TeMA

Journal of
Land Use, Mobility and Environment

This Special Issue of TeMA - Journal of Land Use, Mobility and Environment, collects twenty-seven contributes of international researchers and technicians in form of scenarios, insights, reasoning and research on the relations between the City and the impacts of Covid-19 pandemic, questioning about the development of a new vision and a general rethinking of the structure and urban organization.

TeMA Journal offers papers with a unified approach to planning, mobility and environmental sustainability. With ANVUR resolution of April 2020, TeMA journal and the articles published from 2016 are included in the A category of scientific journals. From 2015, the articles published on TeMA are included in the Core Collection of Web of Science. It is included in Sparc Europe Seal of Open Access Journals, and the Directory of Open Access Journals.

Special Issue

Covid -19 vs City -20

scenarios, insights, reasoning and research

ISSN 1970-9889
University of Naples Federico II

TeMA

Journal of
Land Use, Mobility and Environment

Special Issue

COVID-19 vs CITY-20

SCENARIOS, INSIGHTS, REASONING AND RESEARCH

Published by
Laboratory of Land Use Mobility and Environment
DICEA - Department of Civil, Architectural and Environmental Engineering
University of Naples "Federico II"

TeMA is realized by CAB - Center for Libraries at “Federico II” University of Naples using Open Journal System

Editor-in-chief: Rocco Papa
print ISSN 1970-9889 | on line ISSN 1970-9870
Licence: Cancelleria del Tribunale di Napoli, n° 6 of 29/01/2008

Editorial correspondence
Laboratory of Land Use Mobility and Environment
DICEA - Department of Civil, Architectural and Environmental Engineering
University of Naples "Federico II"
Piazzale Tecchio, 80
80125 Naples
web: www.tema.unina.it
e-mail: redazione.tema@unina.it

Given the short time to produce the volume, the Editorial Board of TeMA Journal carried out the scientific quality audit of the contributions published in this Special Issue.

The cover image is a photo collage of some cities during the Covid-19 pandemic quarantine (March 2020)

TeMA Journal of Land Use, Mobility and Environment offers researches, applications and contributions with a unified approach to planning and mobility and publishes original inter-disciplinary papers on the interaction of land use, mobility and environment. Domains include: engineering, planning, modeling, behavior, economics, geography, regional science, sociology, architecture and design, network science and complex systems.

With ANVUR resolution of April 2020, TeMA Journal and the articles published from 2016 are included in A category of scientific journals. From 2015, the articles published on TeMA are included in the Core Collection of Web of Science. TeMA Journal has also received the *Sparc Europe Seal* for Open Access Journals released by *Scholarly Publishing and Academic Resources Coalition* (SPARC Europe) and the *Directory of Open Access Journals* (DOAJ). TeMA is published under a Creative Commons Attribution 3.0 License and is blind peer reviewed at least by two referees selected among high-profile scientists. TeMA has been published since 2007 and is indexed in the main bibliographical databases and it is present in the catalogues of hundreds of academic and research libraries worldwide.

TeMA

Journal of
Land Use, Mobility and Environment

Special Issue

COVID-19 vs CITY-20
SCENARIOS, INSIGHTS, REASONING AND RESEARCH

Contenets

TeMA Journal of
Land Use, Mobility and Environment

TeMA Special Issue | Covid-19 vs City-20, 189-212
print ISSN 1970-9889, e-ISSN 1970-9870
DOI: 10.6092/1970-9870/6864
Received 15th May 2020, Available online 19th June 2020

The role of the urban settlement system in the spread of Covid-19 pandemic. The Italian case

C. Gargiulo [a], F. Gaglione [b], C. Guida [b], R. Papa [c], F. Zucaro [b], G. Carpentieri [d]*

[a] Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Italy
e-mail: gargiulo@unina.it
ORCID: https://orcid.org/0000-0001-6481-0908

[b] Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Italy
e-mail: federica.gaglione@unina.it, carmen.guida@unina.it, floriana.zucaro@unina.it

[c] Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Italy
e-mail: rpapa@unina.it
ORCID: https://orcid.org/0000-0003-3355-1418

[d] Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Italy
e-mail: gerardo.carpentieri@unina.it
ORCID: https://orcid.org/0000-0002-2111-650X
* Corresponding author

Abstract
The paper proposes a focus on three main aspects related to the spread of the new coronavirus in our country: the correlations that have been established between the spread of the Covid-19 virus and the settlement system of our country; the urban and territorial phenomena that can be associated, positively or negatively, with the diffusion of the virus; and, finally, the correspondence between homogeneous clusters of Italian provinces (due to the current most significant urban phenomena) and the intensity and spread of the infection. The research is divided in four steps: the identification of the scientific and disciplinary approach, the definition of territorial areas and their descriptive variables, the choice of computational models, and the evaluation of the results. The main findings of the study highlight that significant correlations are not always identifiable between settlement characteristics and the spread of the infection. The diffusion of the new coronavirus is closely related to some of the main features of the demographic (e.g. people aged 65 years and above) and socio-economic (e.g. GDP for inhabitant) structure of the urban population.

Keywords
Holistic approach, Settlement system, Spread of Covid-19 virus, Statistic analysis.

How to cite item in APA format
Gargiulo, C., Gaglione, F., Guida, C., Papa, R., Zucaro, F. & Carpentieri, G. (2020). The role of the urban settlement system in the spread of Covid-19 pandemic. The Italian case. *Tema. Journal of Land Use, Mobility and Environment*, 189-212. http://dx.doi.org/10.6092/1970-9870/6864

1. Introduction

"La matematica è lo strumento indispensabile per capire quanto sta accadendo e scrollarsi di dosso le suggestioni. Le epidemie prima di emergenze mediche sono emergenze matematiche perché la matematica non è la scienza dei numeri, è la scienza delle relazioni, descrive i legami e le relazioni tra enti diversi cercando di dimenticare di cosa sono fatti quegli enti astraendoli in lettere, funzioni, vettori, punti e superfici. Il contagio è una infezione della nostra rete di relazioni" (Giordano, 2020).

There are essentially three scientific questions that this work intends to answer. The first concerns the possible correlations that can be established between the intensity and spread (as at 31 May 2020) of Covid-19 infection and the settlement system of our country with reference to the provincial areas and, limited to the available variables, also to the capital cities. The second question relates to the identification of urban and territorial phenomena that may be associated, be it positively or negatively, with the spread of infection. The third question concerns the update (2020) of the main phenomena of settlement nature (*inter alia*) that can be recognised within the system of Italian provinces in order to identify, with reference to the intensity and spread of infection, possible and different correspondences with homogeneous clusters of provinces. This work is divided into six phases: the identification of the scientific-disciplinary paradigm in which the research is developed, the definition of the territorial areas, the phenomena examined and the set of descriptive variables, the identification and implementation of calculation tools, the definition of the analysis models used, the results of the elaborations carried out, the conclusions.

2. The scientific paradigm

The prominent feature of a chaotic system is its high sensitivity to even the smallest actions that can occur at any point of its being and becoming. Therefore, the degree of indeterminacy that can reach a chaotic system is extremely high and, in addition, any phenomenon, even a small phenomenon, can very quickly reach macroscopic proportions. In other words, in the presence of chaos, any prediction can reach relevant inaccuracies (Gargiulo & Papa, 1992).

The theory of dynamic systems can be considered as the essential support in the development of the conceptual framework of reference for the study of "chaos". The definition of a dynamic system is given by the elements and relations between the elements of the system, as well as the laws and criteria of (state) evolution over time (Gargiulo, 2009).

The space of existence of the evolution of the system is known as the space of states or space of phases; this space is a purely conceptual abstraction, the coordinates of which are the components of the state. Of course, the spatial coordinates of the phases change with the context. Although it is recognised that the behaviour of chaotic dynamic systems is unpredictable, the space of states can be useful to represent this behaviour in geometric form.

In order to refer back to our cognitive/interpretative capacity the complex systems and, with them, the city and the territory, these can be broken down into subsystems that, with less difficulty, enable us to find the structure of the system and, therefore, the set of relations that are established between the elements of each subsystem and between these and other subsystems: the organisation (Morin, 1993).

The organisation represents the constituent property of a system and the types of relationships/interactions that are established between the elements of the system determine its functioning.

In order to understand the functioning of a system, it is necessary to adopt a holistic approach and to know it as a whole, in its entirety. Knowledge, as the result of the sum of the behaviours of the individual parts, does not restore the overall behaviour of the system. As Morin states "the organisation connects, in an interrelational way, different elements, or events, or individuals that consequently become components of a whole" (Morin, 1983).

It is a widely shared opinion that the holistic approach, together with the General Systems Theory and the complexity paradigm, offers a valid theoretical support useful to the knowledge and interpretation of the territory and the city, especially during historical times of "bifurcation" such as that we are currently experiencing (Bottero et al., 2017).

In fact, in addition to the speed of transformation processes, the coexistence of apparently contradictory phenomena, the growing spread of the products of technology over the city and throughout the territory, unpredictable catastrophic events occur, such as the Covid-19 (Fang & Wahba, 2020), a pandemic that has exploded in Italy in all its virulence since February 2020, forcing to change our social behaviour and, with it, many relationships that structure the urban and territorial system.

The work presented in this article, which aims to understand if and which characteristics of the territory have contributed to the spread of the Coronavirus, uses the holistic-systemic approach, summarised so far, seeking the relationships/interactions between the numerous elements that come into play in a complex system such as the territorial and urban system.

To this end, the reading and interpretation of the data could only refer to statistical multivariate techniques that allow, through the use of a significant amount of data, a "synthetic" reading, i.e., the reading of the mutual influences of a complex system.

Specifically, in the first part of the work, the objective is specified in the study of the relationships between the descriptive characteristics of the various subsystems (anthropic, functional, economic, geo-environmental and mobility) through multivariate linear regression techniques. The choice of scale, for this reason, was the provincial scale, with some specifications to the urban scale of the provincial capital, to also allow for the reading of the influence of the hegemonic city in the provincial territory.

In the second part of the work, through the method of the principal components, the primary attributes that define and condition the national settlement system on the provincial scale are studied.

In a nutshell, the objective of this analysis is to understand whether the territorial characteristics significantly correlated with the spread of the pandemic play a relevant role in the provincial spatial planning, as appears from the application of factorial analysis and cluster analysis.

3. The definition of variables

3.1 The dependent variable: the extent of infection

In order to identify the dependent variable representative of the distribution and spread of the pandemic in the individual provinces of the national territory, the data provided by the government agencies involved in the management and monitoring of the Covid-19 crisis were used.

Amongst' the data available in open access format, those relating to the number of total infections and deaths due to Covid-19, declared by the individual regional agencies, during the first four months of 2020 have been selected. The numerical values related to infections were obtained from the data repository associated with the WebGIS platform developed by the Italian Civil Protection Department for the Covid-19 crisis on 31 May. The values of deaths due to Covid-19 were taken from the report *Impact of the Covid-19 epidemic on the total mortality of the resident population in the first four months of 2020* prepared by ISTAT (Italian National Institute of Statistics) and ISS (Italian National Institute of Health) and updated on 30 April.

