RURAL DEVELOPMENT INDICATORS FOR REGIONS WITH DIFFERENT DEGREES OF « RURALITY»: A STATISTICAL STUDY[1]

Edoardo PIZZOLI

Istat, Central Directorate of National Accounts

Abstract

Nowadays data from statistical offices allows the calculation of indicators related to several accounts of national development and, typically, economic, social and environmental characteristics can be highlighted. The key limitation of the same data is its territorial detail that, normally, do not match with rural and urban areas as defined by the theory.
The paper, based on available data for Italy, would like to investigate the statistical interaction among: (1) multidimentionality of development and related indicators (that can be approached, for example, alternatively with principal components analysis or composite indicators); (2) multidimentionality in the definition of "rurality"; (3) the degree of rurality of administrative areas (that is the level at which most of the surveys are designed, official statistics are calculated and statistical data is provided also to professional users). In Italy, the standard administrative level of disaggregation for national accounts and other main statistics is "Regions" (NUTS 2 in EU classification).
In detail, the relationships between previously indicated dimensions in the definition of "rurality" and in the description of development (and vice versa) will be studied within the principal components framework. After that, the correspondence analysis approach will be considered to explore overlapping and links between the different degree of rurality of regions and indicators previously selected to describe rural development.
The degree of rurality has been studied in OECD works (1994 - 2009) and in several papers by the author (2007 – 2010). This is also part of a FP7 research project, Blue-Ets (2010-2013), coordinated by Istat. A manual by UN (2005; 2011) dedicated to Rural Development Statistics and related issues will be implemented in the near future in EU and other regions. The Wye City Group of the UN Statistical Division is also working on the same topic. Finally, the specific issue of multidimentionality of indicators and composite indicators has been studied in statistical terms by OECD (2008).
To conclude, the paper would like to test the multivariate techniques and solutions on the previously indicated issues, with the support of different statistical packages and on Italian data available at NUTS 2 and 3 levels.

Résumé

Aujourd'hui les données des offices de statistique permettent le calcul des indicateurs liés à plusieurs dimensions de développement national et typiquement les caractéristiques économiques, sociales et environnementales peuvent être soulignées. La limitation des clés des mêmes données est leurs détail territoriale qui, normalement, ne correspondent pas aux zones rurales et urbaines telles que définies par la théorie.
Ce papier, basé sur les données disponibles pour l'Italie, voudrait étudier l'interaction statistique entre: (1) la multi-dimensionnalité du développement et des indicateurs connexes (qui peut être approché, par exemple, alternativement avec L'analyse en composantes principales ou avec d'indicateurs composites); (2) multi-dimensionnalité dans la définition de «ruralité»; (3) le degré de ruralité des régions administratives (qui est le niveau auquel la plupart des enquêtes sont conçues, les statistiques officielles sont calculés et les données statistiques sont fournies également à les utilisateurs professionnels). En Italie, le niveau administratif standard de désagrégation des comptes nationaux et d'autres statistiques principales sont «Régions» (NUTS2 dans la classification de l'UE).

[1] This paper draws on work carried out under the SSH FP7 Project BLUE-ETS Grant agreement no. 244767, funded by the European Commission, DG Research and Innovation, Unit L Science, Economy and Society, Seventh Framework Programme, Theme 8, Socio-economic Sciences and Humanities.

Dans le détail, les relations entre les dimensions précédentes indiquaient dans la définition de la «ruralité» et dans la description du développement (et vice versa) seront étudiés dans le cadre de l'analyse en composantes principales. Après cela, l'approche de l'analyse factorielle des correspondances seront considérées pour explorer les chevauchements et les liens entre les différents degrés de ruralité des régions et les indicateurs précédemment choisis pour décrire le développement rural.
Le degré de ruralité a été étudiée dans les documents de travail de l'OCDE et dans plusieurs articles de l'auteur (2007 - 2010). Cela fait également partie d'un projet de recherche FP7 coordonné par l'Istat: BLUE-ETS (2010-2013). Un manuel de l'ONU (2005; 2011) est consacrée aux statistiques du développement rural et les questions connexes qui seront mis en œuvre dans un proche avenir dans l'UE et d'autres domaines. A groupe (Wye city group) de la Division de statistique des Nations Unies travaille également sur le même sujet. Enfin, la question spécifique de la multi-dimensionnalité des indicateurs et des indicateurs composites ont été étudiés en termes statistiques de l'OCDE (2008).
Pour conclure, le papier voulez tester les différentes techniques multivariées et des solutions sur les questions précédemment indiqué, basé sur différents logiciels de statistiques et sur les données italienne disponibles au niveau NUTS 2 et 3.

