) Advanced regions; 5) Best performing regions. The typology shows how the dimension and combination of problems differs. And it illustrates that the EU Rural Development policy is confronted with a big task to reach its goals in the new Member States.

Executive Summary

The aim of this Deliverable is to develop a typology, which classifies rural areas in the new Member States (NMS) according to their current stage of development to reveal similarities and differences of demographic, socio-economic and agricultural patterns. At the same time, an instrument shall be provided to differentiate rural areas in the NMS in respect of performance against the European rural development (RD) policy objectives 1 "Improving the competitiveness of agricultural sector", and 3 "Improving the quality of life and encouraging diversification of economic activities". To define rural areas, the OECD definition has been used due to its widespread acceptance and applicability. As a result, 175 (predominantly and significantly) rural NUTS3 regions had to be classified.

The starting point for the typology of rural areas in this Deliverable is the comprehensive database completed in the SCARLED Deliverable 3.1 using harmonised Eurostat data, and the conclusions drawn out of the SCARLED Deliverable 3.2 (Working Paper "Socio-economic, demographic, and agricultural patterns of rural areas in the new Member States"). Five variables on NUTS3 level have been chosen out of the database, well suitable to represent the most important trends and to depict the development stages of a region: 1) Annual average change rate of total population in percent 2000-2005, representing the demographic stability and attractiveness of a region; 2) GDP per capita in EUR at PPP 2004, representing income levels of a region; 3) Annual average change rate of GDP per capita in EUR at PPP 2000-2004, representing the economic dynamic of a region; 4) Share of employment in industry and services in percent 2004, representing the degree of diversification in a region; 5) Share of holdings <2 ESU in total holdings 2005, representing the degree of semi-subsistence agriculture (and need for structural change) in a region. While variable 5 applies to the RD objective 1 (competitiveness of agriculture), variables 1 to 4 refer to the RD objective 3 (quality of life and diversification). Unemployment rate has not been used as variable, since this indicator is connected with many statistical problems and therefore not reliable.

There exist two major kinds of classification methodologies: 1) The inductive, aggregative approaches (including factor and cluster analysis), and 2) the deductive, disaggregative

approaches (including multi-criteria methods). Both have several advantages and disadvantages. As most researchers, the typology developed in this Deliverable uses the aggregative method of cluster analysis, since more indicators can be included in the analysis, the approach is best suitable to reveal the relevant (spatial) patterns and thus to contribute to new insights, and they do not require (as the simple multi-criteria methods do) "arbitrary" thresholds for the indicators used. Cluster analysis is a statistical method to "partition a set of observations into a distinct number of unknown groups or clusters in such a manner that all observations within a group are similar, while observations in different groups are not similar" (Timm 2002, p. 515). In this Deliverable, a hierarchical agglomerative cluster analysis, which is most commonly used in science, is applied to classify rural regions. As an algorithm for clustering the Ward's method together with the distance measure "Squared Euclidean Distance" was chosen, since it is best-suited to result in internally homogenous and externally distinguishable groups. All variables have been standardised before performing the cluster analysis to avoid an unequal weighting due to varying scales. They showed no correlations. To optimise the result of the hierarchical cluster analysis, the non-hierarchical method k-means was applied subsequently.

The cluster analysis revealed five different clusters (regional types) as the most plausible result. Ordered according to the respective cluster average of GDP per capita from lowest to highest, these are:

- Cluster 1: Backward agrarian regions (Agrarian lowest-income regions with pronounced subsistence orientation and strong population decrease)
  Regions in Bulgaria, Romania, and Latvia
- Cluster 2: Dynamic agrarian regions (Agrarian low-income regions with pronounced subsistence orientation and highest economic dynamic)
  Regions in Romania, Bulgaria, Lithuania, and Latvia
- Cluster 3: Intermediate regions (Middle-income regions with subsistence agriculture below average and lowest economic dynamic)
  Regions in Poland, Lithuania, and Slovenia
- Cluster 4: Advanced regions (Rather diversified, middle-income regions with subsistence agriculture above average)
  Regions in Slovakia, Hungary, Poland, Baltic States, Bulgaria, Romania, and the Czech Republic
- Cluster 5: Best performing regions (Diversified, highest-income regions with lowest degree of subsistence agriculture and stable population)
  Regions in Slovenia, Cyprus, Czech Republic, Hungary, Estonia, and Poland

Rural areas of cluster 1 are in all respects the worst performing regions. They urgently need to counteract out-migration, to raise income levels, to support diversification, and to find solutions for semi-subsistence farms. In contrast, cluster 5 contains those rural areas, which feature the best figures in comparison to all rural areas in the NMS. However, even these regions have a still large need to catch up in comparison to the rural EU15. Cluster 2-4 are mixed types in between. Cluster 2 has also big needs on every score, but the best economic dynamic of all types. Cluster 3 is rather intermediate in all respects. Worries cause the comparatively low economic dynamic and the out-migration, which is hidden behind the traditionally high fertility, which decreased only recently. Rural areas of cluster 4 require - despite their progress in income levels and diversification – particularly to cope with small (semi-subsistence) farms and in many regions also to counteract out-migration.

The typology shows how the dimension and combination of problems differs. And it illustrates that the EU RD policy is confronted with a big task to reach its goals in the NMS.

SCARLED Consortium

This document is part of a research project funded by the 6th Framework Programme of the European Commission. The project coordinator is IAMO, represented by Prof. Dr. Gertrud Buchenrieder (buchenrieder@iamo.de).

Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO) – Coordinator
Theodor-Lieser Str. 2
06120 Halle (Saale)
Germany

Contact person: Judith Möllers
E-mail: scarled@iamo.de

Catholic University Leuven (KU Leuven)
LICOS Centre for Institutions and Economic Performance & Department of Economics
Deberiotstraat 34
3000 Leuven.
Belgium

Contact person: Johan Swinnen
E-mail: jo.swinnen@econ.kuleuven.be

University of National and World Economy (UNWE)
St. Town "Chr. Botev"
1700 Sofia
Bulgaria

Contact person : Plamen Mishev
E-mail: mishevp@intech.bg

Corvinus University Budapest (CUB)
Department of Agricultural Economics and Rural Development
Fövám tér 8
1093 Budapest
Hungary

Contact person: Csaba Csáki
E-mail: csaba.csaki@uni-corvinus.hu

Research Institute for Agricultural Economics (AKI)
Zsil u. 3/5
1093 Budapest
Hungary

Contact person: József Popp
E-mail: poppj@akii.hu

Warsaw University, Department of Economic Sciences (WUDES)
Dluga 44/50
00-241 Warsaw
Poland

Contact person: Anna Dominika Milczarek
E-mail: milczarek@wne.uw.edu.pl

Banat's University of Agricultural Sciences and Veterinary Medicine Timisoara (USAMVB)
Calea Aradului 119
300645 Timisoara
Romania

Contact person: Cosmin Salasan
E-mail: cosminsalasan@xnet.ro

University of Ljubljana (UL)
Groblje 3
1230 Domzale
Slovenia

Contact person: Luka Juvančič
E-mail: luka.juvancic@bfro.uni-lj.si

The University of Kent, Kent Business School (UNIKENT)
Canterbury
Kent CT2 7NZ
United Kingdom

Contact person: Sophia Davidova
E-mail: s.davidova@imperial.ac.uk

University of Newcastle upon Tyne, Centre for Rural Economy (UNEW)
Newcastle upon Tyne NE1 7RU
United Kingdom

Contact person: Neil Ward
E-mail: neil.ward@newcastle.ac.uk

CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF MAPS

LIST OF ABBREVIATIONS

ESU	European Size Units
EU15	All EU countries before the enlargements 2004 and 2007
GDP	Gross Domestic Product
NMS	New Member States
NUTS	Nomenclature des Unités Territoriales Statistiques
PPP	Purchasing Power Parities
PR	Predominantly rural
PU	Predominantly urban
RD	Rural Development
SCARLED	Structural change in agriculture and rural livelihoods
SR	Significantly rural
Std. dev.	Standard deviation
WP	Workpackage

1 INTRODUCTION

Rural issues are of high significance in the New Member states (NMS) since more than 90% of the NUTS3 regions are rural with more than 80% of total population. However, this statement does not imply that these areas have homogenous characteristics and problems but that they are divers and manifold. As a result, the substantial inter- and intra-country differences very often do not allow recognising general trends of population, income, employment and agriculture in all rural areas of the NMS (cf. SCARLED Deliverable 3.2). This Deliverable aims at identifying regional types with similar structures and/or problems and thus similar social and farm restructuring needs within the NMS. After a short discussion of the objectives of typologies and the typology construction methods, the methodology and data used for the classification of rural areas in this Deliverable are outlined in chapter 2. The five regional types resulting from the analysis are presented in chapter 3 followed by conclusions in chapter 4.

1.1 Objectives of typologies

The classification of (rural) areas is a method of regional research which aims at characterising regions according to – for the purpose of the respective research question – dominant attributes. The comparability of regions shall be achieved allowing a systematised appraisal. The method is ultimately based on the hypothesis that the diversity of reality can be reduced to few variables (Thiel and Crinius 1990). Typologies are quantitative approaches, demanding large volumes of empirical data, using statistical methodologies such as factor or cluster analysis, or multi-criteria approaches (see Section 1.2). They may distinguish between regions in terms of "rurality", or in terms of several socio-economic characteristics, which may be collectively viewed as indicative of "performance". Typologies can be broad, using many variables, or narrow, within one thematic domain (Copus et al. 2007). There are three main kinds of typology (cf. Thiel and Crinius 1990):

1. Classification of regions according to their previous or current stage of development (most common case). Principally, infinitely many variables are possible, which have to be chosen depending on the research question. "The clarification of the current situation and present development are already important instruments of regional policy. The very presentation of regional types, which are characterised by varying developments and standards, can initiate political effects and support or stimulate argumentations of regional policy" (own translation of Thiel and Crinius 1990, p. 79).
2. Classification of regions according to their deficits, surpluses or necessary measures of regional policy objectives defining needs for action. "Typologies can … (potentially) support territorial impact assessment, by simplifying complex patterns, identifying different regional contexts in which different goals or processes are desirable, or by differentiating regions in terms of performance against goals" (Copus et al. 2007, p.4).
3. Classification of regions according to causes of previous development. Such classification could be in principal suitable for the development of strategies and the selective application of instruments in different regional types. However, it is up to now rather seldom due to numerous methodological problems as a result of the still-insufficient existing theoretical basis of the reasons for socio-economic differences (Klemmer and Junkernheinrich 1990).

The typology developed in this Deliverable corresponds mainly to the first aforementioned kind. It aims at revealing similarities and differences of demographic, socio-economic and

agricultural patterns to identify social and farm restructuring needs in different rural regions in the NMS. However, the typology shall be also linked to the EU rural development (RD) policy. Therefore, the selection of variables will be done within its three axes so that the regions can be differentiated in respect of performance against the European RD policy objectives (see Section 2.1).

1.2 Typology construction methods

Among the range of classification methodologies two major kinds can be distinguished. The first group are aggregative approaches (including factor and cluster analysis). They may be viewed as inductive, since they use mathematical procedures to reveal so far unknown (spatial) structures, and the operator has no direct control over the character of the emerging types. The second group are disaggregative approaches (including multi-criteria methods). They are basically deductive, i.e., the classification is carried out due to certain thresholds, which are defined by models or theories. They may be favoured where the researcher or policy maker has a clear idea what the relevant categories are, and wishes to show how regions fit into them (Bahrenberg et al. 2003, Copus et al. 2007, Thiel and Crinius 1990). The two methodological groups can be compared using six criteria (see Table 1.1). Most researchers have favoured the aggregative methodology, since more indicators can be included in the analysis, the approach is best suitable to reveal the relevant (spatial) patterns and thus to contribute to new insights, and they do not require (as the simple multi-criteria methods do) "arbitrary" thresholds for the indicators used. In contrast, the advantages of the multi-criteria methodology are its transparency, its ease of interpretation, and its robustness of the results over time. However, the small number of indicators, which can be used for classification, might well cause difficulties for this simple approach as well as the definition of thresholds (Copus et al. 2007).