At this stage, a hypothesis was made to develop a synthetic measure of infection that would jointly take into account the number of positive cases and the number of swabs carried out, excluding the number of swabs carried out on individuals already declared positive for the virus to check their recovery. However, the different reliability of the variables involved suggested choosing only one descriptive variable of the infection, whilst being aware of its poor reliability.

3.2 The independent variables: the settlement system

In order to investigate whether and to what extent the characteristics of urban planning have influenced the distribution and spread of infection in Italy, a set of 28 continuous quantitative variables has been defined to describe the "structure" of the system of the Italian provinces, i.e., the main elements and relationships (which can be interpreted through the subsequent use of multivariate statistics techniques) that comprise and determine the organisation and functioning of this territorial level of reference (Masoumi & Shaygan, 2016).

The variables chosen are representative of the main 5 sub-systems that characterise the settlement system and reflect the holistic approach adopted:

- anthropogenic sub-system, consisting of 7 variables related to the demographic structure of the settled population, such as number of inhabitants and settlement density of both provinces and capitals, population rate over the age of 65, average age and natural balance (Gargiulo et al., 2018);
- socio-economic sub-system, consisting of 8 variables related to the economic-productive structure of the settled population, such as elderly dependency ratio, active population, number of unemployed, ratio between the employed and active population, income and GDP per capita, average income per provincial capital;
- functional sub-system, consisting of 2 variables related to the supply of some of the main health and training services in the provincial territories, such as universities and hospital beds;
- mobility sub-system, consisting of 6 variables relating both to mobility supply, in terms of vehicle fleet, public transport fleet (buses and trolley buses) and motorcycle fleet, as well as transport demand, taking into account the number of total daily movements and the number of movements for study and work purposes;
- geo-environmental sub-system, consisting of 5 variables related to the territory over which all the activities, spaces and channels of the provinces are distributed, such as territorial surface area, portion of mountain territorial surface area, average temperature value for the months of March, April and May of the thirty-year period 1981-2010 , distance of the provincial capital of each province from the provincial capital of the area infection by region, distance from the provincial capital of the area of infection (Bergamo) to the capital of each province (Zucaro & Morosini, 2018).

Given that part of the data provided by ISTAT refers to 2011 (the year of the population census), the provinces of Verbano, Forlì-Cesena and Southern Sardinia were excluded from the study, as they were established after that year.

Specifically, as regards the functional sub-system, the variable describing the capacity of the health system has been found from a dataset relating to hospital facilities throughout Europe, the source of which is the European Union Institute of Statistics (EUROSTAT). Being a geo-spatial dataset, the independent variable related to the capacity of health services throughout Europe was obtained from a spatial join operation, performed in terms of GIS, with the shapefile related to the administrative borders of the Italian provinces.

From the reading of the settlement, geographical and territorial information collected, it is possible to record the gap - and not only the economic gap - between the northern provinces and the southern provinces of Italy: Milan, Bolzano, Bologna are at the top of the ranking by GDP per capita, with values even above the European average; at the bottom of the ranking are the provinces of Agrigento, Bari and Cosenza. Provinces such as Biella, Savona and Trieste are "amongst the oldest", unlike Naples, Caserta, where the old age index has the lowest values.

With reference to the functional and mobility sub-systems, the reading of the collected data shows a state of affairs that is not influenced by geographical location, but rather by the metropolitan dimension. Specifically, a greater number of daily movements characterises the Metropolitan Areas (Milan, Naples, Turin, Bari, etc.) compared with the other provincial territories, associated with the greater size of the vehicle fleet (private and local public transport network).

ID	Variable	Source	Year
P01	Covid-19 infections	Civil Protection	2020 (31th May)
P02	Deaths due to Covid-19	ISTAT & ISS	2020 (30th April)
V01	Territorial surface	ISTAT	2020
V02	Territorial surface of the capital	ISTAT	2019
V03	Mountain areas	ISTAT	2011
V04	Resident Population	ISTAT	2019
V05	Resident Population of the capital	ISTAT	2019
V06	Housing density	ISTAT	2019
V07	provincial capital housing density	ISTAT	2019
V08	Average age	ISTAT	2020
V09	Over-65 population	ISTAT	2019
V10	Natural balance	ISTAT	2018
V11	Elderly dependency ratio	ISTAT	2019
V12	Active population per 1,000 inhabitants	ISTAT	2011
V13	Unemployed per 1,000 inhabitants	ISTAT	2011
V14	Average income per capita	ISTAT	2014
V15	Average income per provincial capital	Il Sole 24 Ore	2016
V16	GDP per inhabitant	ISTAT	2016
V17	University locations per 100,000 inhabitants	CRUI	2017
V20	Average temperature value March 1981-2010	ISPRA	2014
V21	Average temperature value April 1981-2010	ISPRA	2014
V22	Average temperature value May 1981-2010	ISPRA	2014
V24	Distance of the provincial capital from the province of first infection	Google Maps	2020
V25	car fleet	MIT	2018
V26	bus and trolley bus fleet	MIT	2018
V27	motorcycle fleet	MIT	2018
V28	Total daily movements	ISTAT	2011
V29	Daily movements for study reasons	ISTAT	2011
V30	Daily movements for work reasons	ISTAT	2011
V31	Hospital beds	Eurostat	2019

Tab.1 The selected dependent and independent variables

3.3 The correlation analysis between variables

After collecting the set of 28 variables, a correlation analysis was carried out to exclude the highly correlated variables and to therefore ensure the significance of the results of the subsequent multivariate statistics analysis. The collinear variables, in fact, do not provide additional information and it is difficult to identify the effect that each of them has on the dependent variable

Table 2 shows the high linear correlation between the variables highlighted in red such as: the resident population, the over-65 population, the vehicle fleet, the total daily movements and the hospital beds. These variables are mainly related to the sub-systems of both functional and anthropic mobility, given that, as it was easy to guess, the resident population is highly correlated with the demand for mobility and the supply of health services. The number of inhabitants is also positively correlated with some of the demographic variables considered, such as the over-65 population and population density.

However, it was decided to keep some variables even if they are highly correlated with others, such as beds and the over-65 population in order to allow for a more immediate interpretation of the final outputs.

Ultimately, 7 variables have been excluded from the initial dataset, which, in terms of significance, do not provide additional elements to the interpretation of the phenomenon under study.

	Resident Population	car fleet	bus and trolley bus fleet	motorcycle fleet	Total daily movements	Daily movements for study purposes	Daily movements for work reasons	Hospital beds
Resident Population	1	0.991	0.930	0.961	0.993	0.993	0.973	0.984
car fleet	0.991	1	0.929	0.954	0.987	0.983	0.970	0.978
bus and trolley bus fleet	0.930	0.929	1	0.890	0.903	0.934	0.869	0.914
motorcycle fleet	0.961	0.954	0.89	1	0.954	0.948	0.938	0.955
Total daily movements	0.993	0.987	0.903	0.954	1	0.976	0.993	0.987
Daily movements for study reasons	0.993	0.983	0.934	0.948	0.976	1	0.944	0.963
Daily movements for work reasons	0.973	0.97	0.869	0.938	0.993	0.944	1	0.981
Hospital beds	0.984	0.978	0.914	0.955	0.987	0.963	0.981	1

Tab.2 Pearson's correlation matrix only with the excluded variables correlation values

3.4 GIS techniques for data verification and control

The tool used for the control and spatial representation of the data and results obtained from this work was the ArcGis computer software. The first part of the research work was dedicated to the collection and management of spatially referenced data, in order to have a cognitive overview of the urban phenomenon (Stillwell & Clarke, 2004). More specifically, a geodatabase divided into five sub-systems (anthropic; socio-economic; functional; mobility and geo-environmental) has been developed on a the provincial scale in relation to the Covid-19 infection index, with the aim of examining the spatial relationships between the characteristics of the urban system and the spread of Covid-19 infection throughout Italy during the cognitive process.

In the second part of the research work, the ArcGIS software has been a tool to support the use of statistical techniques used in order to represent and verify the statistical results obtained, above all, to identify the homogeneous characteristics of the provinces (cluster analysis). Although the use of statistical techniques and the development of GIS tools represent two distinct areas of research, they can both benefit from their joint use (Fotheringham & Rogerson 1993). Their real and functional integration currently represents a frontier of research (Mandal & Mondal,2019) for the definition and analysis of the complexity of urban phenomena, in this case, for the relations between the components of the urban system and the measurement of infection (Hou et al., 2017).

4. Analysis models

4.1 Multiple linear regression

Multiple linear regression is a statistical analysis technique used to make a prediction of the evolution of a dependent variable by taking into account several independent variables (Johnson & Wichern, 1998). The objective is to identify the equation underlying the multiple regression model that can numerically define the linear relationships between the dependent variable and the independent variables considered.

$$Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_n X_{ni} \quad (1)$$

In the event of n independent variables, the multiple regression model assumes the expression reported in equation 1, where β_o is the intercept with the Y_i axis , β_n is the inclination of Y_i with respect to the X_{ni} variable keeping the other independent variables constant.

The IBM SPSS Statistic Version 20 software was used to perform the multiple linear regression analysis. Specifically, the Best-Subsets approach has been used for this application, which enables all possible regression models to be assessed given a number n of independent variables. It has been chosen, as the literature suggests, to select, as the criterion for selecting the best sub-models, the R^2 determination index, the value of which is corrected taking into account the number of independent variables inserted in the model and also considers the width of the data sample (Maxwell, 1981; Muirhead, 1982).

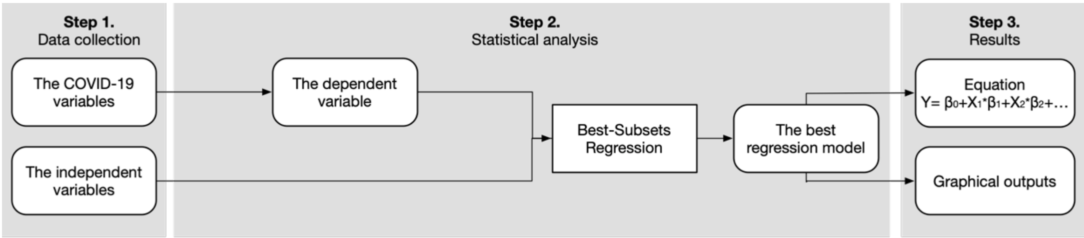

Fig.1 The conceptual procedure for the regression analysis

The use of this specific approach enables the calculation limits of the stepwise approach to be overcome which, in order to identify the best model, makes a limited number of attempts without considering all possible combinations of models. Before proceeding with the regression, the 22 variables used (1 dependent and 21 independent) were normalised using the maximum and minimum values of each variable in order to prevent the different dimensional quantities of the variables influencing the result of the analysis.