Introduction

To produce Rural Development Indicators (RDI) for policy needs at different decision making's levels is becoming a necessity for Statistical Offices (SO) to meet information requirement, for example, for EU structural policies or UN Millennium Development Goals (MDGs).
Several proposals in terms of indicators and characteristics to be covered have been put forward and data on these subjects is regularly produced. A critical mismatch is that the territorial level at which variables should be monitored, in this study the rural and urban areas level as defined by the theory, is not the administrative level at which data is normally available. In Italy, the standard administrative level of disaggregation for national accounts and other main statistics is “Regioni” (Regions, NUTS 2 in EU classification) and only in limited cases is extended to “Province” (Provinces, NUTS 3) and “Comuni” (Italian Municipalities).
Administrative areas in Italy normally incorporate, partially or totally, rural and urban areas in different proportions and rural areas can be found across two or more administrative areas. Consequently, different degrees of “rurality” should be measured also at the statistical level.
In this paper, considering the data available for Italy on some key variables necessary to identify rural areas and calculate RDI at the lowest territorial level, the multivariate approach and related techniques are tested to explore a statistical solution to the previously indicated issues. In the next paragraph the “rurality” definition is briefly summarized. In paragraphs two and three methodology and data are presented. In paragraphs four and five results are reported and discussed.

1. « Rurality » of Administrative Areas

Italy is subdivided in 20 Regions (NUTS2), 106 Provinces (NUTS3) and 8101 Municipalities, each of them with specific administrative borders that overlap with geographical regions. These areas do not correspond and sometimes cut rural or urban areas, creating problems for analysis as official statistics and data are defined only at administrative level.
In any case, administrative areas should be adopted as starting territorial partition of the Country. Their characteristics can be considered to define rural areas following indications from the literature (UN, 2005).
A starting definition consider population density as key variable but further studies demonstrate that also agriculture related and social-environmental characteristics should be consider as classification characteristics.
Finally, even if administrative areas differ one to the other on several economic, social and environmental characteristics, there are spatial correlations that have to be considered in the statistical analysis to produce territorial aggregations useful for policy analysis.

2. Methodology

The methodology of this study is divided into two parts: factorial multivariate analysis was initially used to calculate the rurality factor; next, the local Moran index was applied to identify spatial dependence between counties with respect to rurality.
The factorial Analysis produce the rurality index. Obtaining principal components depends on decomposition of the covariance matrix. Once the covariance matrix is determined, eigenvalues are calculated, indicating the degree of total variance explanation, represented by each component. From this result, the principal component with the highest eigenvalue is expected to be more representative than the others, as well as preferentially represented by all the aforementioned variables.
Next, eigenvalues and loadings are calculated. The latter represent the load of each variable, that is, how much each represents in determining the factor.

Moran's Local index is used for spatial analysis to measure the correlation structure between areas. Each Municipality has a spatial dependence level in relation to neighbouring Municipalities.
These indices are calculated for a certain attribute or variable. Moran's indices in the present study consider the rurality factor, calculated through factorial analysis. The aim is to identify spatial dependent clusters across counties in Italy.

3. Data

Data used in this study are from the main source of information in Italy that is the National Office of Statistics and are produced based on Censuses and sample surveys. The constructed data-set consists of 10 measurements on each Municipality of Italy. The total number of observations is 8101 based on the last available data for all the considered variables (period 2000-2010).
The covariates are as follows:

- land = total area (Km2).
- land_agr = total agricultural area (Km2).
- pop_res = resident population (number).
- pop_pre = present population (number).
- emp_tot = total employment (number).
- emp_agr = employment in agriculture (number).
- emp_agrw = working people in agriculture (number).
- emp_fish = working people in fishing (number).
- build = buildings (number).
- build_re = houses (number).

