Identification of Patterns of Socio-Economic and Spatial Processes via Principal Component Analysis and Clustering Algorithms. The Case Study of Southern Italy

Alfonso Annunziata(✉), Francesco Scorza, Simone Corrado, and Beniamino Murgante

University of Basilicata, Potenza, Italy
annunziata.alfonso@yahoo.it

Abstract. The Implementation of place-based strategies and governance procedures to intervene in inland regions presenting critical socio-economic conditions is a central aspect of cohesion policies at the national and European level. The identification of configurations of socio-economic and spatial conditions specific to distinct areas is a central element for defining targeted policies that combine mitigation and adaptation strategies. The study proposes a set of metrics, based on validated open data set, and combines Principal Component Analysis (PCA) and clustering functions to discerne regions presenting specific configurations of social, economic and spatial conditions. Six Italian regions, including Abruzzo, Basilicata, Calabria, Campania, Molise and Puglia are selected as study areas. The Municipalities are selected as the unit of analysis. The aim of this study is to define a set of variables to understand the social, economic and spatial conditions specific to distinct cluster of municipalities and to discern distinct types of marginal declining areas. The findings of the study provide relevant information for the definition of targeted policies and for balancing adaptation and mitigation strategies.

Keywords: Inland Areas · Principal Component Analysis · Spatial Analysis · Clustering · Open Data

1 Introduction

The decline and marginalization of rural internal areas are emerging as a central issue of European and national policies. The 2019 Politcal Guidelines for the European Commission underline the centrality of rural areas as the place of residence of more than 50%

This paper is the result of the joint work of the authors. In particular, the data collection and analysis were conducted by Alfonso Annunziata and Simone Corrado and the methodological approach was conceived by Beniamino Murgante and Francesco Scorza. 'Discussion and Conclusions' was written jointly by the authors. Simone Corrado wrote the 'Literature Review'. The sections 'Introduction' and 'Results' were written by Alfonso Annunziata. Beniamino Murgante and Francesco Scorza wrote the 'Methodology' section and supervised the study.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14110, pp. 52–69, 2023.
https://doi.org/10.1007/978-3-031-37123-3_5

of European population, as a central element of European identity, society and economy. As a result, the preservation of rural areas landscape, and cultural, social and economic structures, and investment in rural areas' development emerge as relevant aims of European policies [1]. Similarly, Italian strategy for internal areas underlines the centrality of policies aimed at increasing social inclusion, development, and quality of life of internal areas, due to their extension - accounting for 60% of the entire country's surface area, including 52% of the municipalities and 22% of the population - as well as the significance of the environmental and cultural resources located in the rural landscapes. European and National Institutions recognize depopulation, marginalization, locational disadvantage and decline of economic activities as the main issues facing rural areas, and as the target of cohesion policies.

Moreover, consensus emerges on the need of targeted and coordinated strategies, integrating adaptation and mitigation. This perspective is based on the conceptualization of rural decline as a multi-dimensional process resulting from combinations of social, economic and spatial trends specific to individual areas. Hence, the recognition of local specificity is a pre-condition for the formulation of policies. The study proposes a set of metrics, based on validated open data set, and combines Principal Component Analysis (PCA) and clustering functions to discern regions presenting specific configurations of social, economic and spatial conditions. The proposed article describes the preliminary stages of a study, conducted within the framework of the research Project Mitigo, and aimed at investigating the conditions of marginality in the municipalities of the Basilicata Region, resulting from of recognizable spatial and socio-economic tendencies. Six Italian regions, including Abruzzo, Basilicata, Calabria, Campania, Molise and Puglia are selected as study areas. The Municipalities are selected as the unit of analysis. The aim of this study is to develop a set of metrics to discern the social, economic and spatial conditions specific to individual clusters of municipalities.

The article is articulated on five sections. After the introduction, the scientific literature on the topic of rural areas' decline and the research method are described in Sects. 2 and 3. Section 4 presents the results of the analysis. Lastly, the findings of the study, its limitations and its future development are presented in the Conclusions section.

2 Literature Review

Rural depopulation is a relevant topic of national and European policies, of the political debate and of studies in the field of social sciences, environmental science, engineering and agricultural and biological science.

The analysis of the scientific literature, conducted via the Web of Science database and the Bibliometrics tool, returns 273 items containing, in the field topic, the terms sequence 'rural depopulation'.

The selected corpus of articles is cited 7,693 times, excluding self-citations, by 6822 articles, resulting in an average number of citations per item equal to 28.89, and in a H-index of 35. The most frequent terms, in the keywords field, are rural depopulation (55), depopulation (31), agriculture (10), land abandonment (10), rural development (10), land use (7), migration (7), rural population (6), urbanization (6) and land use change (5). Moreover, the analysis of the co-occurrence structures, underline the significant

relation among the terms rural depopulation, land abandonment, land use, migration, urbanization, farmland abandonment, school closure, soil erosion, abandoned hamlet, land degradation. These interdependencies underline an merging focus on the relation among migration, population decrease, closure of basic services, and the depopulation's impact on environmental structures, in terms of land abandonment and degradation and soil erosion. Yet, rural depopulation is a component of a multidimensional process of rural decline or of 'rural shrinkage' [2].