Results of Multiple linear regression

The selection of the independent variables to be included in the elaboration aimed at identifying any significant relationships between independent variables (predictors) and dependent variables in one with the choice of the regression model to be applied are the most important elements in the approach to the elaboration of multiple linear regression. In this case, the selection of the independent variables to be included in the model was made on the basis of the initial set, consisting of 28 variables, subsequently reduced to 21, in order to take into account the strong correlation between some of the most significant variables, a correlation that could have presented problems of multicollinearity, therefore reducing the significance of the model used and especially the results of statistical processing.

The dataset used for the analysis of the multiple linear regression comprises a dependent variable (Covid-19 counts) and the 21 independent variables measured for the 104 provinces under analysis (the provinces of Forlì-Cesena, Verbano-Cusio-Ossola and South Sardinia were excluded due to lack of data).

The application of the Best-Subsets approach to this dataset enabled value of R^2 to be obtained, adapted equal to 0.562. Therefore, the best model identified is able to explain more than 56% of the analysed phenomenon.

ID		β	Significance	Importance
	Intercept	0.632	0.020	
V31	Hospital beds	0.707	0.003	0.268
V08	Average age	-0.186	0.023	0.159
V24	Distance of the provincial capital from the province of first infection	-0.164	0.039	0.129
V10	Natural balance	-0.444	0.042	0.126
V22	Average temperature value May 1981-2010	-0.189	0.065	0.104
V01	Territorial surface	-0.093	0.086	0.089
V16	GDP per inhabitant	0.188	0.086	0.089
V13	Unemployed per 1,000 inhabitants	0.082	0.273	0.036

Tab.3 Statistical significance of predictors

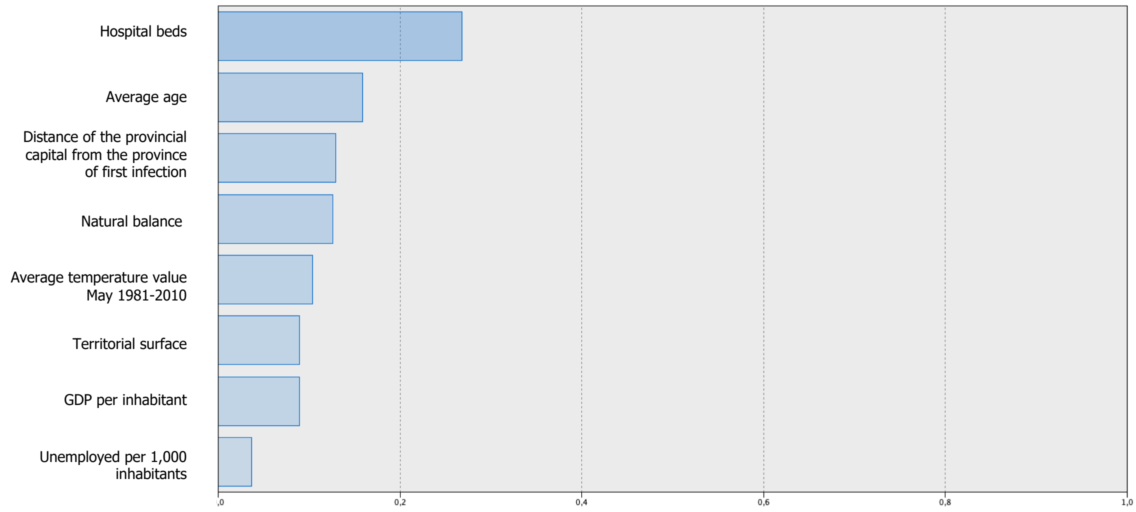

Fig.2 Importance/significance of each predictor

Table 1 shows the list of the eight independent variables selected from the Best-Subsets approach and their values of β_ which are the angular coefficients of the regression model equation. Looking at the values of significance, the eighth variable (Unemployed per 1,000 inhabitants) must be excluded given that the value of significance is greater than 0.1.

The linear model equation estimated from the multiBest-Subsets regression analysis is as follows:

$$Infections=0.632+0.707*V31-0.186*V08-0.164*V24-0.404*V10-0.189*V22-0.093*V01+0.188*V16 \quad (2)$$

It should be noted that the coefficients in a multiple regression model must be considered as net regression coefficients: they measure the variation of the response variable Y in correspondence to the variation of one of the explanatory variables when keeping the others constant.

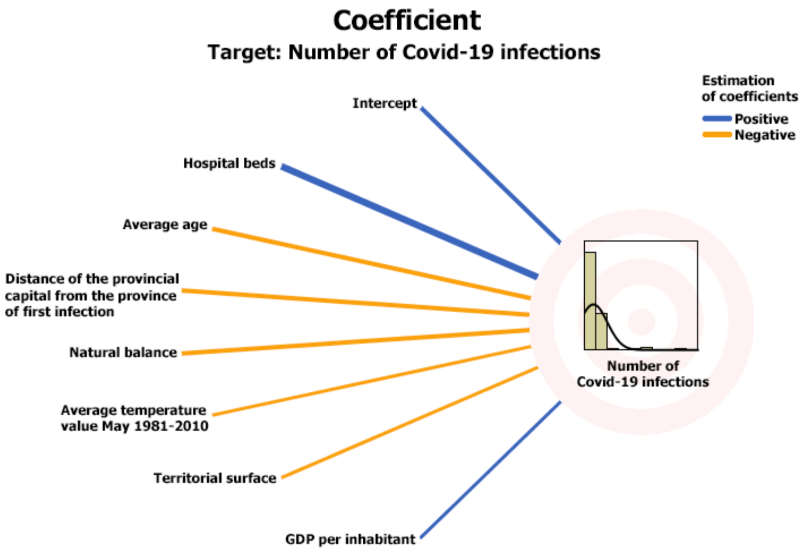

Fig.3 Relationships between the dependent variable and the selected independent variables.

There are numerous outputs of the calculation model used in this study (Best-Subsets) and not all of them are easy to interpret. However, for the objectives of this work, it may be useful to start commenting, in addition to the equation of the estimated linear model shown above, on the synthetic representation below in which the estimation of the coefficients is highlighted with respect to the level of decreasing significance of each one, the sign that assumes the coefficient in the estimation equation and the normalised value of each coefficient.

There are 7 independent variables, selected directly by the model's algorithm, which have certain characteristics - in terms of number, distribution and variability - that can relate to the data related to the dependent variable (measurement of infection) in the Italian provinces as at 31 May 2020, of which two are positively related and 5 are negatively related, with a decreasing level of significance and consequent increasing importance. Unfortunately, none of the variables identified are representative of the characteristics of urban and territorial planning - which is the specific discipline of urban planning in the strict sense. However, other predictors can contribute - in conjunction with the identification of the latent system of the Italian provinces obtained through the analysis of the Main Components and especially with the Cluster Analysis that complete this study - to identify phenomena of coexistence and collinearity between the consistency and distribution of infection on a provincial scale and the settlement system of our country. The variable number of hospital beds is the most important predictor of the entire panel of variables given that its normalised coefficient (0.770) takes the highest value and its significance (0.003) the lowest value and consequently the highest importance. As expressly mentioned in the section of the study relating to the definition of the correlation between the variables used, this data is strongly correlated with other very significant indicators of the settlement structure of our country - resident population, over-65 population, vehicle fleet, daily movements, etc. - and, therefore, the considerations on this predictor must take into account, so to speak, the baggage of information underlying the beds variable. In other words, the number of beds can be identified as a multiplier of the infection, not only because the hospital facilities have become real multipliers of the infection, in conjunction with the RSA (*Residenza Sanitaria Assistenziale* – Extended Care), but also because they are a measure related to the number of elderly people over the age of 65 (age group particularly exposed to the infection and with fatal consequences), the number of residents and the dynamism of the exchanges of each province. The coefficient of the second and final predictor with a positive sign, but with a decidedly marginal importance, is 0.082 and refers to the GDP per settled inhabitant. In a nutshell, it can be said that as the number of beds increased and, with it, as the number of people aged over 65 and the GDP per inhabitant increased, the infection also increased, naturally due to the different coefficients. All the other predictors have instead operated in the sense of a reduction in the number of infections, obviously with different weightings. The most significant was the natural balance with a normalised (negative) coefficient of 0.444, so it can be said that a higher natural balance helped to reduce the total number of infections. The other predictors, again with a negative sign, have remarkably close normalised coefficients values and, therefore, their role in reducing the infection can be considered to have a similar effect. This specifically concerns the distance of the provincial capital from the area of first infection, the average temperature over the last twenty years, the average age and the surface area of each province. In a nutshell and by way of example, it can be said that the value of the geographical distance from the area of first contamination has contributed to reducing the infection. Therefore, an average value of the highest temperature detected in recent years has reduced the spread of the epidemic and the territorial extension of each province has played a role in mitigating the infectiousness of the virus.

The results of this analysis have suggested increasing the knowledge of the latent structure of our settlement system in order to complete the analysis underlying our study and allowing a comparison, with equal effectiveness, of the intensity and spread of infection and the anthropic, socio-economic, functional, spatial and geo-environmental characteristics of the metropolitan areas and provinces of our country.