All of them are correlated to each other and proxy of several other characteristics of the areas.
Based on the previous variables and suggestions from the rural development literature (UN, 2005), the following indicators (RDI) where calculated:

- pop_d = share of present population over total area (population density).
- s_pop_ag = share of employment in agriculture over total employment.
- s_pop_fi = share of working people in agriculture over total work force.
- s_build = share of buildings over total area.
- s_land_a = share of agricultural area over total area.

4. Statistical Analysis

Tab. 1 - Descriptive Statistics of 10 Variables (N = 8101)

Variable	Mean	Median	Min.	Max.	Std. Dev.	Skewness	Kurtosis
LAND	37.196	21.8	.15	1285.	49.83	5.5336	70.213
LAND_AGR	24.201	13.3	0.00	517.	35.32	4.5312	32.933
POP_RES	7033.640	2345.0	0.00	2546804.	39326.86	42.2329	2372.918
POP_PRE	7049.765	2306.0	18.00	2624467.	40325.37	42.6148	2412.490
EMP_TOT	2591.499	854.0	10.00	1002523.	15019.96	.50567	2688.936
EMP_AGR	142.412	62.0	0.00	14782.	333.58	16.0661	524.727
EMP_AGRW	136.581	60.0	0.00	13819.	315.00	15.8311	508.434
EMP_FISH	5.831	1.0	0.00	1533.	39.96	21.1946	598.019
BUILD	3368.966	1267.0	41.00	1151736.	17900.03	42.2913	2358.557
BUILD_RE	3366.113	1267.0	41.00	1150547.	17881.04	42.2947	2358.910

Source: *elaboration on Istat data*

Basic descriptive statistics suggest that the shape of the distribution for the previous variables deviate from normality: median value is lower than the mean and skewness is clearly different from 0, then that distribution is asymmetrical; the kurtosis is clearly different than 0, then the distribution is more peaked than normal.

Tab. 2 – Correlation Matrix of 10 Variables (Marked correlations are significant at p < .05)

Variable	LAND	LAND_AGR	POP_RES	POP_PRE	EMP_TOT	EMP_AGR	EMP_AGW	EMP_FISH	BUILD	BUILD_RE
LAND	1.00	.92*	.38*	.38*	.58*	.37*	.58*	.28*	.39*	.39*
LAND_AGR	.92*	1.00	.23*	.23*	.50*	.23*	.51*	.15*	.24*	.24*
POP_RES	.38*	.23*	1.00	1.00*	.74*	.99*	.72*	.49*	.99*	.99*
POP_PRE	.38*	.23*	1.00*	1.00	.74*	.99*	.72*	.49*	.99*	.99*
EMP_TOT	.58*	.50*	.74*	.74*	1.00	.72*	.99*	.51*	.73*	.73*
EMP_AGR	.37*	.23*	.99*	.99*	.72*	1.00	.70*	.47*	.99*	.99*
EMP_AGRW	.58*	.51*	.72*	.72*	.99*	.70*	1.00	.42*	.71*	.71*
EMP_FISH	.28*	.15*	.49*	.49*	.51*	.47*	.42*	1.00	.49*	.49*
BUILD	.39*	.24*	.99*	.99*	.73*	.99*	.71*	.49*	1.00	1.00*
BUILD_RE	.39*	.24*	.99*	.99*	.73*	.99*	.71*	.49*	1.00*	1.00

Source: *elaboration on Istat data*

High correlation coefficients are a signal that some of the previous variables can be omitted without loss of information. couple of variable are:

- resident population and present population;
- employment in agriculture and working people in agriculture;
- buildings and houses.

Tab. 3 – Descriptive Statistics of 5 Indicators (N = 8101)

Variable	Mean	Median	Min.	Max.	Std. Dev.	Skewness	Kurtosis
POP_D	.277457	.102782	.000967	13.29358	.616930	8.06509	103.485
S_POP_AG	.100567	.071205	.000000	.68506	.091515	1.67542	3.534
S_POP_FI	.001859	.000302	.000000	.51365	.009716	28.43155	1177.981
S_BUILD	.127836	.055794	.001141	4.73761	.243695	6.84444	72.630
S_LAND_A	.062803	.064285	.000000	.88857	.034363	4.08870	65.780
Transformed indicators: natural logarithm (square root for s_land_a)							
POP_D	-2.21110	-2.27515	-6.94124	2.587282	1.316203	.176687	.181186
S_POP_AG	-2.72421	-2.64219	-6.90776	-.378256	.999032	-.444331	.109812
S_BUILD	-2.76027	-2.88609	-6.77559	1.555533	1.092806	.488904	.294762
S_LAND_A	.24092	.25355	.00000	.942637	.068994	-.201006	4.147594

Source: *elaboration on Istat data*

The five indicators also deviate from normality with a high asymmetry and peaks of their distribution.