Grasland et al. [3] define rural shrinkage as a process resulting in a significant decrease of population over a period equal to or superior than a generation. A distinction is to be made discerning active shrinkage, as a result of out migration, and legacy shrinkage, intended as a consequence of natural population decrease and regions' age structure. Moreover, out-migration can be determined by regional rural-urban processes or by transnational, globalized movements. A more relevant distinction discerns simple shrinkage, related to simple population decrease, and complex shrinkage, resulting from the interaction of social, economic and spatial processes [4, 5]. Copus et al. [5] identify four types of complex decline. These types of shrinkage processes can be co-present and interact within a region, determining specific configurations of cumulative causation of rural depopulation and decline. The four identified types of complex shrinkage processes are denoted as: economic restructuring, locational disadvantage, peripherisation and disruptive events. More precisely, economic restructuring refers to the decline in traditional labour-intensive agricultural practices and in extractive and manufacturing activities, resulting in land abandonment. Land abandonment, in turns, is related to relevant impacts on social, cultural and economic structures and on environmental systems, including depopulation, land degradation and soil erosion, decrease in biodiversity, and loss of cultural landscapes [6–8]. Locational disadvantage refers to local negative conditions, including isolation, sparsity, scarcity of natural resources, fragmentation of agricultural land, minor land productivity, determining negative impacts on economic development [7, 9, 10]. Peripherisation regards the macro-scale processes of spatial reorganization of economic activities and globalization [4]. Lastly, disruptive events, include historical events, political regime transitions, or political integration process. The Eu integration process, and related transnational rural and agricultural policy initiatives, the end of the socialist regimes era, and wars in the Balkan and in Ukraine represent major disruptive events [5, 6, 11, 12].

The complex interaction of social, economic and cultural factors related to rural depopulation is recognized in the definition of internal areas proposed by the Italian National Strategies for Internal Areas (SNAI). Internal areas are defined as regions that present conditions of marginalisation, in terms of distance from services of education, healthcare, and transportation, and where trends of contraction of human presence are recognizable. These trends result in trends of depopulation, population ageing, unemployment, land abandonment and decrease in the use of local capital. Lastly, a further aspect concerns the evolution of European rural and agricultural policies. This evolution can be described in terms of a transition on two directions: from an exogenous paradigm based on outer inputs aimed at sustaining rural economies and favour population retention to endogenous and neo-endogenous solutions, aimed at supporting the valorization of the territorial capital; and from mitigation strategies, aimed at inverting depopulation

tendencies, to adaptation strategies aimed at increasing residents' quality of life in the context of the inevitability of local processes of rural decline. A consensus emerges in the political debate and in the scientific discourse, on the need of targeted strategies that combine adaptation and mitigation policies [4, 5, 13–15]. In order to identify areas presenting specific socio-economic trend and spatial processes, the subsequent section presents a method for discerning spatial clusters based on socio-economic and spatial variables.

3 Methodology

The study develops a set of indicators of socio-economic and spatial variables and utilizes clustering algorithms and statistical analysis to identify clusters of municipalities presenting similar conditions. The study is articulated on seven stages: i) selection of the area of study and of the unit of analysis; ii) definition of categories of socio-economic and spatial variables; iii) definition and measurement of relevant indicators; iv) Implementation of Principal Component Analysis (PCA) to reduce the dimensionality of the variable space; v) Construction of the elbow plot and determination of the optimal k-number of clusters, applied for implementing the K-means clustering function; vi) implementation of the K-means clustering function and categorization of the units of analysis into k distinct clusters; vii) exploratory statistical analysis of the distribution of the values of a sub-set of indicators within each cluster (see Fig. 1). The area of study includes six Italian region – Abruzzo, Basilicata, Calabria, Campania, Molise and Puglia - comprised in the macro-area South Italy defined by the Italian National Institute for Statistics. The 1783 Municipalities comprised in the study area, are selected as the unit of analysis. The focus on the municipal scale is aimed at recognizing broad regional processes and configurations of socio-economic and spatial variables, and at the same time, at providing a detailed description of the area of study that captures local specific conditions and trends. The socio-economic and spatial conditions investigated include conditions of access to services, education, Residential buildings, individual socio-economic status, digital infrastructure, Municipal infrastructures, Agriculture, Service and Industry sectors, and population (see Table 1). In particular, Access to basic services is measured via indicators of proximity to four categories of services: rail stations, hospitals, educational facilities, pharmacies.

Proximity is formalized as the number, weighted by distance, of services of the i-th category, located within a pre-determined distance from the centroid of the administrative area of each municipality. The radius is equal to 40 km. This value is identified as a relevant distance limit for evaluating proximity to occasionally accessed functions [16]. The four indicators of access to services AR40ST, AR40ED, AR40OSP, AR40FAR are calculated via the Attraction reach function of the Place Syntax Tool (PST) [17]. The Education system is described in terms of Percentage of resident population with a secondary education degree and in terms of Percentage of resident population with a tertiary education degree, measured for the year 2020. These data are retrieved from ISTAT open dataset. Moreover, the Quantity of services offered by the municipality compared to the average of municipalities comparable in terms of population, is utilized as a metric of public investment in reducing inequality in the education sector.



Fig. 1. Methodology and organization of the study

Table 1. Indicators of socio-economic and spatial tendencies in the area of study

Category	Indicator Code	Indicator Definition
Access to Service	AR40ST	Coverage of transportation services (RFI Stations), Radius 40 km
	AR40ED	Coverage of Education facilities, Radius 40 km
	AR40OSP	Coverage of healthcare services (Hospitals), Radius 40 km
	AR40FAR	Coverage of healthcare services (Pharmacies), Radius 40 km

(*continued*)

Table 1. (*continued*)

Category	Indicator Code	Indicator Definition
Population	RP6520	Proportion of the Population over 65 years of age, in 2020
	DPOP	Variation of the resident population over the 1981–2020 period
	P1000ED	Number of residents per 1000 housing units
	SMI_19_02	Aggregated Internal Migration Balance over the 2002–2019 period
	SME_19_02	Aggregated External Migration Balance over the 2002–2019 period
	SN_19_02	Aggregated Natural Balance over the 2002–2019 period
Socio-economic Status	R_C_10	Proportion of resident population with annual income in the range 0–10000 € - 2020
	R_IMP	Average annual income per taxpayer - 2020
	R_INAT	Proportion of inactive people to the resident population
Access to Internet	SP_GRE	Average Maximum download speed
	SP30HH	Number of Households served by internet network with speed in the range 0–30 Mbps
Municipal Infrastructures and Facilities	D_OUTGN	Amount of services offered by the municipality compared to the average of municipalities with equivalent resident population - 2018
	SSTDABGN	Standard Expenditure per resident for the provision of basic services - 2018
	SSTOABGN	Historical Expenditure per resident for the provision of basic services - 2018
Building Stock	P_ED_DIS	Proportion of available housing unitis to total housing units - average 2015,2016,2017,2018 (%)
	QIR22	Average Quotation of Housing units - 2022