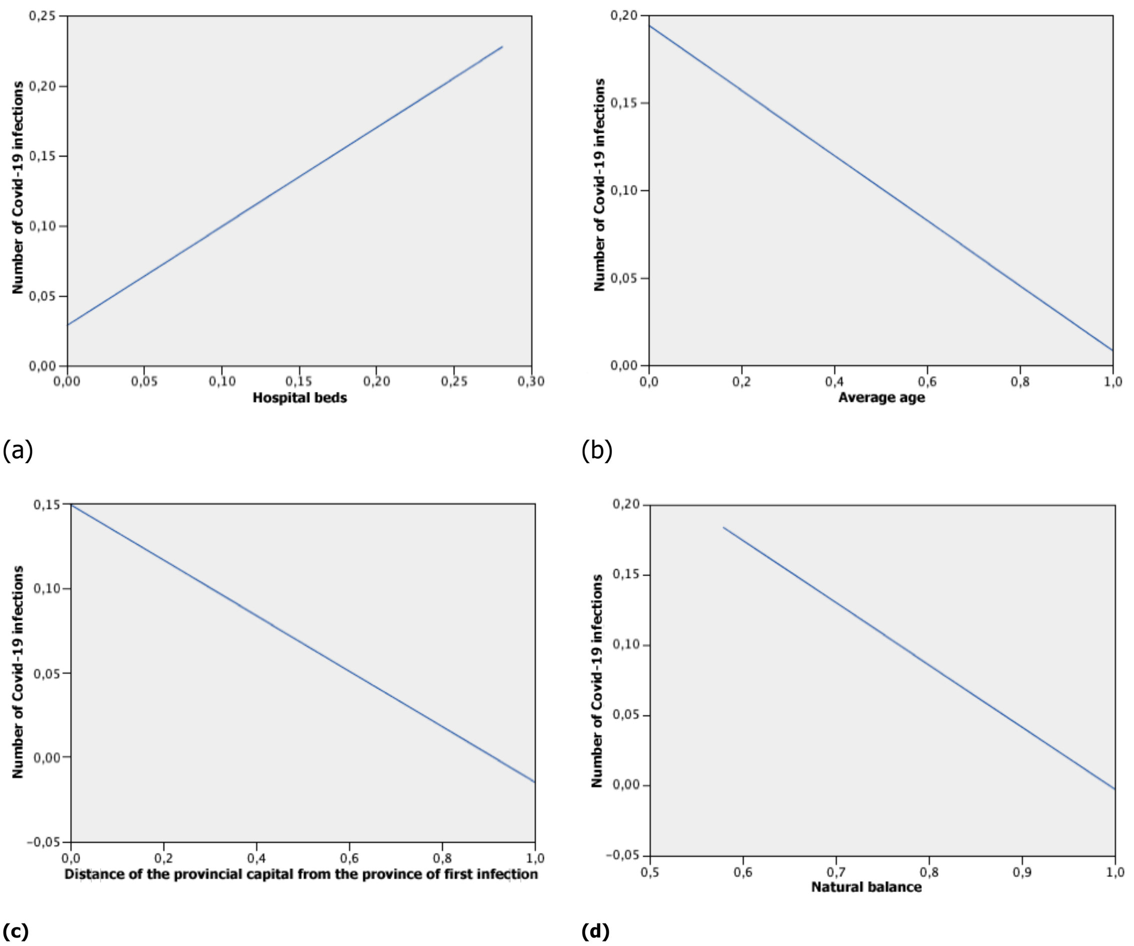

Fig.4 Performance of the significant variables with respect to the dependent variable

4.2 Principal components analysis

The principal components analysis (PCA) provides a description of the variability of a set of variables by means of a subset of new variables, known as main components, which are independent of each other, obtained as linear combinations of the original variables and ordered according to the maximum share of variability (Sadocchi, 1981).

The PCA with reference to p variables, $X_1, X_2, \dots, X_i, \dots, X_n$ with $i = 1, 2, \dots, p$ (random multivariate vector), enables as many p variables (other than the first ones) to be identified, $Y_1, Y_2, \dots, Y_i, \dots, Y_n$ with $i = 1, 2, \dots, p$ (multivariate vector), each linear combination of the p starting variables. The objective of the PCA is to identify appropriate linear transformations Y_i of the observed variables that can be easily interpreted and capable of highlighting and summarising the information inherent in the initial data (Härdle & Simar, 2003; Chakraborty & Reader, 2010).

The application of the PCA to the set of 21 variables selected for this study enables new variables to be obtained that can provide a concise explanation of the variability observed in each of the 21 initial variables. In other words, this method provides a description of the phenomenon under study by means of a small number of unrelated dimensions, ordered according to their importance, the coefficient of which represents the weighting that each variable has in the determination of the component, enabling an easier interpretation (Papa & Piscopo, 1988).
The IBM SPSS Statistic Version 20 software was used to perform the factorial analysis of the main components. The statistical analysis was carried out through the method of factor extraction, setting the threshold value of the eigenvalue, for the choice of the factors, equal to 1. The method chosen for factor rotation (Varimax) maximised the variance of the factor weightings. The outputs, numerical values, matrices and graphs that the software provides have been used in the interpretation phases of the results.

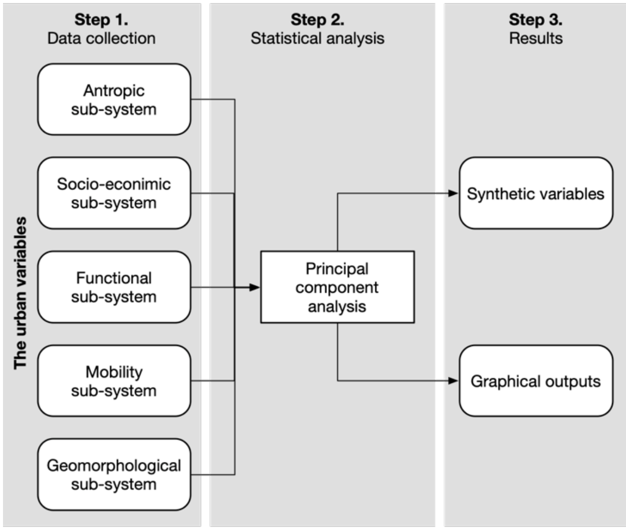

Fig.5 The conceptual procedure for the principal component analysis

For the execution of the analysis of the main components (PCA), data related to the 21 continuous quantitative variables were used as inputs to describe the urban phenomena of the Italian provinces.
The graph shown in Fig. 6 shows that the representation of the eigenvalues for each component has an "elbow" in correspondence of the fifth component that recommended to stop the search for significant components at this value.

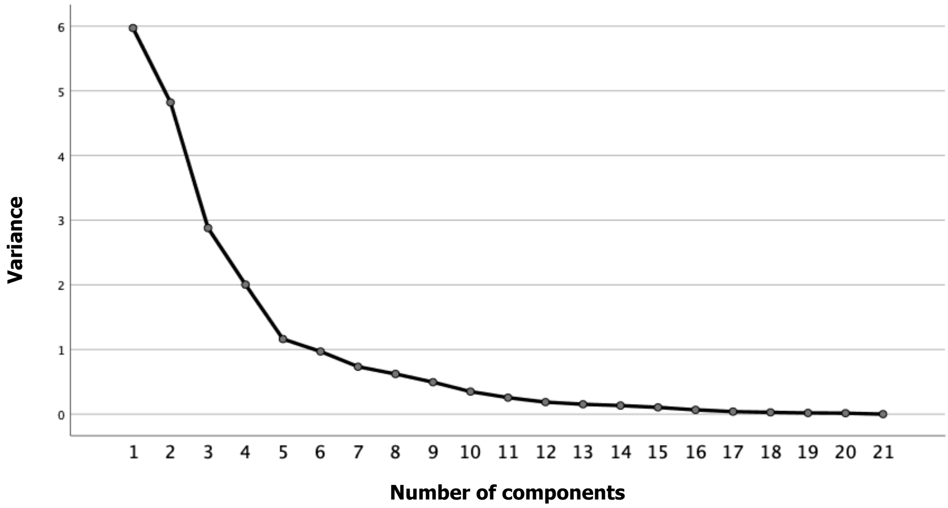

Fig.6 Scree-graph of principal component analysis

The following table shows the percentages of variance of PCA eigenvalues relative to the rotated matrix of the total explained variance. In fact, it should be noted that the distribution of the variables on the components presents some lack of homogeneity as they are grouped around values that are not vastly different from each other. Therefore, it was necessary to consider the sorting of the variables in the rotated factor matrix, which provides a more meaningful sorting of the variables. The first two components "explain" more than 20%, with a cumulative variance of more than 51%.

Factor	Rotation Sums of Squared Loadings		
	Total	**% of variance**	**% cumulated**
1	4.717	22.46	22.46
2	4.180	19.90	42.36
3	3.880	18.47	60.84
4	2.326	11.07	71.91
5	1.730	8.23	80.15

Tab.4 The rotated variance explained

Results of principal components analysis

The multivariate statistical analysis of the principal components, as mentioned above, was conducted from the same variables used for the multiple linear regression. The aim of this part of the work is to identify which characteristics (settlement, anthropic, functional, social, economic, etc.) play a relevant role in the urban and territorial structure of the Italian provinces.

The principal component method provides three outputs: the factorial axes that cut the case cloud along the main axes of inertia, the character coordinates (variables) and object coordinates (104 provinces) on the set of axes defined by the main components.

The first step for the analysis of the outputs consists of the analysis of the eigenvalue vector that enables the percentages of variance explained by each component in relation to the total to be obtained.

In this case, the system is definitely structured as the first 5 components explain 80% of the whole system variance.

The variance explained for each of the first three components assumes a value close to 20% (over 22% for the first, just under 20% for the second and almost 19% for the third) (Tab. 4).

The first component, which explains more than 22% of the variance of the system of variables taken into account, can be defined as an index of the metropolitan city effect. This index is characterised not only by the number of elderly population and the number of hospital beds that occupy the first two places in the ranking, but also by the population living in the capital and the density of settlements in the capital and the province.

It is a component that, albeit not significantly prevailing over the two components that follow it, is nevertheless relevant for the analysis of the territorial system in relation to the Covid-19 crisis, to which reference is made in the following cluster analysis.

The correlation values between the individual variables and the first component show that 15 are positively correlated and 6 are inversely correlated (three of which - active population, distance from the capital of the infection and universities - have values very close to zero and, therefore, not very significant for the interpretation of the index). Overall, there are 8 variables close to zero, with a plus or minus sign (with a value between +0.092 and -0.092).

Variables	Components				
	1	2	3	4	5
Territorial surface	0.230	-0.200	-0.574	0.131	0.557
Territorial surface of the capital	0.319	-0.184	0.215	-0.088	0.713
Mountain areas	0.043	-0.006	-0.904	0.115	0.185
Resident Population of the capital	0.877	0.058	0.079	0.063	0.244
Housing density	0.611	0.249	0.293	0.260	-0.405
Provincial capital housing density	0.770	0.174	-0.063	0.212	-0.417
Average age	-0.154	0.225	0.004	-0.929	-0.002
Over-65 population	0.964	0.105	-0.013	0.123	0.093
Natural balance	-0.773	-0.049	0.040	0.455	-0.214
Elderly dependency ratio	-0.145	0.241	-0.024	-0.931	-0.052
Active population per 1,000 inhabitants	-0.092	0.731	0.044	0.015	0.080
Unemployed per 1,000 inhabitants	0.025	-0.795	0.205	0.162	0.036
Average income per capita	0.209	0.645	-0.203	-0.166	0.047
Average income per provincial capital	0.404	0.758	-0.131	0.055	-0.031
GDP per province inhabitant	0.319	0.821	-0.226	-0.150	-0.001
University locations per 100,000 inhabitants	-0.022	0.165	-0.019	0.062	0.568
Average temperature value March 1981-2010	0.092	-0.396	0.867	0.063	0.114
Average temperature value April 1981-2010	0.082	-0.308	0.905	0.084	0.111
Average temperature value May 1981-2010	0.077	-0.245	0.912	0.098	0.120
Distance of the provincial capital from the province of first infection	-0.065	-0.838	0.181	0.300	0.165
Hospital beds	0.953	0.126	0.030	0.157	0.143

Tab.5 The rotated variance explained for each component

The variables that most characterise the first component refer to the age of the resident population. Specifically, the population over the age of 65 occupies first place in the order of variables, followed by the number of beds in hospital facilities and slightly spaced out from the resident population and population density of the province and the capital. The per capita income of the capital and the GDP of the province also contribute to the positive definition of this component.