Tab. 4 – Correlation Matrix of 5 Indicators (Marked correlations are significant at p < .05)

Variable	POP_D	S_POP_AG	S_POP_FI	S_BUILD	S_LAND_A
POP_D	1.00	-.25*	.01	.96*	.19*
S_POP_AG	-.25*	1.00	.10*	-.26*	.25*
S_POP_FI	.01	.10*	1.00	.03*	.06*
S_BUILD	.96*	-.26*	.03*	1.00	-.22*
S_LAND_A	-.19*	.25*	-.06*	-.22*	1.00

Source: *elaboration on Istat data*

The correlation coefficient show an low correlation of the share of employment in agriculture over total employment with the other variables and can be omitted in our model calculation.
The transformed indicators in Tab. 3 have a closer shape with respect to normality and coefficient of correlations remain significantly high (Tab. 5).

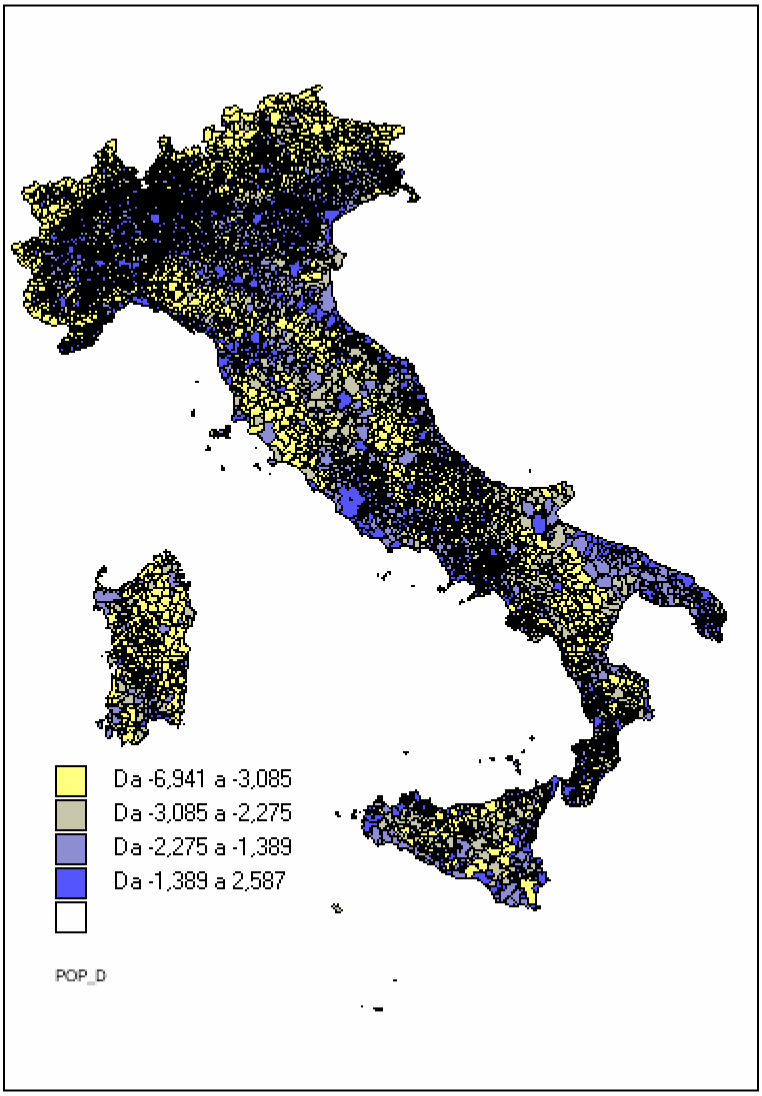
Tab. 5 – Correlation Matrix of 5 Transformed Indicators (Marked corr. are significant at p < .05)