Table 1.1 Assessment of the two common approaches of typology construction

	Cluster and factor analysis (aggregative method)	Simple multi-criteria method (disaggregative method)
Simplicity/transparency	-	+
Degree of objectivity, replicability of results	-	+
Ease of interpretation	-	+
Ability to reveal the "relevant" spatial patterns	+	-
Ability to consider a large number of indicators	+	-
Robustness of the results over time	-	+

Source: Copus et al. 2007, p. 82-83

The typology developed in this Deliverable will use the aggregative method of cluster analysis (see Section 2.2) due to its aforementioned advantages. Furthermore, the EU RD policy does not define such clear policy objectives to allow the definition of justified thresholds as required by multi-criteria approaches.

2 METHODOLOGY AND DATA

The first step in the typology of rural areas is the definition of "rural". For this purpose, the OECD definition (OECD 1994 and 2005, cf. Deliverable 3.2) has been used due to its widespread acceptance and applicability. As a result, 175 rural NUTS3 regions[2] had to be classified. In Section 2.1 below, the selection of variables will be motivated, followed by a description of the applied method of cluster analysis (Section 2.2).

2.1 Selection of variables

In principle, the variables used for classification have to meet two main criteria: They have to be relevant for the research question, and they have to be available, in a harmonised form, across all 12 NMS, at an appropriate level of regional distribution. The starting point for the typology of rural areas in this Deliverable is the comprehensive database completed in the SCARLED Deliverable 3.1 using harmonised Eurostat data, and the conclusions drawn out of the SCARLED Deliverable 3.2 (Working Paper "Socio-economic, demographic, and agricultural patterns of rural areas in the new Member States"). The least aggregated level, for which data is available and which is therefore used, is NUTS3[3]. On this level, some regions are still not as internally homogeneous as desired. In addition, not all indicators which might be relevant for assessing the demographic, socio-economic and agricultural patterns are available. Nevertheless, for each thematic domain one or more variables could be chosen, well suitable to represent the most important trends and to depict the development stages of a region. These variables are:

1. Annual average change rate of total population in percent 2000-2005[4]: Demographic indicator representing the demographic stability and attractiveness of a region.
2. GDP per capita (PPP, EUR) 2004: Socio-economic indicator representing the income level (and therewith also partly the quality of life and infrastructure level) of a region.
3. Annual average change rate of GDP per capita (PPP, EUR) 2000-2004: Socio-economic indicator representing the economic dynamic of a region.
4. Share of employment in industry and services in percent 2004[5]: Socio-economic indicator representing the degree of diversification (non-agricultural employment) in a region.
5. Share of holdings <2 ESU[6] in total holdings 2005: Agricultural indicator representing the degree of subsistence agriculture (and need for structural change) in a region.

---

[2] Rural regions refer to the categories "predominantly rural" and "significantly rural" of the OECD definition. The category "predominantly urban" has not been considered in the analysis.

[3] NUTS = Nomenclature des Unités Territoriales Statistiques/Nomenclature of territorial units for statistics: NUTS0 and NUTS1 refer in the NMS to the whole country. The NUTS2 level represents in the NMS 55 administrative units each with approx. 800,000 to 3,000,000 inhabitants. The NUTS3 level encompasses in the NMS 193 regions each with approx. 150,000 to 800,000 inhabitants.

[4] Hungary 2001-2005

[5] For Romania national data have been used, since these data are not available in Eurostat on NUTS3 level.

These variables also refer to two major objectives of the EU RD policy. In the period 2007-2013, the European RD policy focuses on three core policy objectives each corresponding to a thematic axis (and complemented by a "methodological" LEADER axis): 1) Improving the competitiveness of agricultural and forestry sector, 2) Improving the environment and the countryside through land management, 3) Improving the quality of life and encouraging diversification of economic activities (European Commission 2006). While variable 5 applies to objective 1 (competitiveness of agriculture), variables 1 to 4 allow differentiating regions in terms of performance against the objective 3 (quality of life and diversification). Objective 2 (environment) is not considered, since it requires more natural scientific indicators.

Unemployment rate has not been used as variable, although it is available on NUTS3 level and generally considered as important socio-economic parameter. However, this indicator is connected with many statistical problems and therefore not reliable. Unemployment rates are highly influenced by definitions or approaches of collecting statistical data as well as the incentives to register as unemployed in the respective countries. It is not able to record hidden unemployment which has a high significance in many parts of Central and Eastern Europe. In Romania for example, people does not regard themselves as unemployed even if they have hardly anything to do. Incentives to register as unemployed are low, small family farms have a high importance for employment, and there exist measures like shortened work schedules to keep unemployment low, so that the low unemployment rate in Romania conceals the real social problems.

Finally, five variables have been used to classify the 175 rural NUTS3 regions of the NMS by the means of cluster analysis.

2.2 Cluster analysis

Cluster analysis is a statistical method to "partition a set of observations into a distinct number of unknown groups or clusters in such a manner that all observations within a group are similar, while observations in different groups are not similar" (Timm 2002, p. 515). There are various different procedures for the grouping of observations (here: regions). In general, there are hierarchical and non-hierarchical procedures. The hierarchical procedures are best suited if no information exists about the optimal number of clusters. While *agglomerative* hierarchical methods start with a set of *N* regions and join step by step the most similar regions or clusters to new clusters ending with one cluster that involves all regions, the *divisive* hierarchical methods operate the opposite way. In practice, only the agglomerative methods are of significance. Non-hierarchical procedures are based on a given partition of regions in *G* clusters which has to be optimised. They build clusters around a set of starting points (so-called cluster seeds). For example, when four cluster seeds are specified four clusters will be formed. In practice, non-hierarchical procedures are often applied to improve the results of hierarchical cluster analyses (see Backhaus et al. 2003, Eckey et al. 2002, and Hair et al. 2006 for details about cluster analysis).