(*continued*)

Table 1. (*continued*)

Category	Indicator Code	Indicator Definition
Education	D_OUT_IS	Amount of educational services offered by the municipality compared to the average of municipalities with equivalent resident population - 2018
	SSSG	Percentage of resident population with a secondary education degree - 2020
	TSTRZ	Percentage of resident population with a tertiary education degree - 2020
Agriculture	N_FARM_D20	Number of Agricultural Enterprise per Km^2 - 2020
	R_AGR18	Proportion of Agricultural land to the total surface area of the municipality - 2018
Local Economy – Industrial and Service sectors	VA_ADD	Value added per employee of local units (apparent labor productivity)
	ADD_MOL+	Percentage of employees in local units with a non-negative corrected Gross Operating Margin
	TUR1000A	Tourist presences - average 2015,2016,2017,2018 – Number of tourists per 1,000 residents

Lastly, demographic factors include population variation over the 1981–2021 period, percentage of population of age superior to 65 years, internal migration balance, external migration balance, and natural balance, calculated over the 2002–2021 period. Data related to population trends, level of education and to productivity of industrial activities are retrieved from the ISTAT open data set. Data related to investment for services at the municipal scale are retrieved from the Open Civitas data set. Data related to the number of farms are retrieved from the Italian Agricultural Accounting Information Network (RICA) data set. Lastly, data related to individual income are retrieved from the open data set of the Italian Ministry of economy and finance. For each variable, the most recent data, or the data covering the broadest time span are considered.

The dimensionality of the variable space is then reduced via techniques of principal component analysis, implemented via the GeoDa tool. The singular value decomposition, and the z-transformation are selected as the method and the standardization functions for the determination of the Principal components. Each principal component is calculated as a linear combination of the original variables, of the i-th category, that maximizes the explained variance of the set of the original variables [18]. The principal components are selected as the input variables for grouping the 1783 municipalities into k clusters, via partitioning clustering functions. The objective of Clustering functions is to determine compact set of similar observation, such that the intra-cluster similarity is maximized

and the similarity among cluster is minimized. Similarity is formalized in terms of distance, in the multi-dimensional variable space, among each pair of observations i and j. A distinction is made between the distances of each observation in the j-th cluster from the other observations contained in the same cluster (intra cluster dissimilarity) and the distances of each observation in cluster j-th from the observations not contained in the j-th cluster (between-cluster dissimilarity). In the case of K-means clustering functions, which are utilized for this study, dissimilarity is formalized as the sum of squared Euclidean distances across the p-dimensions of the variable space. As a consequence, the objective of minimizing the loss function, is formalized as minimizing the sum of squared distances among observations within each cluster or, alternatively, as reducing the sum of squared deviations from the mean in each cluster, the within sum of squared errors. The compactness of the clusters is then assessed by calculating the ratio of the between clusters sum of squared errors to the total sum of squared errors. This measure is denoted as the Sum of Squares Ratio. Superior values denote an improved separation of the clusters. The value of the sum of squares ratio increases with the number of clusters, k. As a result, the definition of the optimal number k of clusters is a central aspect for the implementation of the k-means clustering function. An elbow-plot is calculated for representing the variation of the sum of squares ratio against the value of k. The rationale is that the optimal value of k is indicated by the point of the curve at which the improvement of the fit decreases. In the last stage, a set of measures of central tendency and variability of the original variables is calculated for each cluster and scatter-plots and probability distributions are used to model the distribution of data values of the original variables within each cluster. The objective is to infer spatial and socio-economic tendencies, specific to each cluster of municipalities.

4 Results

The Principal component analysis, reduces the dimensions of the variable spaces to 13 principal components. The set of principal components is selected by using the 95% limit criterion, that is, by selecting the minimum set of principal components required to explain the 95% of variance in the set of the original variables. The determination of the optimal number k of clusters returns a k equal to 7. As a consequence, The K-means clustering function is implemented to group the 1783 units of analysis into seven clusters, based on the similarity of the social, economic and spatial conditions, represented by the 13 principal components. A second clustering is implemented, considering a number of clusters equal to 8. The results, for $k = 7$ presents a sum of squares ratio equal to 0.499, thus indicating an adequate level of separation of the sets of units of analysis. For $k = 8$, the sum of squares ratio is equal to 0.517 (Table 2).

A preliminary interpretation of the results underlines the emergence of a compact cluster of municipalities contiguous to the City of Naples as a recognizable set (see Fig. 2). A second cluster, consisting of 202 municipalities, include the metropolitan area of Bari, and major urbanized areas, including the cities of Avellino, Battipaglia, Campobasso, Catanzaro, Chieti, Foggia, Isernia, L'Aquila, Lecce, Matera, Pescara, Potenza, Reggio Calabria, Salerno, Vibo Valentia, and industrial poles, including Melfi. Moreover, compact sub-sets comprised in cluster 3 can be recognized, including municipalities in the plain areas of the Apulia Region and of the Calabria Region. More precisely,