The interpretation of the factor also seems to be confirmed by the variables that are on the opposite side of the order and correlated with the minus sign with the first component, such as the natural balance and, at a significant distance, the average age of the population and the old-age dependency ratio.

Three main considerations that derive from the observation of this output can be useful to note that only 30 of the 104 provinces are positively correlated to this component and that, in addition, in the first five places of the order, the 4 most populous metropolitan areas of the country are ranked: Rome, Milan, Turin and Naples. Following these, at a significant distance, is Genoa and, in addition, a group of 5 provinces composed of Palermo, Florence, Bologna, Monza and Bari. The presence of Monza is highlighted, which manages to climb the ranking, likely due to a large number of residents and especially due to the high population density.

The second component, which explains almost 20% of the variance of the system of variables taken into account, can be defined as the inverse index of the development delay, giving more weighting to the variables related to the minus sign.

Amongst the variables positively correlated with the highest values are the variables that describe the economic system: GDP, average income per capital, working population and per capita income per province.

The variable related to the distance from the area of high infection that assumes the highest value amongst all the variables with negative correlation is very important, which, together with the unemployed and the average temperature values, seems to define the southern provinces with high unemployment and average temperature.

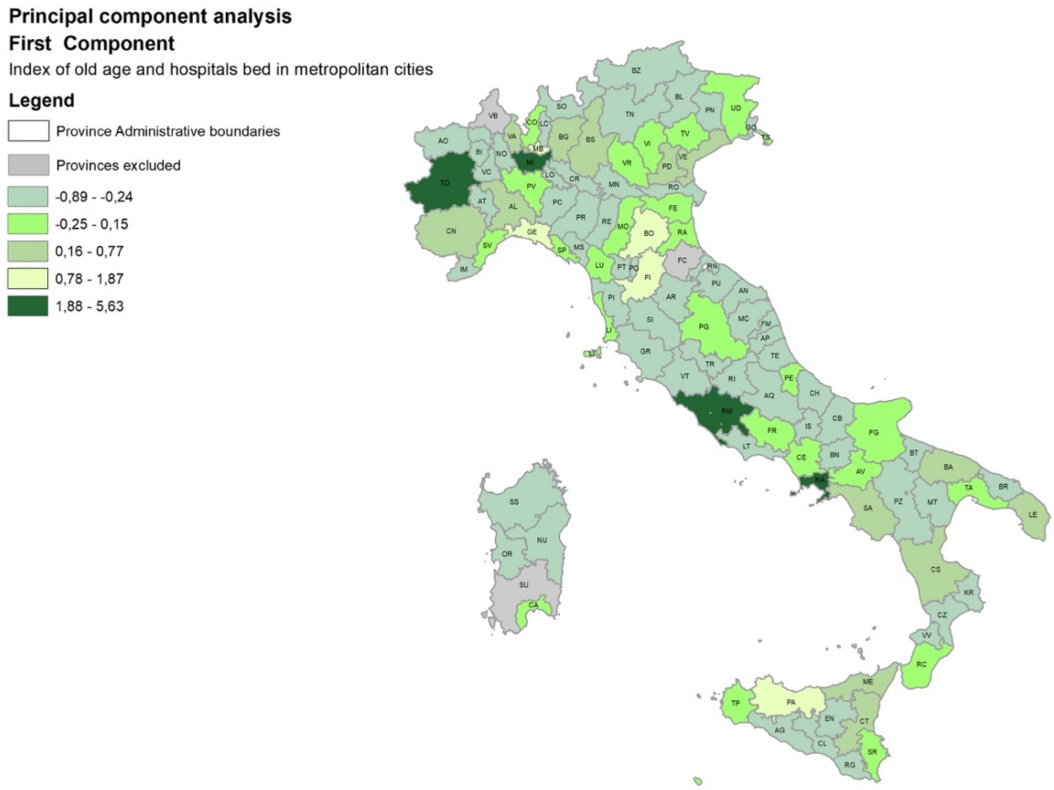

Fig. 7 The values of principal component analysis (PCA) of the first component for the Italian provinces

Specifically, GDP represents gross domestic product, i.e., the sum of all goods and services traded in a country, considered according to their value in euros, as opposed to per capita income, which is the average income of each citizen.

In addition, this component focuses on a variable that is particularly important for the spread of infection: the distance from the capital of the area of maximum infection.

In fact, in the ordering of the variables with respect to the second component, this variable assumes the highest value amongst those negatively correlated.

The correlation ratios between the individual variables and the second component show that, out of the 21 variables, 12 are positively correlated and 9 are inversely correlated; the variables closest to zero and, therefore, insignificant for the interpretation of the index are the eight variables between +0.174 and -0.184, amongst which we find universities, beds, elderly population and land area.

The output relating to the sorting of cases (104 provinces) according to the correlation of each one with the second component enables information to be added that seems to confirm this reading.

In the first place of the order is Milan, followed at a distance by Monza and then by a group of 5 provinces that assume remarkably close values: Parma, Siena, Pisa, Bolzano and Modena. It is worth pointing out that the first province in the South ranks 69 with Caserta.

For the considerations of this territorial system in relation to the Covid-19 crisis, please refer to the following cluster analysis; in fact, this factor finds full correspondence with the first cluster.

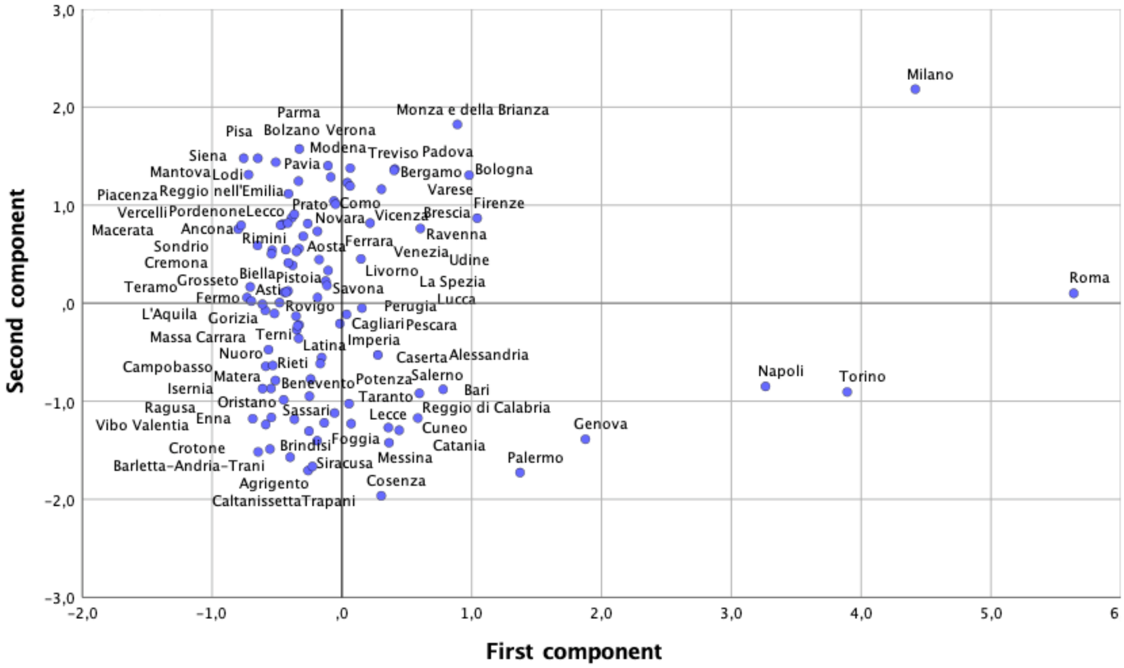

Fig. 8 Biplot graph of the principal component analysis (PCA) between the first and second components

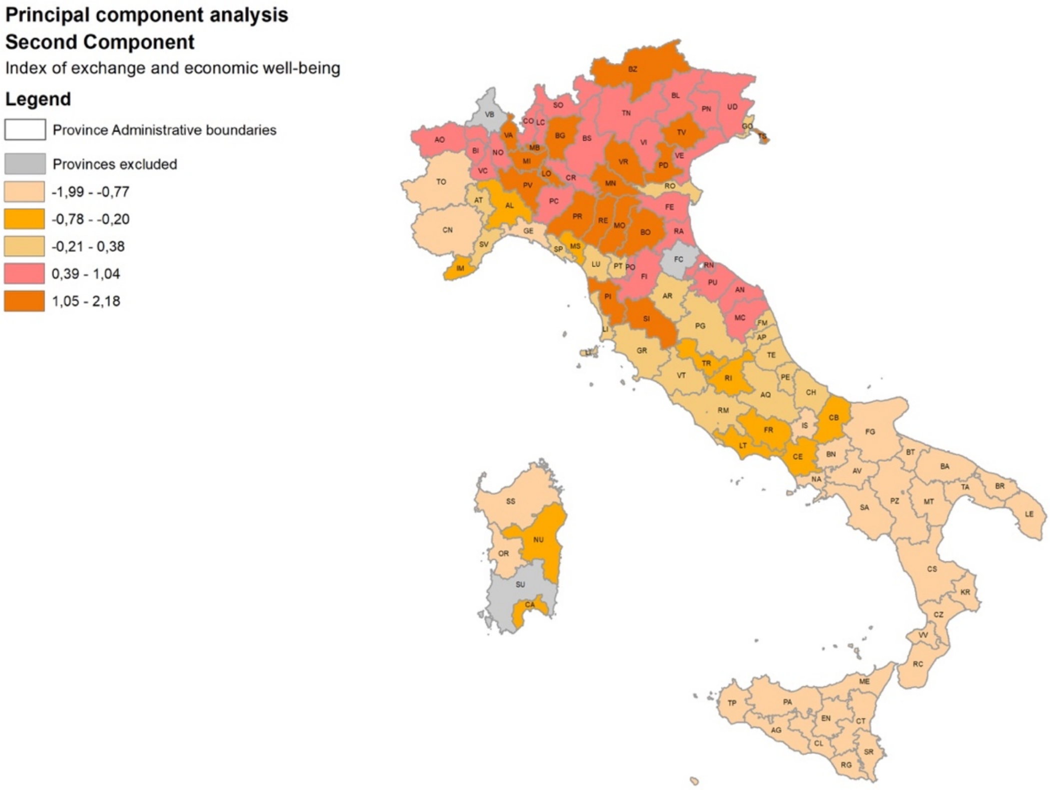

Fig. 9 The values of principal component analysis (PCA) of the second component for the Italian provinces

The third component, which explains almost 19% of the variance of the system of variables taken into consideration, can be defined as the inverse index of the dynamism of the production system.