In this Deliverable, a hierarchical agglomerative cluster analysis, which is most commonly used in science, is applied to classify rural regions. To create groups (or clusters) with high

[6] The economic size of farms is expressed in terms of European Size Units (ESU). The value of one ESU is defined as a fixed number of EUR of Farm Gross Margin. Currently, these are 1200 EUR per ESU (European Commission 2008, p. 6/78).

internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity this procedure assumes a multi-dimensional co-ordinate system where each axis represents one feature (such as GDP per capita) and each region is definitely positioned according to its characteristics. The closer to each other regions are, the more likely they are to be grouped into the same cluster (Bahrenberg et al. 2003). The distance between regions within the multi-dimensional space can be measured differently, and there exist various different clustering algorithms. Each combination of distance measure and agglomerative method may cause different results for the same data set and a lot of research has been done to propose the most effective combination. In this Deliverable, the Ward's method was chosen as an algorithm for clustering together with the distance measure "Squared Euclidean Distance", since it is best-suited to result in internally homogenous and externally distinguishable groups (Bergs 1981 cited in Backhaus et al. 2003) and has been often used for grouping regions (see e.g., Barjak 2000, Baum, Trapp and Weingarten 2004, Rovan and Sambt 2003). Ward's method uses a variance criterion for building up clusters. Two clusters are joined when the increase of the error sum of squares (sum of squared deviations from the cluster centroid) is minimal taking all other solutions into account. The centroid of a cluster is defined as the arithmetic mean for all variables and regions within the cluster (Backhaus et al. 2003, Bahrenberg et al. 2003, Eckey et al. 2002).

One precondition for cluster analysis is that the used variables have to be uncorrelated. Correlated variables act as weights in the clustering process, e.g. one attribute could be considered three times due to three correlated variables whereas others are considered only once. This causes solutions that are biased with respect to the correlated variables (Backhaus et al. 2003, Hair et al. 2006). Pearson's correlation coefficients for the five variables, which are used for classification in this Deliverable (see Section 2.1), are all below 0.8 (most even below 0.5) and, thus, the variables can be considered as uncorrelated according to Hübler (1989).[7] Furthermore, all variables have been standardised before performing the cluster analysis to avoid an unequal weighting due to varying scales.[8]

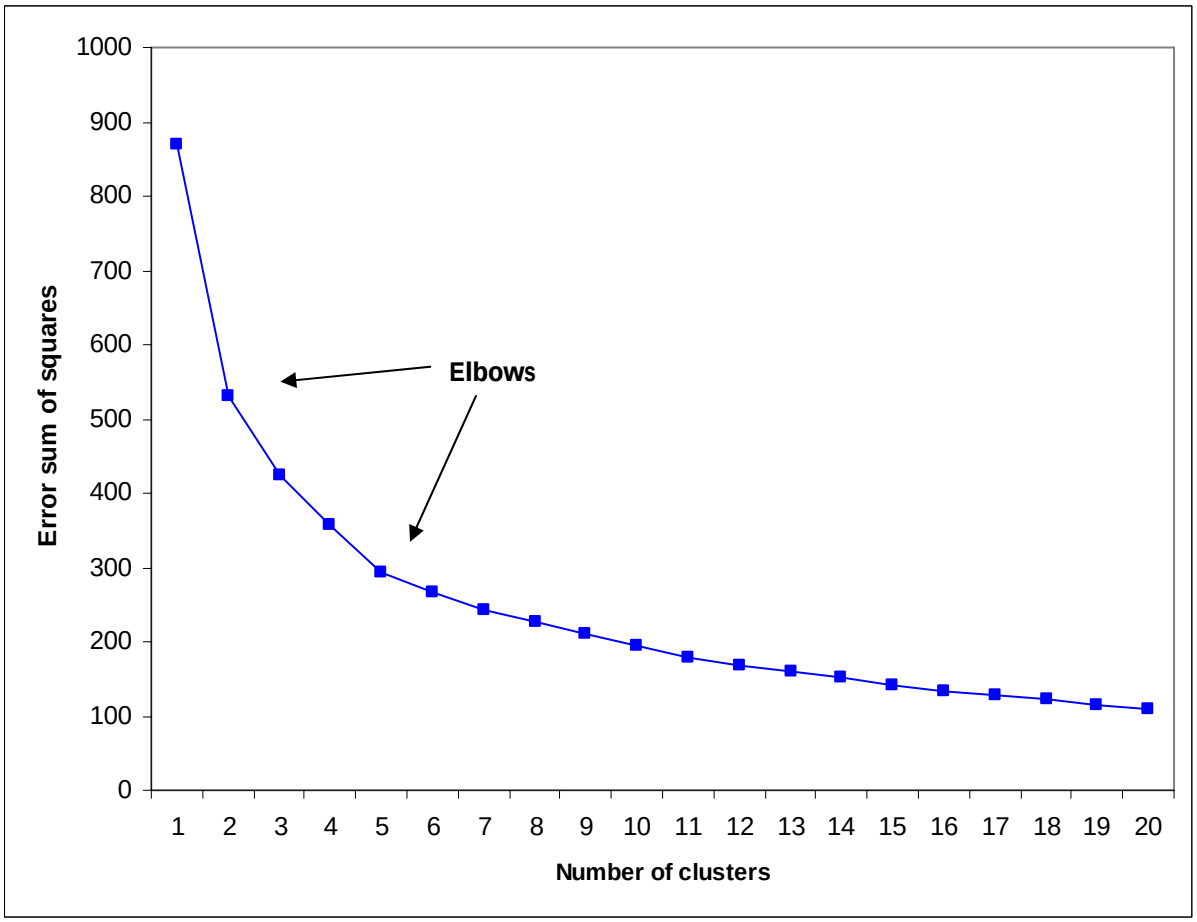
An agglomerative hierarchical cluster analysis does not automatically result in one optimal number of clusters. During the clustering process, the number of clusters is reduced one by one by merging two existing clusters. In the first step, each region represents a single cluster. After the last step, all regions are included in one cluster. A dendrogram visualises the steps in a hierarchical clustering procedure. There is no singular measure to decide on the most appropriate number of clusters for the research problem investigated. The elbow criterion, i.e., a sudden jump upwards in the agglomeration coefficients (at Ward's method these are the error sum of squares measured at several clustering steps), provides an indication of the step at which to stop the clustering procedure. The dendrogram is an additional mean of deciding on the number of clusters. Since the expert is given the responsibility of choosing the distance measure and the clustering algorithm, as well as the most appropriate number of groups, the results of a cluster analysis are always to some degree subjective (Backhaus et al. 2003, Hair et al. 2006).