Table 2. Centroids of the Clusters

Principal Components	Clusters						
	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
PC1_SPA	−0.83	−0.51	−0.27	0.40	−0.86	6.72	−0.45
PC1_EDU	−0.89	0.68	−0.29	1.56	−0.40	−0.20	0.88
PC2_EDU	0.20	0.39	−0.44	−0.24	0.46	−0.93	0.10
PC1_DIG	−0.60	−0.73	0.94	0.77	−0.94	1.87	−0.16
PC1_SERV	−0.05	0.12	−0.57	−0.26	2.35	−0.78	0.98
PC1_BLD	−0.55	−0.27	0.02	1.04	−0.91	1.90	2.25
PC1_SES	−1,33	0.66	−0.11	2.26	−0.98	0.74	0.32
PC1_AGR	−0.29	−0.13	0.63	0.29	−0.75	0.15	−0.80
PC1_ECO	−0.36	0.46	−0.18	0.27	−0.48	0.22	1.42
PC2_ECO	−0.10	−0.13	−0.04	−0.12	0.02	−0.27	5.82
PC1_POP	1.25	0.39	−0.97	−1.99	2.92	−3.19	−0.52
PC2_POP	−0.38	0.27	−0.22	0.42	0.84	−0.43	0.47
PC3_POP	0.00	−0.18	0.27	−0.61	0.04	0.75	−0.04

	Within Cluster Sum of Squares						
	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
	4023.08	3828.89	3671.93	2828.34	2277.98	1991.59	811.66
	Sum of Squares Ratio						
SSR	0.499024						

areas comprised in Cluster 3 can be qualified as areas of agricultural production. The municipalities comprised in cluster 3 present, in fact, a superior average density of farms. Mountain areas, presenting a median average elevation of 551.86 and of 502.36 m above sea level, are comprised in cluster 1, 2 and 5. Lastly, Cluster 7 comprises 30 municipalities presenting a significant tourism potential, denoted by values of the indicator average number of tourists per 1000 residents, ranging from 83.79 to 602.15 and resulting in a mean value of 254.99. Decrease of population is more significant in municipalities comprised in clusters 1,2 and 5. More precisely, the mean and median variation of population over the 1981–2021 period are equal to −0.311 and −0.322 for cluster 1 municipalities, to −0.119 and to −0.127 for cluster 2 municipalities and −0.473 and −0.489 for area comprised in Cluster 5. Population is stable in plain areas (mean equal to 0.036 and median equal to 0.010) and in the 30 municipalities comprised in Cluster 7 (mean equal to 0.057 and median equal to 0.064 respectively). A significant increase of population is observed in urbanized areas comprised in Cluster 4 and 5.

The mean is equal to 0.363 and the median to 0.280 in regional-scale poles included in Cluster 4 and to 0.351 and to 0.258 respectively in the Municipalities comprised in the Naples metropolitan area (see Table 3).

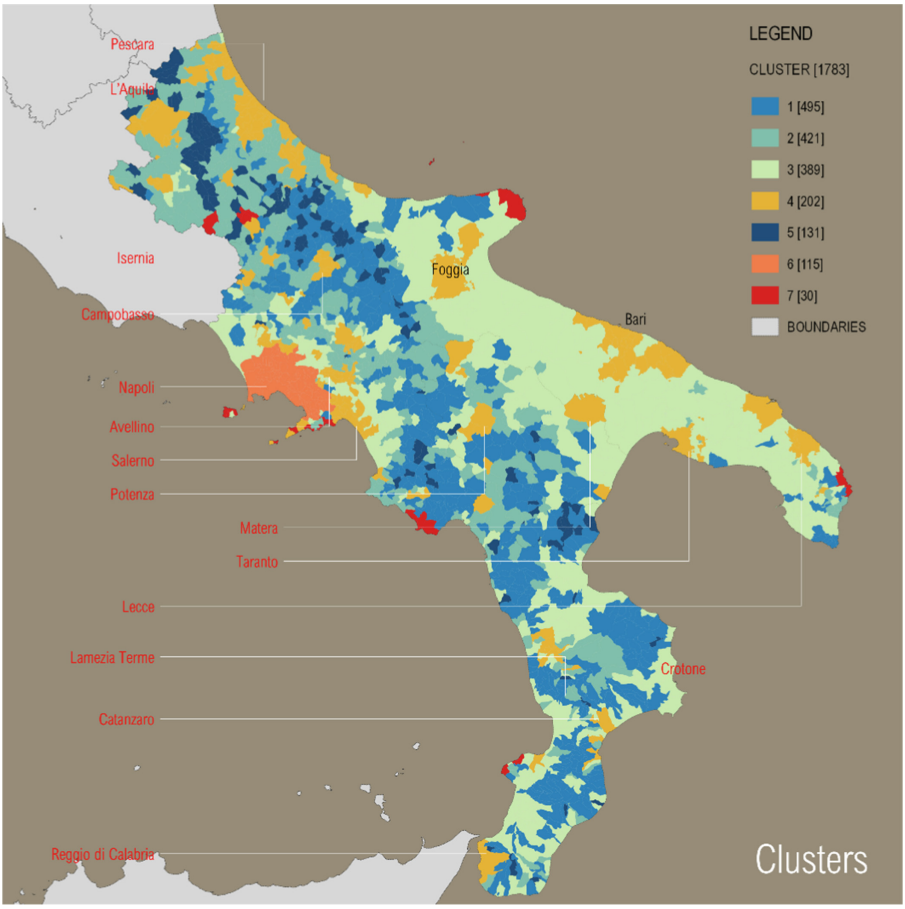

Fig. 2. Visualization of the seven clusters determined via a K-means Clustering algorithm

Table 3. Measures of Central tendency and of variance

D_POP							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
1	−0.660	0.221	0.881	−0.311	−0.322	0.138	0.187
2	−0.653	0.450	1.103	−0.119	−0.128	0.170	0.242
3	−0.407	2.573	2.980	0.037	0.010	0.223	0.195
4	−0.389	2.588	2.977	0.364	0.281	0.414	0.453
5	−0.770	0.078	0.848	−0.473	−0.490	0.140	0.175
6	−0.338	1.718	2.056	0.352	0.258	0.417	0.450