Variable	POP_D	S_POP_AG	S_BUILD	S_LAND_A
POP_D	1.00	-.49*	.94*	.16*
S_POP_AG	-.49*	1.00	-.54*	.41*
S_BUILD	.96*	-.54*	1.00	-.27*
S_LAND_A	-.16*	.41*	-.27*	1.00

Source: *elaboration on Istat data*

Taking into account the population density the map of Italy divided into municipalities is the following:

Fig. 1 – Municipalities in Italy According to the Population Density



This indicator is not considered a sufficient one to identify rural areas: for example, Municipalities in Puglia, in the South of Italy, are densely populated but have a well developed agriculture and several rural characteristics according to rural areas definition (UN, 2005; 2011).

Factor analysis can be performed on Municipalities based on the previous 4 indicators. The following table shows the eigenvalue of the first principal component:

Tab. 6 – Eigenvalue and Variance Explained (%)

Factor	Eigenvalue	% total Variance
1	2.485775	62.14437

Source: *elaboration on Istat data*

Factor 1 explain 62.14% of the total variance and can be selected as a rurality index of Municipalities.

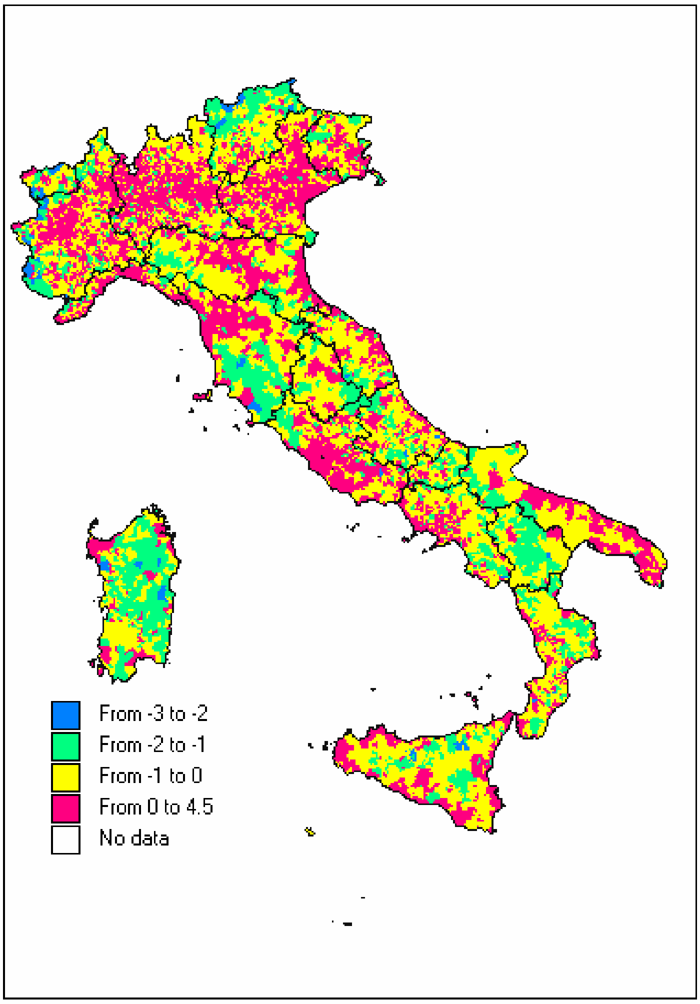
Tab. 7 – Variables Loadings for the First Factor

Variable	Factor 1 (Marked loadings are > .700000)
S_BUILD	.932545*
POP_D	.894806*
S_POP_AG	-.764067*
S_LAND_A	-.481308
Expl. Variance	2.485775
Prp. Total	.621444

Source: *elaboration on Istat data*

Resulting loadings indicate that all variables have a high weight in explaining factor 1. As such, this factor can be interpreted as an inverse indicator of rurality, in which the population density and the share of buildings over total area have a positive value. In other words, the higher the buildings density and population of a municipality, less it has the features of a rural area and it tends to be rural. With respect to the other two variables, share of employment in agriculture over total employment and share of agricultural area over total area, there is a direct relationship with rurality. In Figure 2 illustrated the map of Italy, divided in municipalities and regions, according to the calculated rurality index.