[7] Therefore, it was not necessary to exclude variables or to perform a factor analysis prior to the cluster analysis. These are both ways to overcome correlations in the dataset (Backhaus et al. 2003).

[8] By a Z-transformation, a variable is standardised such that its mean equals 0 and its standard derivation equals 1.

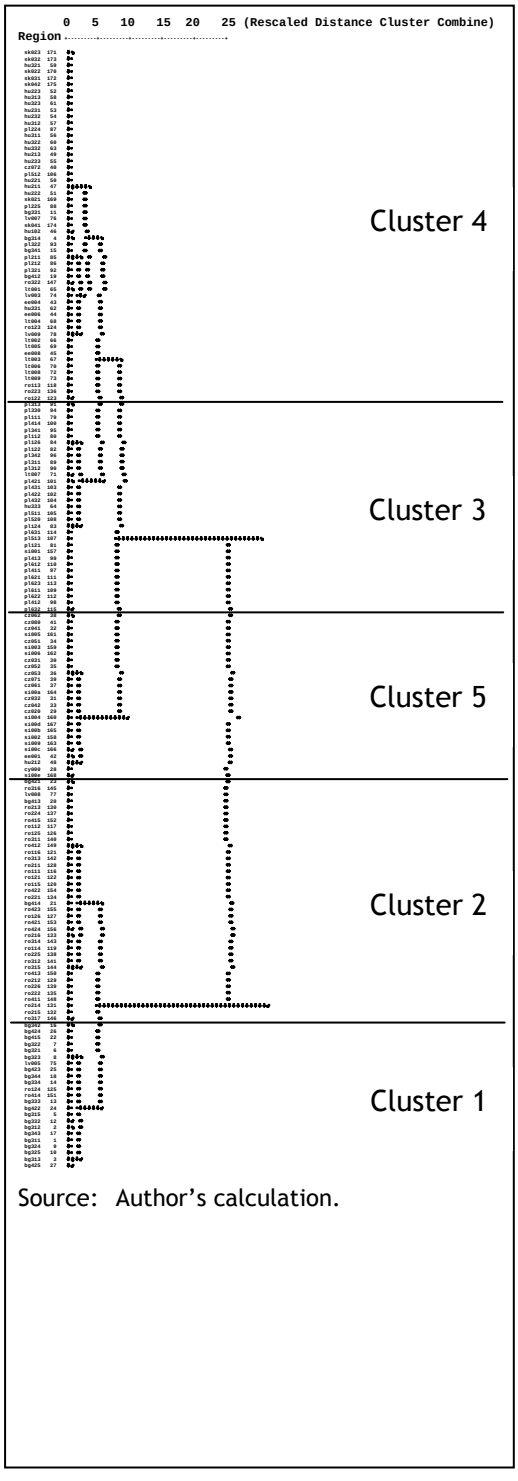
Figure 2.1 shows the elbow criterion plot of the applied cluster analysis. Two "elbows" are visible, suggesting a 2- or 5-cluster solution. Since the transition of the 2-cluster solution to the 1-cluster solution shows always the strongest increase of heterogeneity (Backhaus et al. 2003) and a 2-cluster solution offers not the desired differentiation of regions, the 5-cluster solution was chosen. It is also visualised in the dendrogram (see Figure 2.2).

Figure 2.1 Elbow criterion plot, cluster analysis of 175 rural NUTS3 regions in the NMS



Source: Author's calculation.

Figure 2.2 Dendrogram of the cluster analysis of 175 rural NUTS3 regions in the NMS



Source: Author's calculation.

One drawback of hierarchical cluster methods is that they do not reallocate the regions during the cluster process. At a certain point within the cluster process it could happen that a region that was allocated to a certain cluster some steps before would be better allocated into another subsequent cluster. As this is not done automatically during clustering it should be done afterwards otherwise the clusters will not be as clearly separated as it is needed for further analysis. For this purpose, the non-hierarchical cluster method k-means will be used. It forms a given number of clusters around the so-called cluster seeds, which represent in this Deliverable the arithmetic means for all variables within each cluster taken from the hierarchical cluster procedure. The k-means method computes the Euclidean distances of all regions to the various cluster seeds. Then, each region is assigned to the nearest cluster seed and the cluster seeds are calculated again. This process is repeated until changes in cluster seeds become small (no reassignments in cluster membership) or a specified number of iterations are done (Backhaus et al. 2003, Bahrenberg et al. 2003, Eckey et al. 2002, Hair et al. 2006). The application of the k-means method subsequent to the hierarchical cluster analysis resulted in the regrouping of 17 regions located in Hungary, Lithuania, Poland, and Romania.

3 RESULTS OF ANALYSIS

The applied cluster analysis of the 175 rural NUTS3 regions in the NMS revealed five different clusters (regional types) as the most plausible result. Ordered according to the respective cluster average of GDP per capita from lowest to highest, these are:

- Cluster 1: Backward agrarian regions
- Cluster 2: Dynamic agrarian regions
- Cluster 3: Intermediate regions
- Cluster 4: Advanced regions
- Cluster 5: Best performing regions

The five clusters are visualised in Map 3.1 and characterised in Table 3.1. In the following, they will be shortly described. The characteristics of the clusters are based on the cluster averages, which can contain a more or less pronounced variation (see Table 3.1).