(*continued*)

Table 3. (*continued*)

D_POP							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
7	−0.299	0.792	1.091	0.058	0.065	0.222	0.296

RP6520							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
1	0.192	0.442	0.250	0.280	0.276	0.039	0.052
2	0.178	0.361	0.183	0.258	0.257	0.033	0.043
3	0.142	0.303	0.161	0.220	0.220	0.028	0.040
4	0.119	0.287	0.168	0.212	0.212	0.028	0.038
5	0.230	0.564	0.334	0.349	0.339	0.065	0.073
6	0.096	0.245	0.149	0.169	0.167	0.028	0.041
7	0.185	0.287	0.102	0.234	0.235	0.024	0.030

SN_19_02							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
1	−320.65	30.61	351.26	−111.47	−107.93	56.72	75.08
2	−230.97	38.71	269.68	−89.09	−86.55	49.17	68.31
3	−122.20	168.85	291.05	−7.22	−10.68	43.97	61.17
4	−132.72	228.75	361.47	8.06	6.55	49.01	64.24
5	−560.10	−20.35	539.74	−232.59	−217.76	90.93	117.08
6	−38.66	379.64	418.30	74.66	68.60	56.64	71.63
7	−146.63	60.40	207.03	−26.67	−31.07	47.60	67.84

SMI_19_02							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
1	−399.41	126.81	526.21	−108.88	−108.38	59.85	74.07
2	−213.10	313.50	526.60	−49.59	−58.78	65.71	71.87
3	−310.13	246.43	556.56	−61.71	−62.28	64.52	67.98
4	−144.35	492.37	636.71	42.29	28.70	104.60	135.85
5	−914.29	245.49	1159.77	−116.50	−117.81	126.11	120.78
6	−238.52	677.69	916.21	−47.66	−67.65	107.72	99.25
7	−153.18	64.52	217.69	−34.70	−39.38	48.28	76.39

AR_40_0SP							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
1	0.00	10.90	10.90	1.97	1.63	1.57	1.96
2	0.00	24.23	24.23	3.92	2.90	3.14	3.74
3	0.00	25.59	25.59	5.24	4.03	4.51	4.93

(*continued*)

Table 3. (*continued*)

AR_40_0SP							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
4	0.04	36.89	36.85	9.61	8.81	6.29	9.02
5	0.00	9.89	9.89	1.81	1.32	1.73	1.56
6	20.76	70.45	49.70	43.96	43.94	13.55	22.39
7	0.00	9.51	9.51	4.04	3.22	2.99	5.38

AR_40_EDU							
CLUSTER	MIN	MAX	RANGE	MEAN	MEDIAN	STDDEV	IQR
1	0.00	56.14	56.14	8.43	6.08	8.04	7.11
2	0.00	119.53	119.53	18.07	13.75	15.39	18.86
3	0.00	117.71	117.71	25.27	18.35	22.07	25.75
4	0.00	156.69	156.69	45.86	37.92	29.71	41.49
5	0.00	40.05	40.05	6.41	4.45	6.26	4.48
6	108.64	348.57	239.93	217.96	218.12	64.89	112.69
7	0.00	47.22	47.22	19.51	19.93	15.67	28.31

In terms of proportion of the elderly population on the total population, the measures of central tendency for individual clusters underlines the critical conditions of municipalities comprised in cluster 5. In these areas, in fact, the mean value of the proportion of the elderly population is equal to 0.348 and the median to 0.339. Values, in particular, range from 0.23, observed in Monteleone di Puglia to 0.564, measured in San Benedetto in Perillis. As a result, 105 out of 131 municipalities comprised in cluster 5 present a proportion of elderly population superior to 0.3. This value represents a critical point, indicating the loss of the municipality's endogenous capacity to maintain adequate levels of demographic vitality. Critical conditions are also observed in municipalities comprised in Clusters 1 and 2. The mean and median values of the proportion of the elderly population are equal, respectively, to 0.280 and 0.276 in Cluster 1 municipalities, and to 0.258 and to 0.257 in Cluster 2 areas. Values are comprised in the ranges 0.192–0.442 and 0.178–0.361. A distinct condition is observed in the Metropolitan areas of Naples. The mean and median values are equal to 0.167, and the proportion of the elderly population ranges from 0.096 to 0.245. A significant component of population decrease is represented by the natural balance. This is measured via an aggregated natural balance metric, calculated as the proportion of the aggregated natural balance over the 2002–2019 period on the total population measured in the year 2002 (see Table 3). A significant, negative trend is observed in the municipalities comprised in Clusters 1,2 and 5 representing mountain internal areas. In particular the mean and the median are equal to -0.232 and to -0.217 in Cluster 5 municipalities. Viceversa, urbanized areas comprised in Cluster 4 present a stationary condition (mean equal to 0.008 and median equal 0.006) and the Naples metropolitan area present a moderate positive natural trend (mean equal to 0.074 and median equal to 0.068). A significant variance of values of the internal migration balance is observed for all clusters. Yet, a relevant negative trend can be recognized

in Cluster 5 and in Cluster 1: In cluster 5 74 out of 131 municipalities (56%) present an Internal migration inferior than −0.100. In cluster 1 275 municipalities out of 495 present a significant negative internal migration balance. Lastly, the external migration balance underlines a general positive trend, in particular in the Naples metropolitan area (mean equal to 0.048 and median equal to 0.038, in Cluster 3 areas (mean and median are respectively equal to 0.050 and 0.041), in urbanized areas included in Cluster 4 (mean equal to 0.056 and median equal to 0.048) and in internal areas comprised in Cluster 5 (mean equal to 0.093 and median equal to 0.072). Clusters 1, 2 and 5 present also critical conditions in terms of access to basic services. In particular, the average number of hospital in a 40 km range is equal to 1.97 for Cluster 1 Municipalities, to 3.92 for cluster 2 areas and to 1.81 for municipalities comprised in Cluster 5. Municipalities comprised in the Naples metropolitan area present the most favorable conditions of access to hospital. The average number of destinations in the 40 km range is equal, in fact, to 43.96. Lastly, a similar condition emerges regarding educational facilities. The average number of destinations is equal to 8.43 for areas in cluster 1, to 18.06 for municipalities in cluster 2, to 6.41 for areas included in Cluster 5. Viceversa, favorable conditions of access to education is observed in the Naples metropolitan area (mean equal to 217.96) and in urbanized areas in Cluster 4 (mean equal to 45.86).