At the top of the order and with values much higher than the following variables, we find the average temperature data for the months of May, April and March (referring to the years 1981-2010). Confirming the interpretation of the index is the variable related to the mountain area, correlated to the third factor with high values but with a negative sign.

Amongst the positively correlated variables are population density and the territorial surface area of the provincial capital and amongst the negatively correlated variables are the territorial surface area of the province and the GDP per inhabitant.

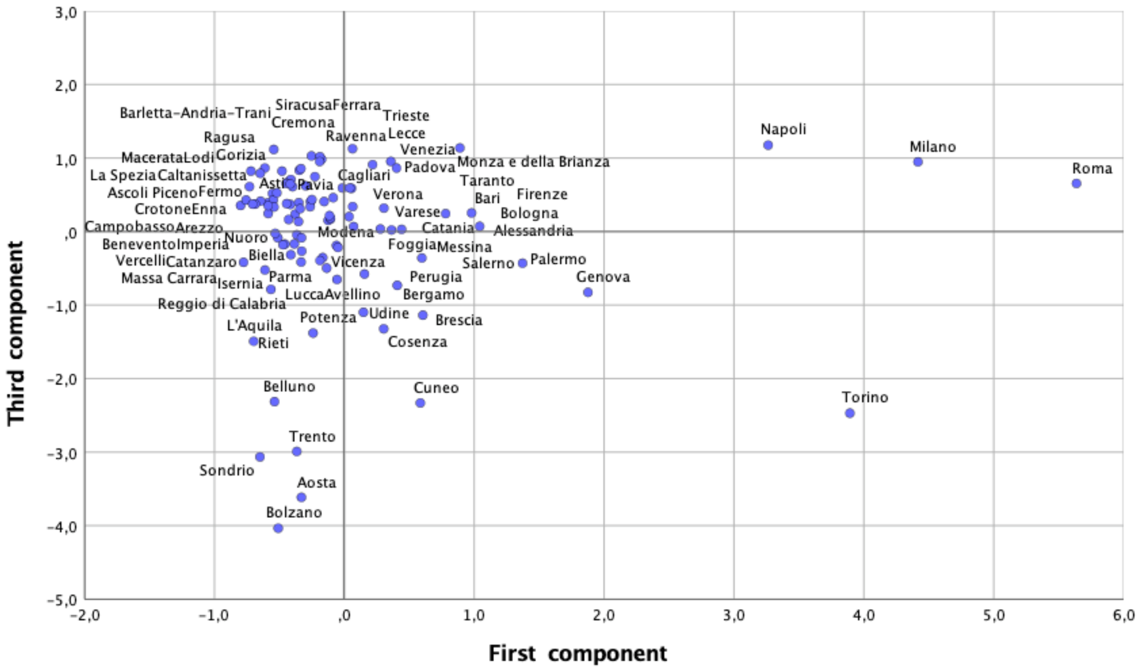

Fig. 10 Biplot graph of the principal component analysis (PCA) between the first and third components

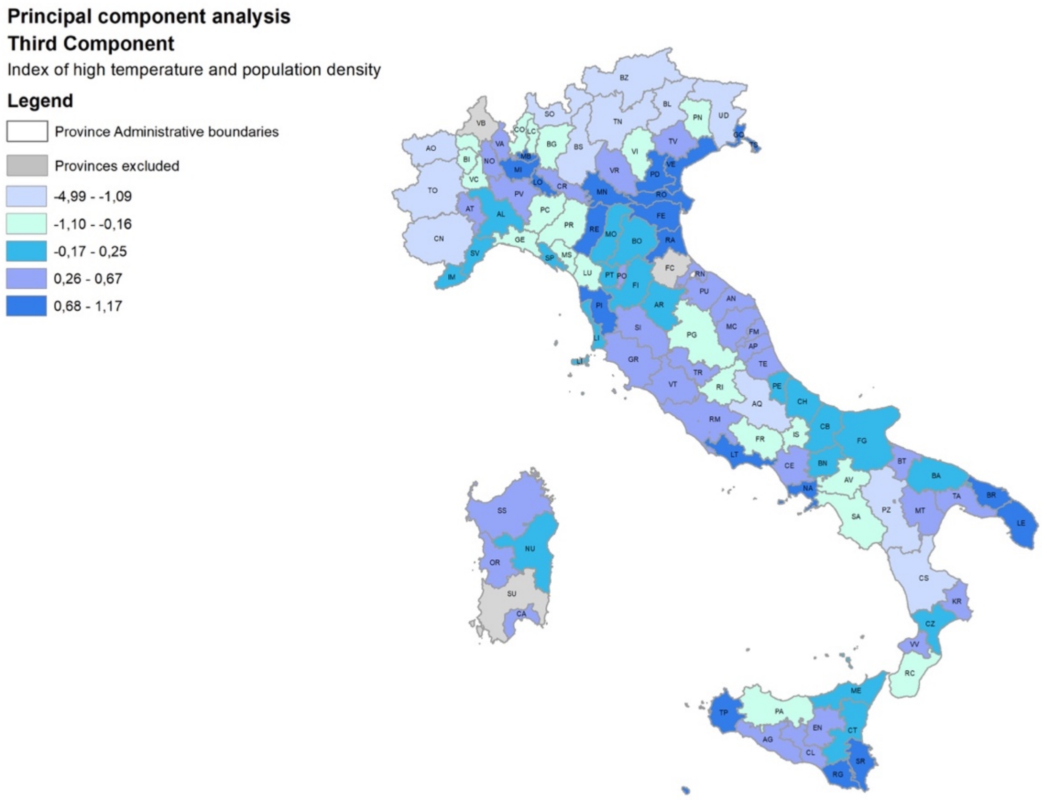

Fig. 11 The values of principal component analysis (PCA) of the third component for the Italian provinces

The output related to the order of cases (provinces) based on the correlation of each one with the third component confirms this reading, placing Naples in last place in the order, followed by Trieste (city in structural crisis), Ragusa and Brindisi. At the top of the ranking are the provinces of Bolzano and Aosta, Sondrio, Trento, Belluno, not surprisingly all located in the pre-alpine area.

As regards the other main components, please refer to the comments proposed for the third cluster for the identification of phenomena of urban and territorial nature that may be associated, positively or negatively, to the spread of infection.

As regards the other main components, please refer to the comments proposed for the third cluster for the identification of phenomena of urban and territorial nature that may be associated, positively or negatively, to the spread of infection.

The fourth and fifth components play a relatively marginal role in the definition of the latent structure of the system, as they explain, respectively, only 11% and 8% of the total. However, the respective variances, added to the values of the components already analysed, reach (cumulate) more than 80% of the variance explained in the interpretation of the entire set of variables under examination. As regards the fourth component, the strong influences (with the minus sign) of the old-age dependency ratio and of the average age values are to be highlighted, whilst, on the contrary, with the plus sign, the natural balance is highlighted. The other values are relatively low (to be considered as irrelevant) and in many cases close to zero on the scale. In terms of a synthetic reading of the structure of the system, it can be said that the fourth component gathers in itself the dynamics (negative and positive) of the growing resident "ageing population". The fifth component refers, positively, to the size of the area affected by the provincial capital and the presence of universities. On the contrary, it explains, but with relatively low values, the settlement density of the capital and the province as a whole. In other words, it is possible to recognise the characteristic features of the "large metropolitan areas" not only due to the territorial extension but, above all, due to the strong influence of the provincial capital with respect to the entire province

4.3 The Cluster Analysis

The multivariate analysis techniques aimed at defining clusters are made up of a set of procedures useful to identify, amongst objects of any nature, some clusters that are mutually exclusive and tend to be homogeneous within them in terms of characteristics and quantity. To define the unambiguous belonging to a group, a distance from a representative point of the cluster is used, having set the number of groups to be obtained (partitional clustering). Alternatively, a hierarchy of partitions characterised by an increasing number of groups is constructed, which can be visualised by means of a tree diagram (dendrogram), in which the grouping (hierarchical clustering) is represented (Morrison, 1990).

The method chosen for the objectives of the study is the two-step analysis of the partitional clustering that enables natural groupings to be revealed within a data set that otherwise would not be evident. The algorithm used differs from traditional clustering techniques due to the ability to create clusters based on both discrete and continuous variables and for the ability to analyse large datasets. In addition, the two-step analysis allows the identification, using the Bayesian Information Criterion (BIC), of the optimal number of clusters into which to divide the data sample (Carpentieri & Papa, 2018).

The IBM SPSS Statistic Version 20 software was used to perform the two-step cluster analysis. Performing the two-step cluster analysis with this software requires some preliminary operations. As with the statistical analyses presented in the previous paragraphs, the dataset used (104 provinces and 21 variables) has been normalised in order to prevent the different units of measurement of the variables used from influencing the interpretation of the model outputs.

To enable data processing, data normalisation, cluster creation and the likelihood method for measuring the distance between clusters have been implemented in rapid succession.

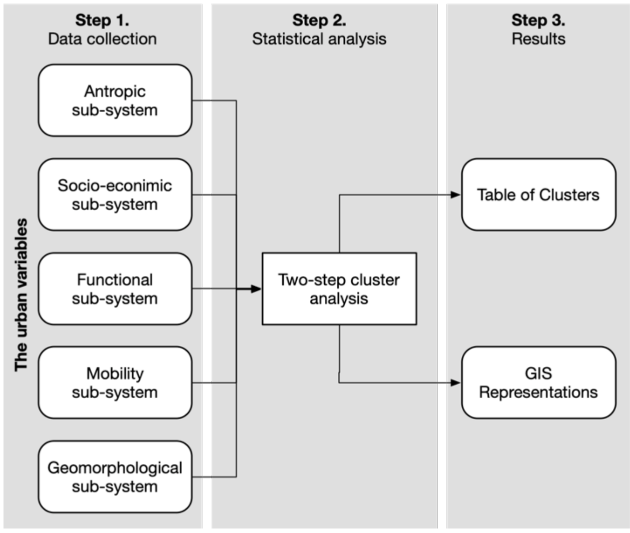

Fig. 12 The conceptual procedure for the Two-step cluster analysis

To complete the procedure, the software provides, as output, the optimal number of clusters into which to divide the data set, the table showing the value assumed by each variable in the different clusters and the table associating the cases (provinces) to the clusters. This last output also allows the realisation of graphic representations and/or on a map of the results of the division into clusters of the analysed elements.
After verifying the feasibility of clustering analysis, this process is defined us assessing cluster tendency and the model was developed.
The first application of the two-step cluster analysis provided three as the optimal cluster number through the Bayesian Information Criterion (BIC). Following the reading of the results obtained from this first application, it was decided to make further attempts at clustering, which would allow, for the same measure of validity (identified by the software in the internal cohesion), for an easier interpretation of the results considering the characteristics of the data and associated information. The number of clusters identified is six.