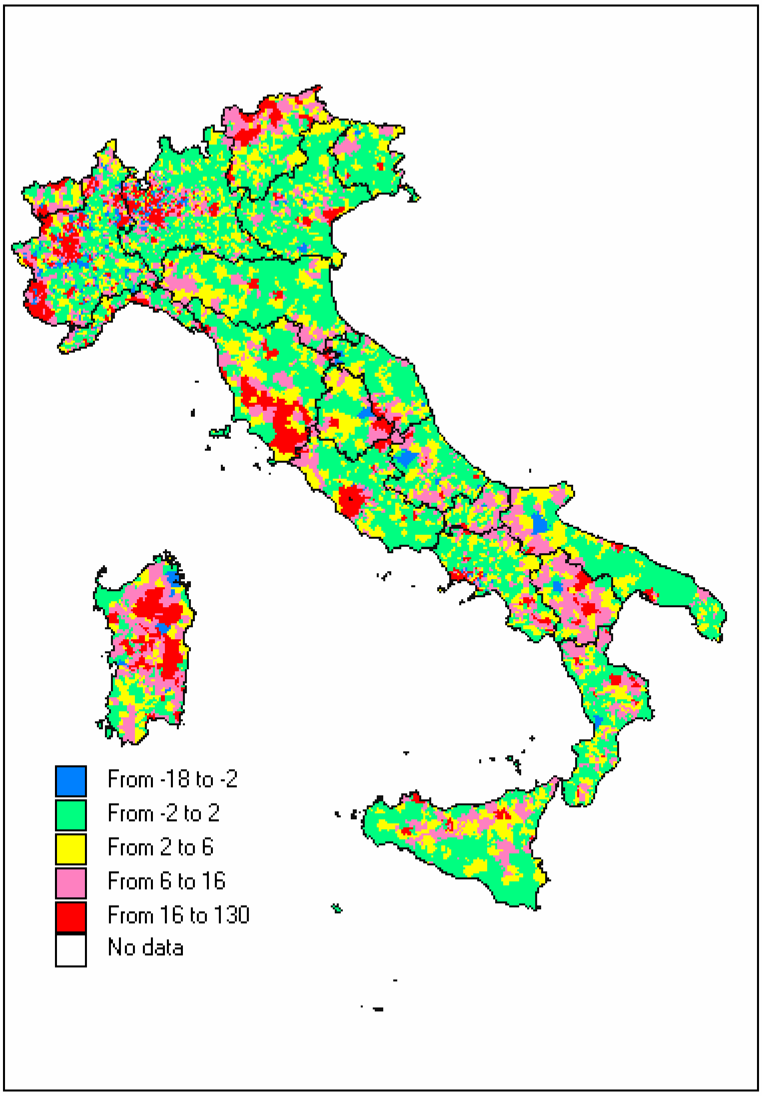
Fig. 2 – Municipalities in Italy According to the Rurality Index



With this approach, rural areas are concentrated in less populated areas, for example in mountain municipalities, but also in agricultural areas, for example in Emilia-Romagna, Tuscany, Umbria, Puglia and Sardinia regions.

Given this rurality distribution over municipalities, It is interesting to discover the integration of rural and urban settings where population resides and work. The spatial correlation can be measured with the local Moral index, indicating tendency to the formation of clusters of municipalities.

Fig. 3 – Municipalities in Italy According to Local Moran Indices



The map shows the existence of a strong spatial correlation among municipalities over large areas of Italy and across Regions, formed by both rural and non-rural administrative areas. As expected, spatial dependency is much more intense in the metropolitan regions with greater connections in economic activities and movement of people, with respect to geographically isolated mainly rural areas.