Cluster 1: Backward agrarian regions

Agrarian lowest-income regions with pronounced subsistence orientation and strong population decrease

This cluster contains those regions of Bulgaria – besides Latgale in Eastern Latvia and three regions in Romania – where very unfavourable factors coincide: The high share of agriculture in employment (36% on average), which is characterised by a high subsistence orientation (92% of all holdings <2ESU on average), is connected with the lowest GDP per capita among all groups (5458 EUR at PPP on average). The population has decreased dramatically between 2000 and 2005 (annually -1.9% on average); a high out-migration – beside the birth deficit – can be assumed. This reduces the future development potential of these regions, since it will lead to an overaged population structure. The annual average change rate of GDP per capita (annually 6.7% between 2000 and 2004 on average, varying between 0.7% and 11.9% within the group) is slightly below the average of all rural regions in the NMS (but of course strongly above the rural EU15 average of 2.9%). Measures to create job opportunities for young people are badly needed to prevent that these regions found themselves back in a dead end.

Cluster 2: Dynamic agrarian regions

Agrarian low-income regions with pronounced subsistence orientation and highest economic dynamic

In addition to one Latvian, three Lithuanian and three Bulgarian regions, the regions of cluster 2 are mainly located in Romania. They are similar to those of cluster 1 in respect of high agricultural employment (35% on average) and subsistence orientation (90% of all holdings <2ESU on average). However, in contrast to cluster 1, this group has, on average, both, a higher GDP per capita (6586 EUR at PPP) and the highest annual change rate of GDP per capita among all groups (annually +11.3% between 2000 and 2004). Furthermore, the population decrease is less pronounced (annually -0.7% between 2000 and 2005 on average). While the high economic dynamic raise hopes for the future regional potential, the issues of low diversification, subsistence agriculture and population decrease call for support in structural change. Furthermore, it can be questioned, whether small villages benefit from the growth in income levels or whether it is rather limited to regional centres.

Cluster 3: Intermediate regions

Middle-income regions with subsistence agriculture below average and lowest economic dynamic

Cluster 3 comprises primarily rural areas in Poland, supplemented by two Lithuanian regions and Pomurska in Northeast Slovenia. It is characterised by a medium level of GDP per capita (8609 EUR at PPP on average) with the lowest growth rate among all groups (annually +4.5% between 2000 and 2004 on average). Agriculture accounts for 28% of employment on average and has a subsistence orientation below the average of all rural areas in the NMS (57% of all holdings <2ESU on average). The population is rather stable (change rate annually -0.02% between 2000 and 2005 on average), caused by a still high natural population growth, which countervails the net out-migration. This phenomenon will not continue in the long term, since the total fertility rates have fallen meantime below reproductive level also in these regions (cf. Deliverable 3.2). Thus, these regions run an obvious risk that migration induced changes in age structure and fertility will result in natural change (and with it the total population change) turning negative.

Cluster 4: Advanced regions

Rather diversified, middle-income regions with subsistence agriculture above average

Rural areas of cluster 4 are already more diversified compared with the mean of all NMS (87% of employed in industry and services on average). They encompass all rural areas in Slovakia, nearly all rural areas in Hungary, large parts of the Baltic States, Western and Southern Poland, and some regions in Bulgaria, Romania, and the Czech Republic. The income level is above average (9414 EUR at PPP) with a growth rate below average (annually +5.4% between 2000 and 2004 on average). Agriculture shows a still high subsistence orientation (85% of all holdings <2ESU on average), and the population is slightly decreasing (change rate annually -0.2% between 2000 and 2005 on average). Higher levels in diversification and income offer generally better preconditions for overcoming structural change and stimulating new economic activity than in poorer, agrarian regions. Nevertheless, rural areas of cluster 4 need support in using their potential and in the adjustment of the large number of small (semi-subsistence) farms.

Cluster 5: Best performing regions

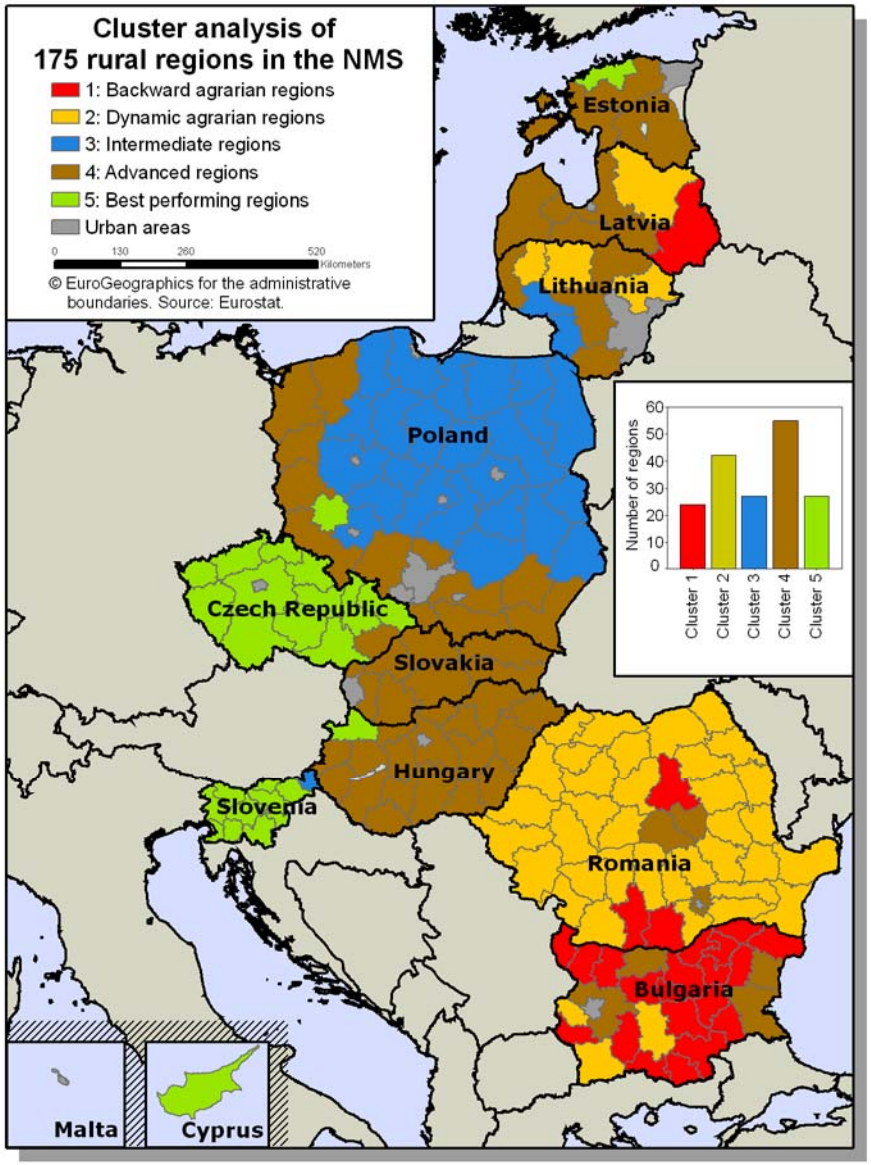
Diversified, highest-income regions with lowest degree of subsistence agriculture and stable population