5 Discussion and Conclusion

Clustering functions emerge as relevant tools for investigating processes that incorporate a spatial dimension. In particular, the analysis demonstrates the relevance of the proposed metrics and of k-means clustering functions as a tool for preliminary analyses aimed at discerning specific socio-economic processes co-present in a region.

The results, in fact, underline seven specific configurations of spatial and socio-economic conditions: i) a vast metropolitan area, emerging as a supra-local centrality; ii) a set of urbanized regions, emerging as local poles of services, and presenting co-presence of agricultural and industrial activities; iii) municipalities presenting a significant tourism potential; iv) areas presenting a significant concentration of agricultural activities; v) internal areas presenting distinct forms of decline and marginalisation.

As a result, the clustering functions, based on the principal components derived from the set of socio-economic and spatial variables, discern as specific clusters the declining rural areas, and, more precisely, identifies three distinct forms of decline. Municipalities comprised in cluster 5, in particular, present the most significant trends of population decrease and ageing. The decrease and ageing of population is reflected by economic conditions: Economic metrics, in fact, underline the superior incidence of income deriving from retirement pensions in cluster 5 areas (mean proportion of pension income equal to 0.397) and in cluster 1 area (mean equal to 0.358). Similarly, proportion of inactive individual is significant in cluster 5 areas, and is equal to 0.592. Viceversa, the proportion of income deriving from employment is inferior, and equal, on average, to 0.279 in cluster 5 municipalities and to 0.326 in Cluster 1 areas.

Moreover, areas comprised in clusters 1 and 5 present inferior average income, equal, respectively, to 12827.31 € per year and to 13585.70 € per year, and a superior proportion of residents earning an income of less than 10000 euros per year, equal, on

average, to 0.442 in Cluster 5 municipalities and to 0.472 in Cluster 1 municipalities. Lastly, the level of productivity is inferior in Clusters 1 and 5, and is equal, on average, to 23.01 and to 23.45.

Viceversa, Cluster 2 municipalities present a moderate decrease of population (mean value equal to −0.118), combined with a relevant proportion of the elderly population (mean value of 0.258), and, compared to Clusters 1 and 5 areas, moderately inferior proportion of residents earning less than 10000 € per year, minor proportion of pension income (mean of 0.304), superior average income, equal to 15579.63 € per year, and productivity (equal to 27.15).

A preliminary interpretation of the results can lead to the qualification of the clusters of internal mountain areas in terms of distinct levels of decline: Cluster 5 municipalities reveal a condition of decline considerable as non-reversible, Cluster 1 areas reveal an advanced level of decline, and Cluster 2 a condition of mild decline. The discerning of distinct forms of decline underlines the need for targeted policies, based on specific combinations of adaptation and mitigation strategies. Considerations based on these preliminary results indicate the prevailing need, in Cluster 5 areas, for adaptation strategies, hence, for policies aimed at increasing the quality of life of residents. In Clusters 1 and 2 a combination of actions aimed at increasing quality of life and of mitigation strategies aimed at promoting economic development at limiting tendencies to depopulation can be considered as favorable. In particular, mitigation interventions in municipalities comprised in Cluster 2 and in Cluster 3 could privilege mitigation strategies aimed at valorizing the local cultural, natural and landscape capital and the residual economic potential, related to the agricultural and tourism sectors.

Inadequate conditions of access to basic services emerges as a relevant critical aspect of municipalities included in Clusters 1, 2 and 5, and as a significant element of marginalization and decrease in quality of life. Consequently, the implementation of alternative forms of provision of basic services emerges as a central aspect of adaptation strategies. For instance, the Italian National Strategy for Internal Areas (SNAI) prefigures a reorganization of the healthcare system based on a decentralized infrastructure of local facilities, including pharmacies, general practitioners, and alternative forms of service provision, including health homes, mobile health services, integrated home care.

Moreover, the SNAI prefigures the improvement of telemedicine services. Concerning transport services, the modest transport demand, locational disadvantages, and dispersed urbanization of internal rural areas, determine the need for the development of a multi-modal transport model that integrates public transit [19], active mobility and flexible and semi-flexible demand responsive shared transport services [20]. These solutions can include car-sharing and split car-sharing services, carpooling solutions based on autonomous driving vehicles [21], ride-hailing and shared ride-hailing [22]. A relevant aspect, central to the implementation of mitigation and adaptation strategies, concerns the improvement of the internet infrastructures. In clusters 1, 2 and 5 areas, in fact, the average proportion of households served by an internet infrastructure providing speeds ranging from 0 to 30 Mbps is equal, respectively, to 0.836, 0.884 and 0.953, and the maximum average download speed is equal to, respectively, 24.722 Mbps, 20.459 Mbps and 12.624 Mbps. Mitigation strategies, and in general terms actions aimed at promoting economic development, are frequently related to the valorization of sustainable forms of