5. Conclusions

The results of this empirical study at Municipality level in Italy, on rurality estimation of administrative areas and spatial correlation in larger clusters of administrative areas, are very promising from different perspectives.
The degree of rurality has been assessed based on a multivariate analysis approach, at a low geographical level and the territorial aggregations produced with Moran index become very informative on territorial relationships and areas homogeneity.
As a next step, an aggregation criteria should be developed to move from lower to larger administrative areas classification (from NUTS 3 to NUTS 2).
This approach demonstrates to be informative and to consider the complexity of rural-urban interrelation at territorial level; the need of a set of indicators to take account of multidimensionality of "rurality", without any oversimplification in the classification of administrative areas.
Further results and analysis should be produced on specific study areas to understand the power and reliability of this multivariate approach.

References

[1] Benzécri, J.-P. et Coll. (1986) *Pratique de l'analyse des données*, Economie, Dunod, Paris, France.
[2] Greenacre, M. and Blasius, J. (2006) *Multiple Correspondence Analysis and related techniques*, eds, Chapman and Hall, London, UK.
[3] Husson, F., Lê S., et Pagès J. (2009) *Analyse de données avec R*, Presses Universitaires de Rennes, Rennes Cedex, France.
[4] Lebart, L., Morineau , A. and Warwick, K. (1984) *Multivariate Descriptive Statistical Analysis*, John Wiley and sons, New-York, USA.
[5] Lebart, l., Morineau a. et Fénelon, J.P. (1986) *Traitement des Données Statistiques*, Dunod, Paris, France.
[6] Moreau, J., Doudin, P-A. et Cazes, P. (2000) *L'analyse des correspondances et les techniques connexes*, Mathématiques et Applications, Springer, Berlin, Germany.
[7] Pizzoli, E. and Palmegiani, G. (2007) *Rural Development Statistics (RDS) for Policy Monitoring: a Rural-Urban Territorial Classification and Farmers' Income Data*, African Commission on Agricultural Statistics (AFCAS), Alger, Algeria. www.fao.org/es/ess/meetings
[8] Pizzoli, E. and Gong, X. (2007) *How to Best Classify Rural and Urban?*, Forth International Conference on Agriculture Statistics (ICAS-4), National Bureau of Statistics of China, Beijing, China www.stats.gov.cn/english/icas
[9] Pizzoli, E. (2008) *Rural Area Definition*, First Wye City Group Meeting, York, UK. www.fao.org/fileadmin/templates/ess/pages/rural/wye_city_group/2008/index.htm
[10] Pizzoli, E., Pianura, P. and Piersante, A. (2009) *Implications of Rural Areas Definition on Households Income Analysis: the Mediterranean Area*, Società Italiana degli Economisti Agrari (SIDEA), Napoli, Italy. www.depa.unina.it/sidea2008
[11] Pizzoli, E., Gianfaldoni, V. (2009) *Social Well-Being, Economic Development and Sustainability in Rural and Urban Areas: a Comparison of Indicators*, Third Wye City Group Meeting, Washington, USA.
typo3.fao.org/fileadmin/templates/ess/pages/rural/wye_city_group/2010/3rd_Wye_Conference.html
[12] Pizzoli, E. (2010) *Implications of spatial aggregation on economic variables: probability distribution forms and bias*, 10th Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius, Lithuania. www.vilniusconference10.com
[13] OECD (1994) *Creating Rural Indicators for Shaping Territorial Policy*, Paris, France.
[14] OECD (2002) *OECD Territorial Reviews: Canada*, Paris, France.
[15] OECD (2006) *The New Rural Paradigm: Policies and Governance*, Paris, France.
[16] OECD (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide*, Paris, France. www.oecdbookshop.org/oecd/display.asp?K=5KZN79PVDJ5J&DS=Handbook-on-Constructing-Composite-Indicators-Methodology-and-User-Guide
[17] OECD (2009) *Rural Policy Review: Italy*, Paris, France.
[18] Rizzi, A (1995) *Analisi dei dati. Applicazioni dell'informatica alla statistica*, La Nuova Italia Scientifica, Roma, Italy.
[19] UN (2005) *Rural Households' Livelihood and Well-Being. Statistics on Rural Development and Agriculture Household Income*, Geneva, Switzerland.
www.fao.org/fileadmin/templates/ess/pages/rural/index.htm

[20] UN (2011) The Wye Group Handbook - Statistics on rural development and agricultural household income, 2nd edition, FAO, Rome, Italy.
www.fao.org/fileadmin/templates/ess/pages/rural/wye_city_group/wyehandbook.html