The best ranking have rural areas of cluster 5, which includes Cyprus, large parts of Slovenia and the Czech Republic, one region in Northwest Hungary, Legnicki in Southwest Poland, and the region of Tallinn in Estonia. They feature the highest GDP per capita (15301 EUR at PPP on average), the highest degree of diversification (93% of employed in industry and services on average) and the lowest subsistence orientation in agriculture (54% of all holdings <2ESU on average) among all groups. The economic dynamic is below the NMS average (change rate of GDP per capita annually 5.2% between 2000 and 2004 on average), but in virtually all regions greater than the rural EU15 average of +2.9%. The population is stable (change rate annually +0.07% between 2000 and 2005 on average; in most regions value near zero), and strongly growing in Cyprus (+1.8%). That these regions come off well in comparison to all rural regions in the NMS should not hide the fact that structural change has to be supported here, too. With the exception of the Slovenian

region Osrednjeslovenska (Ljubljana), the GDP per capita is in all regions below the rural EU15 average (19819 EUR at PPP in 2004), and the farm structure much more fragmented than in the EU15.

To summarise the results, the cluster analysis of 175 rural NUTS3 regions in the NMS resulted in five different types of regions as the most adequate result. The typology shows how the dimension and combination of problems differs. However, it should be noted that other indicators or procedures could lead to other classifications.

Map 3.1 Regional distribution of the 5 clusters as a result of a cluster analysis of 175 rural NUTS3 regions in the NMS



Source: Author's calculation based on Eurostat Regio data

Table 3.1 Characteristics of the 5 clusters and of all rural NUTS3 regions in the NMS

Cluster (number of regions)		Population: Annual average change rate of population in % 2000-2005 [1]	Income: GDP p.c. (PPP, EUR) 2004	Dynamic: Annual average change rate of GDP p.c. (PPP, EUR) 2000-2004	Diversification: Percentage share of employment in industry and services 2004 [2]	Subsistence: Share of holdings <2 ESU in total holdings 2005
1: Backward agrarian regions (24)	**Average [3]**	**-1.93**	**5457.5**	**6.69**	**63.8**	**91.7**
	Minimum	-4.58	4541.1	0.71	41.2	82.5
	Maximum	-0.88	7185.2	11.94	85.0	96.4
	Std. dev.	0.92	812.9	2.30	10.9	3.9
2: Dynamic agrarian regions (42)	**Average [3]**	**-0.73**	**6585.5**	**11.26**	**64.6**	**89.8**
	Minimum	-1.65	3627.5	6.73	42.3	76.3
	Maximum	0.15	10130.0	15.81	81.8	96.7
	Std. dev.	0.32	1552.8	2.15	10.4	4.4
3: Inter-mediate regions (27)	**Average [3]**	**-0.02**	**8608.5**	**4.52**	**71.9**	**57.2**
	Minimum	-0.59	5780.3	3.05	52.0	33.7
	Maximum	0.76	12364.2	6.85	86.3	71.3
	Std. dev.	0.33	1575.8	0.97	10.1	9.8
4: Advanced regions (55)	**Average [3]**	**-0.22**	**9414.2**	**5.42**	**87.0**	**84.9**
	Minimum	-1.08	4905.4	0.77	65.7	68.8
	Maximum	1.58	15436.2	12.30	96.8	95.4
	Std. dev.	0.46	2158.1	2.19	9.0	7.5
5: Best performing regions (27)	**Average [3]**	**0.07**	**15301.4**	**5.19**	**92.7**	**53.7**
	Minimum	-0.44	12634.9	2.80	83.0	39.5
	Maximum	1.77	25595.8	10.47	99.0	81.7
	Std. dev.	0.42	2788.3	1.47	4.4	10.6
All (175)	**Average [3]**	**-0.50**	**8976.7**	**6.82**	**77.0**	**77.9**
	Minimum	-4.58	3627.5	0.71	41.2	33.7
	Maximum	1.77	25595.8	15.81	99.0	96.7
	Std. dev.	0.81	3607.1	3.22	14.8	17.0
Rural EU15 average [4]		**0.55**	**19819.2**	**2.88**	**92.4**	**29.5 [5]**

Source: Author's calculation based on Eurostat Regio data

Notes: [1] Hungary 2001-2005. [2] Romania: national data (National Institute of Statistics, 2006: Romanian Statistical Yearbook 2005. P.180-183). [3] Unweighted arithmetic mean value. [4] Average of PR and SR regions in the EU15 countries. Due to missing values only 644 (for the diversification indicator: 532) regions (out of 690) could be used for calculation. Countries with missing values are BE, DE, DK, ES, FI, IT, UK. [5] Value for total EU15 (including PU regions).