Results of the Cluster Analysis

After carrying out the PCA that enabled the structure of the settlement system to be determined, the Cluster Analysis was developed to define the "aggregation criteria" between territories, i.e., the distribution of settlement characteristics with respect to their geographical location. The comparison between the results of the cluster analysis and the data collected on the pandemic, which describe the intensity and spread of the infection in the Italian provinces, enables an association of the characteristics of the subsystems in which the urban and territorial system of the provinces has been discretized (settlement, social, economic, geographical, etc.) to the evolutionary phenomenon of the epidemic, in order to understand possible and different correspondences with homogeneous clusters of provinces.

Cluster 1 - Late Southern Italy development

The first cluster includes a large number of provinces (32) that geographically encompass the entire South, part of Sardinia and almost all of Lazio. The variables that mainly condition this cluster refer to the economic-productive characteristics, which assume values well below the national average. The distance from the area of first infection (the province of Bergamo), together with the variables related to the average temperature, above the national average values, contribute to the geographical definition of this cluster.
The analysis of the data on the intensity and spread of infection, both in absolute value and in relation to the population, shows that, in the provinces that constitute the first cluster, the epidemic on average has been

very low. The great distance from the areas of first infection, together with the measures of containment and social distancing adopted promptly during the lockdown period, were likely the two key factors to stem the pandemic in southern Italy. A second favourable circumstance, but only in times of pandemic, could be represented by the lack of economic-productive dynamism that has allowed the measures to be respected on time, reducing the possibility of infection.

Cluster 2 - The Italian province

The second cluster includes a large number of provinces (30) that are mainly distributed in Central and North-Eastern Italy. The prevailing feature that distinguishes this cluster is its small size. In fact, the size, both in demographic terms and in terms of territorial extension, assumes values far below the national average, both with reference to the province and the provincial capital. Further significant elements for the description of this cluster are the descriptive variables of the population structure, which show a high average age and a high old-age dependency ratio.

In comparison with the data by province on the number of infected people - both in absolute value and in relation to the population - and on the number of deaths, it shows that, in the provinces that comprise the second cluster, the epidemic on average remained at low values.

The low population size (together with its characteristics) and the low settlement density seem to have played an important role in containing the infection. Furthermore, the lack of economic-productive dynamism that characterises (even if to a lesser extent than the previous one) this cluster has allowed a stricter respect of the lockdown rules, reducing the possibility of infection.

Cluster 3 - The pre-alpine system

The third cluster comprises a small number of provinces (5) which are distributed exclusively in the territory of the Alpine crown. Given the population density, resident population and elderly population which have values well below the national average, the pre-Alpine system is characterised by a strong dynamism - both in terms of economy and production (income, GDP and active population) and in terms of demography (natural balance), with very high values compared with the national average, in provinces by territorial extension and by size of the relevant mountain area.

In comparison with the data by province on the number of infected people - both in absolute value and in relation to the population - and the number of deaths, it emerges that the pre-Alpine system was strongly affected by the epidemic.

Unlike cluster 2, in this case, the low population size and low settlement density failed to play in favour of containing the infection. This characteristic, therefore, does not seem to guarantee a viral containment if, at the same time, productive economic conditions of an opposite sign to those of the cluster under examination do not occur. In other words, it seems that the strong economic-productive dynamism, which distinguishes the pre-Alpine system, has contributed to the expansion of the infection, beyond the demographic and settlement characteristics which, therefore, lose relevance.

Cluster 4 - The economic-productive engine of Italy

The fourth cluster that we can define, without a shadow of a doubt, the economic-productive engine of the country, includes 20 provinces that, in the shape of a circular crown, develop around the Milan area and are in Lombardy (the predominant part) as well as in Veneto and Emilia Romagna. This cluster is characterised by some peak values, which are much higher than the national average, in the sectors concerning the percentage of the active population and the value of the average income per capita. With a natural balance higher than the Italian average and an average age lower than the average of the entire country, there is a GDP per inhabitant that is second only to that of the large metropolitan cities. These territories are characterised by a relatively low settlement density with small provincial capitals and a relatively small number of university sites. The number of beds per inhabitant is below the national average, as is the presence of the over-65 population. The infection hit these areas hard, not only because they were close to the initial outbreaks or because they themselves fuelled new outbreaks, but also because the restrictive measures did not interrupt production activities of strategic interest, a characteristic recognised to all companies that with a self-declaration claimed to belong to a strategic production chain. The initial indecision as to whether such drastic measures should be taken for what many of the institutions and companies within this cluster considered, in the initial phase, the Coronavirus epidemic to be little more than a trivial flu, contributed decisively to the situation in these territories.

Cluster 5 – "Middle" Italy

13 provinces form part of this cluster, the fifth of the six clusters that constitute the output of the model used in this study, located almost exclusively on the Apennine ridge of Central Italy and on the Veneto-Romagna coast to the Trieste area. This grouping can be concisely defined as Middle Italy. The values that it presents with respect to the 21 variables that constitute the input of the model are very close to those of the national average, with some exceptions that, however, characterise these territories and enable the latent structure to be read. Firstly, these are areas that act as a crown outside the provinces of cluster four and are therefore relatively farther away from the areas of first and major infection in inner Lombardy. Secondly, they have average per capita income values aligned with those of the richest areas of the country, they have an over-65 population lower than the national average and they are characterised by an average age significantly higher than the national average in line with the areas of central Italy (luster 2). Cluster 5 is characterised by the highest percentage of universities in the whole country and an exceptionally low settlement density of the provincial capital. Also with regard to the Covid-19 infection, these areas have confirmed themselves as a median area between the most affected cities and territories and those where the infection has been relatively low.

Cluster 6 - The system of metropolitan areas

The sixth and final cluster includes the large metropolitan areas of Italy: Milan, Rome, Naples and Turin. This group is characterised by the abundantly predictable role of some variables typical of large metropolitan areas, such as the settlement density of the provincial capital, the average per capita income, the percentage of the over-65 population, the strong presence of hospital beds and, of course, the average income of the capital and the GDP per inhabitant. However, there are some anomalies concerning University locations where metropolitan cities are outclassed by the provinces of cluster 5, which has an almost double value and, above all, a natural balance that is not only far below the national average, but is equal to half of the values present in all the other clusters. The Covid-19 pandemic in these metropolitan areas has developed all its potential for infection and death due to the high density of settlements, the reluctance to accept, firstly in Milan and Turin, the need to interrupt not only social contacts but, above all, economic and productive activities. The widespread presence of the elderly population in these cities, with the exclusion of Naples, has been devastated by the virulence of the infection and by errors in emergency management, especially as regards the protocols of acceptance of hospital facilities (which have become real hotbeds of infection) and the guilty transfer of sick and/or recovering elderly people to care homes, with the natural consequence of a general transmission of the infection to most of the residents of these facilities. The metropolitan cities of Rome and Naples have managed to contain the impact of the infection as they were delayed by the spread of the virus, a delay that has enabled them to implement an effective preventive lockdown and to benefit from the mistakes made in the areas most affected by the epidemic. Finally, the immediate closure of schools and universities, in conjunction with the prevailing commitment in the services businesses of city users and, therefore, the widespread possibility of replacing office work with work from home, further contributed to the containment of infection.

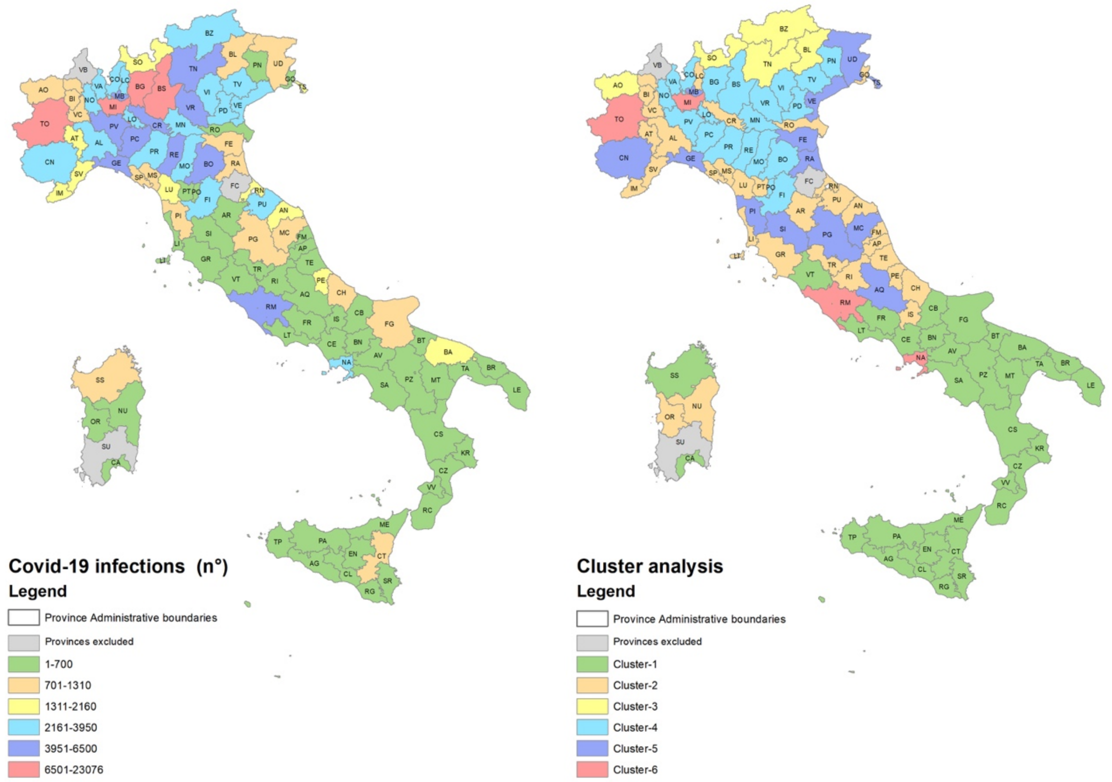

Fig.13 Comparison between the distribution of the infection at May 31, 2020 and the clusters of the Italian provinces

4. Conclusions

There are essentially three scientific questions that this work wanted to answer. The first concerns the possible correlations that can be established between the intensity and spread (as at 31 May 2020) of Covid-19 infection and the settlement system of our country. The second question relates to the identification of urban and territorial phenomena that may be associated, either positively or negatively, with the spread of infection. The third question concerns the update (2020) of the main phenomena of settlement nature (*inter alia*) that can be recognised within the system of Italian provinces in order to identify, with reference to the intensity and spread of infection, possible and different correspondences with homogeneous clusters of provinces.
Unfortunately, none of the variables identified are representative of the characteristics of urban and territorial planning - which is the specific discipline of urban planning in the strict sense.
However other predictors contributed - in conjunction with the identification of the latent system of the Italian provinces obtained through the analysis of the Principal Components and especially with the Cluster Analysis that complete this study - to identify phenomena of coexistence and collinearity between the consistency and distribution of infection and the settlement system of our country. The variable number of hospital beds is the most important predictor of the entire panel of variables given that its normalised coefficient (0.770) takes the highest value and its significance (0.003) the lowest value and consequently the highest importance. As expressly mentioned in the section of the study relating to the definition of the correlation between the variables used, this data is strongly correlated with other very significant indicators of the settlement structure of our country - resident population, over-65 population, vehicle fleet, daily movements, etc. - and, therefore,

the considerations on this predictor must take into account, the baggage of information underlying the beds variable. In other words, the number of beds can be identified as a multiplier of the infection, not only because the hospital facilities have become real multipliers of the infection, in conjunction with the RSA, but also because they are a measure related to the number of elderly people over the age of 65 (age group particularly exposed to the infection and with fatal consequences), the number of residents and the dynamism of the exchanges of each province.