tourism, including eco-tourism and smart tourism, and to the incentivization of agricultural practices. In this perspective, the development of an agricultural production model based on multifunctional farms emerges as a relevant policy. Multifunctional farms are recognized as fundamental elements for the conservation and valorisation of the local capital and of traditional, sustainable forms of production [23–26]. The emergence of multifunctional entrepreneurial strategies in the primary production sector is considered a form of resilience and resistance to the crisis in the agriculture industry, and a valuable element for the conservation and maintenance of environmental and socio-cultural landscape values. The adoption of multifunctional strategies results in the diversification of agricultural production, particularly based on traditional crops and on conservation, organic, and biological farming practices. It aims at integrating agricultural production and complementary functions, including transformation and distribution of products, catering, accommodation, environmental education, nature experience, cultural services and maintenance of the rural landscape components. The notion of multifunctionality underlines the diverse potential functions of the agricultural sector in contemporary society, and in particular in marginalized areas: maintaining the vitality and adequate conditions for socio-economic development of rural areas, ensuring food security for the population, diversifying food production, and preserving the landscape. Moreover, the concept of multifunctionality in agricultural practices is related to the concept of sustainable tourism. Multifunctional farms, in fact, emerge as tourism experience providers in the context of alternative forms of sustainable tourism, including eco-tourism, geo-tourism and smart tourism. Eco-tourism refers to a form of responsible travel based on sustainable practices of nature experience, and aimed at promoting the local economic and social development, while minimizing the impact on the environment. It focuses primarily on experiencing and understanding the landscape, its flora, fauna and the cultural artifacts from the locality [27], and it aims to foster the involvement of local enterprises and service providers.

Tourism economy development can also be enable by strategies of tourism informatization, and in particular, by smart tourism; Smart tourism, in fact, refers to the integration of digital platforms and tools for improving the tourist's experience - by providing personalized and engaging experiences for tourists - and for enabling the instant integration of data related to tourist activities, product consumption, and tourism and social resources [28, 29]. Final consideration concerns the relevance of selected indicators. In particular, a main limitation to this study regards the limited availability of complete, integrated and open data set. In particular, data on the structure of economic activities, in the manufacturing and tourism sectors is provided at NUTS-3 regions level, thus not providing useful information for discerning social and economic trends specific to compact service poles, intermediate urbanized areas, areas of dispersed urbanization and marginal rural areas. Moreover, variations in the definition of categories, or in the procedures for the calculation of indicators, prevent the comparison of data related to distinct periods and, consequently, the reconstruction of trends at the municipal level. For instance, due to a revised definition of active enterprise, the data related to the Number of employees of local units of active enterprises calculated from the year 2019, are not comparable to data related to previous years. Lastly, the number of employees itself is defined as a continuous variable in surveys from 2012, and as a discrete variable in 2001 and 2011

census data. Moreover, privacy regulations require that data related to the structure of agricultural production at the municipal scale – including Net labor productivity, Net productivity of land, Incidence of public support, Net labor profitability, Net value added of labor and Net value added of land - are not available for municipalities where less than 5 enterprises are located. Similar criteria limit the availability of data relevant to the understanding of patterns of educational poverty: data, aggregated at the municipal level, related to the results of the Invalsi test, for the measurement of students' competences, are available for municipalities where two or more schools are located. As a consequence, the future development of this study will focus on the definition of metrics for understanding the socio-economic tendencies at the local scale, and for qualifying the clusters of declining rural areas in terms of specific social and economic conditions [30, 31]. The relevance of studies aimed at defining metrics for the analysis of the spatial distribution of social and economic trends and of conditions of access to basic services relies in their contribution in increasing the transparency and impact of public policies. Spatial information, in fact, can facilitate informed decisions in the context of the definition and implementation of cohesion policies [32]. In particular, spatial and statistical analysis techniques can assist public agencies and local governments in identifying distinct forms of complex rural shrinkage, defining adaptation and mitigation objectives, formulating targeted strategies, and measuring the outcomes of public interventions.

Acknowledgements. The study was supported by the Italian National Operational Programme on Research and Innovation ('PON', MIUR, 2014-2020) through the Research Project "MitiGO - Mitigazione dei rischi naturali per la sicurezza e la mobilità nelle aree montane del Mezzogiorno", ARS01_00964; CUP B44I20000300005.

References

1. European Commission, Directorate-General for Communication, Leyen, U.: A union that strives for more: my agenda for Europe: political guidelines for the next European Commission 2019–2024. Publications Office (2019). https://doi.org/10.2775/018127
2. European Committee of the Regions: The impact of demographic change on European regions. Publications Office (2016). https://doi.org/10.2863/26932
3. Grasland, C., et al.: Shrinking regions: a paradigm shift in demography and territorial development. Parlement Européen; Direction Générale des politiques internes de l'Union (2008)
4. Dolton-Thornton, N.: Viewpoint: how should policy respond to land abandonment in Europe? Land Use Policy **102**, 105269 (2021). https://doi.org/10.1016/j.landusepol.2020.105269
5. Copus, A., et al.: ESCAPE. European shrinking rural areas: challenges, actions and perspectives for territorial governance: applied research. Final report. Version. 21, 2020 (2020)
6. Lasanta, T., Arnáez, J., Pascual, N., Ruiz-Flaño, P., Errea, M.P., Lana-Renault, N.: Space–time process and drivers of land abandonment in Europe. CATENA **149**, 810–823 (2017). https://doi.org/10.1016/j.catena.2016.02.024
7. MacDonald, D., et al.: Agricultural abandonment in mountain areas of Europe: environmental consequences and policy response. J. Environ. Manag. **59**, 47–69 (2000). https://doi.org/10.1006/jema.1999.0335