4 CONCLUDING REMARKS

Rural areas in the NMS cannot be considered homogeneous and need adapted policy strategies. In this Deliverable, five types of rural areas with similar structures and/or problems have been identified by the means of cluster analysis based on five variables. The variables reflect the main needs in the rural areas of the NMS. These are:

1. To counteract out-migration,
2. To raise income levels,
3. To support diversification,
4. To facilitate farm restructuring – find solutions for semi-subsistence farms.

Rural areas of cluster 1 are in all respects the worst performing regions. In contrast, cluster 5 contains those rural areas, which feature the best figures in comparison to all rural areas in the NMS. However, even these regions have a still large need to catch up in comparison to the rural EU15. Cluster 2-4 are mixed types in between. Cluster 2 has also big needs on every score, but the best economic dynamic of all types. Cluster 3 is rather intermediate in all respects. Worries cause the comparatively low economic dynamic and the out-migration, which is hidden behind the traditionally high fertility, which decreased only recently. Rural areas of cluster 4 require - despite their progress in income levels and diversification – particularly to cope with small (semi-subsistence) farms and in many regions also to counteract out-migration.

The typology illustrates that the EU RD policy is confronted with a big task to reach its goals in the NMS. The conducted classification differentiates rural areas in terms of performance against the RD goals "Improving the competitiveness of agricultural sector", and "Improving the quality of life and encouraging diversification of economic activities". It shows that the dimension and combination of problems differs. To analyse the levels of single needs in the rural areas of the NMS, the maps of Deliverable 3.2 could be used as well.[9] They can be regarded also as small typologies of one specific topic. The results of the survey (WP4 "Design and implementation of a survey instrument") and the subsequent analysis in the Workpackages[10] of the SCARLED project can provide important insights in the structural adjustment processes of the surveyed regions, representing different regional types. In addition, further research could try to refine the typology with an extended set of variables (e.g., to specify farm restructuring needs).

[9] See Map 2.1, 2.3, 3.1, 3.3, and 4.3 in Deliverable 3.2.

[10] WP5 "Farm structure evolution"; WP6 "Socio-economic functions of subsistence farming and co-operation among farmers", and WP7 "Rural labour markets and diversification of rural economies".

List of references

Backhaus, K., Erichson, B., Plinke, W. and R. Weiber (2003): Multivariate Analysemethoden: eine anwendungsorientierte Einführung. 10th edition, Berlin et al., Germany: Springer.

Bahrenberg, G., Giese, E., and J. Nipper (2003): Statistische Methoden in der Geographie. Band 2, Multivariate Statistik. Berlin, Stuttgart, Germany: Gebrüder Borntraeger Verlagsbuchhandlung.

Barjak, F. (2000): Differences in the economic capability of regions - a typology for East Germany and Poland. Discussion Papers No. 121. Halle (Saale), Germany: Institute for Economic Research Halle.

Baum, S., Trapp, C. and P. Weingarten (2004): Typology of rural areas in the Central and Eastern European EU new Member States. IAMO Discussion Paper No. 72. Halle (Saale), Germany: IAMO.

Copus, A., Psaltopoulus, D., Skuras, D., Terluin, I. and P. Weingarten (2007): Common Features of Diverse Rural Areas: Review of Approaches to Rural Typology. Final Report Version 1.3 (Contract: 150669-2007 F1SCUK).

Eckey, H.-F., Kosfeld, R. and M. Rengers (2002): Multivariate Statistik. Grundlagen – Methoden – Beispiele. Wiesbaden, Germany: Gabler.

European Commission (2006): The EU Rural Development Policy 2007–2013. Fact Sheet. Luxembourg: Office for Official Publications of the European Communities.

European Commission (2008): Farm Accounting Data Network - An A to Z of methodology. Version 28/03/2008 08:58:40. Brussels, Belgium: European Commission, DG Agri http://ec.europa.eu/agriculture/rica/pdf/site_en.pdf (accessed April 2008)

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. and R. L. Tatham (2006): Multivariate Data Analysis. 6th edition. Prentice Hall, Upper Saddle River, New Jersey, USA: Pearson.

Hübler, O. (1989): Ökonometrie. Stuttgart, Germany: Gustav Fischer Verlag.

Klemmer, P. and M. Junkernheinrich (1990): Regionstypenbezogene Fortentwicklung der Raumentwicklungspolitik. Typisierung von Arbeitsmarktregionen anhand ausgewählter Bestimmungsfaktoren des regionalen Entwicklungspotentials. In Räumliche Typisierung für die Raumentwicklungspolitik. Akademie für Raumforschung und Landesplanung (ed.): p. 1-61, Hannover, Germany.

OECD (1994): Creating rural indicators for shaping territorial policy. Paris, France: Organization for Economic Cooperation and Development (OECD).

OECD (2005): Regions at the glance. Paris, France: Organization for Economic Cooperation and Development (OECD).

Rovan, J., Sambt, J. (2003): Socio-economic Differences Among Slovenian Municipalities: A Cluster Analysis Approach. In Developments in Applied Statistics. Ferligoj, A., and A. Mrvar (eds.): p. 265-278, Metodološki zvezki 19, Ljubljana, Slovenia.

Thiel, E. and W. Crinius (1990): Raumkategorien und raumordnungspolitische Maßnahmen. Probleme einer regionstypenbezogenen Raumentwicklungspolitik. In Räumliche Typisierung für die Raumentwicklungspolitik. Akademie für Raumforschung und Landesplanung (ed.): p. 63-111, Hannover, Germany.

Timm, N.H. (2002): Applied Multivariate Analysis. New York: Springer.