The results of this analysis have suggested increasing the knowledge of the latent structure of our settlement system in order to complete the analysis underlying our study and allowing a comparison of the intensity and spread of infection and the anthropic, socio-economic, functional, spatial and geo-environmental characteristics of the metropolitan areas and provinces of our country. The Principal Component Analysis and Cluster Analysis highlighted the "aggregation criteria" between the different territories. The comparison between the results of and the data collected about the pandemic, which describe the intensity and spread of the infection in the Italian provinces, allowed an association of the characteristics of the subsystems in which the urban and territorial system of the provinces has been discretized (settlement, social, economic, geographical, etc.) to the evolutionary phenomenon of the epidemic, in order to understand possible and different correspondences with homogeneous clusters of provinces.

Author Contributions

The work, although the result of a common reflection, was divided as follows: Carmela Gargiulo: conceptualization and methodology; Federica Gaglione, Carmen Guida and Floriana Zucaro: data curation, investigation and visualization; Rocco Papa: supervision; Gerardo Carpentieri, corresponding author, software, formal analysis and investigation.

References

Bottero, M., Mondini, G., & Datola, G. (2017). Decision-making tools for urban regeneration processes: from Stakeholders Analysis to Stated Preference Methods. *Tema. Journal of Land Use, Mobility and Environment, 10(2)*, 193-212. https://doi.org/10.6092/1970-9870/5163

Carpentieri, G., & Papa, R. (2018). Classifying railway station catchment areas. An application of node-place model to the Campania region. In Leone, A., & Gargiulo, C. (Editors). *Environmental and territorial modelling for planning and design* (Vol. 4). FedOA-Federico II University Press. 229-307. https://doi.org/10.6093/978-88-6887-054-6

Chakraborty, J., & Reader, S. (2010). Multivariate Analysis Methods. In B. Warf (Ed.), *Encyclopedia of Geography*, pp. 962-1962. Thousand Oaks, CA: SAGE Publications. http://dx.doi.org/10.4135/9781412939591.n793

EUROSTAT – European Statistics (2020). Healthcare services in Europe. Kirchberg, Luxemburg: European Commission. Retrieved from: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/healthcare-services

Fang, W. & Wahba, S. (2020). Urban Density Is Not an Enemy in the Coronavirus Fight: Evidence from China. Retrieved from: https://blogs.worldbank.org/sustainablecities/urban-density-not-enemy-coronavirus-fight-evidence-china

Fotheringham, S. A., & Rogerson, P. A. (1993). GIS and spatial analytical problems. *International Journal of Geographical Information Science*, *7*(1), 3-19. https://doi.org/10.1080/02693799308901936

Fotheringham, S., & Rogerson, P. (2013). *Spatial analysis and GIS*. Florida, US: CRC Press. ISBN: 1482272466

Gargiulo, C. (2009). Sistema Urbano e complessità. In Papa R. *Il governo delle trasformazioni urbane e territoriali – metodi tecniche e strumenti*, 23-46. FrancoAngeli Editore. ISBN 978-88-568-1147-6.

Gargiulo, C., Zucaro, F., & Gaglione, F. (2018). A Set of Variables for the Elderly Accessibility in Urban Areas. *TeMA-Journal of Land Use, Mobility and Environment*, 53-66. https://doi.org/10.6092/1970-9870/5738

Gargiulo, C., & Papa, R. (1992). Caos e caos: la città come fenomeno complesso. In AA. VV. *Per il XXI Secolo: una enciclopedia e un progetto,* 297-306. DiPiST-Università degli Studi di Napoli Federico II. ISBN: 978-88-97110-15-6

Giordano, P. (2020). *Nel contagio*. Torino: Giulio Einaudi editore. ISBN: 885843420X.

Härdle, W., & Simar, L. (2003). *Applied multivariate statistical analysis*, 22007, 1051-8215. Berlin: Springer. https://doi.org/10.1007/978-3-662-45171-7

Hou, D., O'Connor, D., Nathanail, P., Tian, L., & Ma, Y. (2017). Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. *Environmental Pollution, 231*, 1188-1200. http://dx.doi.org/10.1016/j.envpol.2017.07.021

ISPRA – Istituto Superiore per la Protezione e la Ricerca Ambientale (2014). Il consumo di suolo in Italia - Edizione 2014. Roma: ISPRA – Rapporti. ISBN 978-88-448-0646-0

ISTAT - Istituto Nazionale di Statistica (2011). Statistiche Istat. Popolazione residente—Censimento 2011—Superficie Delle Abitazioni Occupate da Persone Residenti. Retrived from: http://daticensimentopopolazione. istat.it/

ISTAT & ISS - Istituto Nazionale di Statistica & Istituto Superiore di Sanità (2020). *Impatto dell'epidemia COVID-19 sulla mortalità totale della popolazione residente nel primo quadrimestre 2020.* Retrieved fromhttps://www.epicentro.iss.it/coronavirus/pdf/Rapp_Istat_Iss_3Giugno.pdf

Johnson, R. A. & Wichern, D. W. (1998). *Applied Multivariate Analysis*, 4th ed., New Jersey: Pearson Prentice Hall. ISBN: 0-13-187715-1

Mandal, S., & Mondal, S. (2019). Logistic Regression (LR) Model and Landslide Susceptibility: A RS and GIS-Based Approach. In S. Mandal & S. Mondal (Eds.) *Statistical Approaches for Landslide Susceptibility Assessment and Prediction,* pp. 107-121. Switzerland: Springer Cham. https://doi.org/10.1007/978-3-319-93897-4_4

Masoumi, H. E., & Shaygan, M. (2016). A Longitudinal Analysis of Densities within the Pedestrian Sheds around Metro Stations. The Case of Tehran. *Tema. Journal of Land Use, Mobility and Environment*, 5-20. https://doi.org/10.6092/1970-9870/3908

Maxwell, A.E, (1981). *Analisi multivariata nella ricerca sociale*, Milano: Franco Angeli. ISBN: 8820419017

MIT – Ministero delle Infrastrutture e dei Trasporti (2019). Open data. Retrieved from: http://dati.mit.gov.it/

Morin, E. (1983) Il Metodo, trad. it., Idee/Feltrinelli, Milano. ISBN: 8807090023

Morin, E. (1993). *Introduzione al pensiero complesso*. Milano: Sperling & Kupfer. ISBN: 882001520X

Morrison, D. F. (1990). *Multivariate statistical methods*. New-York: McGraw–Hill. ISBN: 0070431876

Muirhead, R. J. (1982). *Aspects of Multivariate Statistics*. New York: John Wiley and Sons. http://dx.doi.org/10.1002/9780470316559

Papa, R. & Piscopo, O. (1988). *Città intermedia e innovazione tecnologica*. Napoli: Dipist- Università di Napoli Federico II, Collana di Urbanistica. ISBN: 9788897110064

Sadocchi, S. (1981). *Manuale di analisi statistica multivariata per le scienze sociali*. Milano: Franco Angeli. ISBN: 8820437260

Stillwell, J., & Clarke, G. (Eds.).(2004). *Applied GIS and spatial analysis*. Chichester:Wiley.ISBN:978-0-470-84409-0

Zucaro, F., & Morosini, R. (2018). Sustainable land use and climate adaptation: a review of European local plans. *TeMA Journal of Land Use, Mobility and Environment, 11(1)*, 7-26. https://doi.org/10.6092/1970-9870/5343

Website

Il Sole 24 Ore (2017). La mappa dei redditi nelle città. Retrieved from: https://lab24.ilsole24ore.com/mappaRedditi/index.html

Author's profile

Carmela Gargiulo

She is full professor of Urban Planning at the University of Naples Federico II. Since 1987 she has been involved in studies on the management of urban and territorial transformations. Her research interests focus on the processes of urban requalification, on relationships between urban transformations and mobility, and on the estate exploitation produced by urban transformations. On these subjects she has co-ordinated several research teams. Author of more than 150 publications.

Federica Gaglione

She is an engineer, Ph.D. student in Civil Systems Engineering at the University of Naples Federico II. Her research concerns the topic of urban accessibility. From August to December 2019 she served as a Visiting Researcher at the University of Aberdeen (UK) undertaking a significant amount of research regarding pedestrian accessibility for older persons.

Carmen Guida

She is an engineer, Ph.D. student in Civil Systems Engineering at Department of Civil, Architectural and Environmental Engineering of University of Naples Federico II. Currently, her Ph.D. research concerns accessibility to urban services for elderly people with the aim of minimizing social exclusion and inequalities within urban areas.

Rocco Papa

Professor Emeritus of Land Use Planning at the University of Naples Federico II, Polytechnic School and Basic Sciences. Full Professor of Land Use Planning at University of Naples Federico II, from 1993 to 2017. Director of the Department of Urban and Regional Planning DiPiST of University of Naples Federico II, from 1999 to 2005. Vice-Mayor of the Municipality of Naples, from 2001 to 2006. City Councilor for Livability for the Naples Municipality, from 1997 to 2001. Research activity, carried out continuously since 1974, has developed according to the following four main lines: the study of the interactions between urban and mobility systems; the management and governance of metropolitan areas. He is the author of more than 160 publications.

Floriana Zucaro

She is an engineer, she received a M.Sc. in Environmental and Territorial Engineering at the University of Naples Federico II with a specialization in management of urban and territorial transformations. In April 2015 she holds a PhD in Hydraulic, Transport and Territorial Systems Engineering at University of Naples Federico II. Since 2018 she has been a research fellow on the "MOBILAGE - Mobility and aging: daily life and welfare supportive networks at the neighborhood level" project financed by Fondazione Cariplo.

Gerardo Carpentieri

He is an engineer, Ph.D. in Civil Systems Engineering at University of Naples Federico II and Research Fellow of Land Use Planning at the Department of Civil, Architectural and Environmental Engineering of the University of Naples Federico II. From 2020, he is a member of Unina research group for the ERASMUS+ Key Action2: Project "Development of a Master Programme in the Management of Industrial Entrepreneurship for Transition Countries" (MIETC).