8. Rizzo, A.: Declining, transition and slow rural territories in southern Italy characterizing the intra-rural divides. Eur. Plan. Stud. **24**, 231–253 (2016). https://doi.org/10.1080/09654313.2015.1079588
9. Alamá-Sabater, L., Budí, V., Álvarez-Coque, J.M.G., Roig-Tierno, N.: Using mixed research approaches to understand rural depopulation. Economía Agraria y Recursos Naturales Agric. Resour. Econ. **19**, 99–120 (2019)
10. Reynaud, C., Miccoli, S., Benassi, F., Naccarato, A., Salvati, L.: Unravelling a demographic 'Mosaic': spatial patterns and contextual factors of depopulation in Italian municipalities, 1981–2011. Ecol. Ind. **115**, 106356 (2020). https://doi.org/10.1016/j.ecolind.2020.106356
11. Pašakarnis, G., Morley, D., Malienė, V.: Rural development and challenges establishing sustainable land use in Eastern European countries. Land Use Policy **30**, 703–710 (2013). https://doi.org/10.1016/j.landusepol.2012.05.011
12. Ioffe, G., Nefedova, T., Kirsten, D.B.: Land abandonment in Russia. Eurasian Geogr. Econ. **53**, 527–549 (2012). https://doi.org/10.2747/1539-7216.53.4.527
13. Syssner, J.: Arguments for a local adaptation policy. In: Syssner, J. (ed.) Pathways to Demographic Adaptation: Perspectives on Policy and Planning in Depopulating Areas in Northern Europe, pp. 69–86. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-34046-9_6
14. Dax, T., Fischer, M.: An alternative policy approach to rural development in regions facing population decline. Eur. Plan. Stud. **26**, 297–315 (2018). https://doi.org/10.1080/09654313.2017.1361596
15. Horlings, L.G., Marsden, T.K.: Exploring the 'new rural paradigm' in Europe: eco-economic strategies as a counterforce to the global competitiveness agenda. Eur. Urban Reg. Stud. **21**, 4–20 (2014). https://doi.org/10.1177/0969776412441934
16. Yamu, C., Frankhauser, P.: Spatial accessibility to amenities, natural areas and urban green spaces: using a multiscale, multifractal simulation model for managing urban sprawl. Environ. Plann. B Plann. Des. **42**, 1054–1078 (2015). https://doi.org/10.1068/b130171p
17. Stahle, A., Marcus, L., Karlström, A.: Place syntax tool — GIS software for analysing geographic accessibility with axial lines. Presented at the New Developments in Space Syntax Software, pp. 35–42 (2007)
18. Anselin, L.: A local indicator of multivariate spatial association: extending Geary's c. Geogr. Anal. **51**, 133–150 (2019). https://doi.org/10.1111/gean.12164
19. Harz, J., Sommer, C.: Determinants of success and constraints of integrated ridesharing in rural areas. Transp. Res. Interdiscip. Perspect. (2020)
20. Bruzzone, F., Scorrano, M., Nocera, S.: The combination of e-bike-sharing and demand-responsive transport systems in rural areas: a case study of Velenje. Res. Transp. Bus. Manag. **40**, 100570 (2021). https://doi.org/10.1016/j.rtbm.2020.100570
21. von Mörner, M.: 4 - Demand-oriented mobility solutions for rural areas using autonomous vehicles. In: Coppola, P., Esztergár-Kiss, D. (eds.) Autonomous Vehicles and Future Mobility, pp. 43–56. Elsevier (2019). https://doi.org/10.1016/B978-0-12-817696-2.00004-4
22. Elting, S., Ehmke, J.F.: Potential of shared taxi services in rural areas – a case study. Transp. Res. Procedia **52**, 661–668 (2021). https://doi.org/10.1016/j.trpro.2021.01.079
23. Organisation for Economic Co-operation and Development: Multifunctionality: Towards an Analytical Framework. OECD Publishing, Paris (2001)
24. Organisation for Economic Co-operation and Development: The New Rural Paradigm: Policies and Governance. OECD Publishing, Paris (2006)
25. Belletti, G., Brunori, G., Marescotti, A., Rossi, A.: Individual and Collective Levels in Multifunctional Agriculture. University of Florence, Florence (2002)
26. Meloni, B., Pulina, P. (eds.): Turismo sostenibile e sistemi rurali: multifunzionalità, reti di impresa e percorsi. Rosenberg & Sellier, Torino (2020)
27. Kiper, T.: Role of ecotourism in sustainable development. In: Özyavuz, M. (ed.) Advances in Landscape Architecture, Ch. 31. IntechOpen, Rijeka (2013). https://doi.org/10.5772/55749

28. Annunziata, A., Desogus, G., Garau, C.: Smart tourism governance for urban bioregion: an evaluating approach to the relationship between coastal and inland areas of South Sardinia. In: Building the Urban Bioregion Governance Scenarios for Urban and Territorial Planning, p. 290. SdT Edizioni, Firenze, Italy (2022)
29. Wang, D., Wang, F.: Contributions of the usage and affective experience of the residential environment to residential satisfaction. Null. **31**, 42–60 (2016). https://doi.org/10.1080/02673037.2015.1025372
30. Corrado, S., Scorza, F.: Machine learning based approach to assess territorial marginality. In: Gervasi, O., Murgante, B., Hendrix, E.M.T., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2022. LNCS, vol. 13376, pp. 292–302. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-10450-3_25
31. Santarsiero, V., Nolè, G., Scorza, F., Murgante, B.: Geographic information and socio-economic indicators: a reading of recent territorial processes in the test area of Basilicata region. In: Calabrò, F., Della Spina, L., Piñeira Mantiñán, M.J. (eds.) NMP 2022. LNNS, vol. 482, pp. 2104–2111. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-06825-6_202
32. Murgante, B., Borruso, G., Lapucci, A.: Geocomputation and urban planning. In: Murgante, B., Borruso, G., Lapucci, A. (eds.) Geocomputation and Urban Planning. SCI vol. 176, pp. 1–17. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-89930-3_1