Friulian Journal of Science **24.** 2017, 45-57

Territorial reorganisation issues at NUTS-2 level: reflections on the Friuli Venezia Giulia experience

GIAN PIETRO ZACCOMER*

Abstract. The main purpose of this article is to put forward a few general reflections regarding the problem of territorial reorganisation of a NUTS-2 region in light of the teachings of Prof. Marzio Strassoldo. More specifically, this work follows his belief that the territory, in statistical modelling, should not be considered as a mere spatial constraint to be fulfilled or underlying assumption to be taken for granted. In this sense, in his thinking you can find a kind of *territory supremacy* that should not only be understood in the sense of a territory's ability to modify its socioeconomic values, which even "under the same conditions" can change from territory to territory, but also to make, with its full component of relationships, certain models recommended in the literature only partially applicable to the regional case being studied. This trans disciplinary view, already present in his thinking at the end of the Eighties, was extremely uncommon among statisticians, including those working in economics, and can almost certainly be traced back to his first scientific interests that focused more on Economic Geography than Statistics. It is not by chance that he believed that skills in several scientific fields were necessary to fully analyse a territorial phenomenon, itself often covering a range of disciplines. Indeed, he recognised an increasing need for interdisciplinary figures, both in the professional and academic field (*"borderline" researchers* as he liked to call them, implicitly playing on the geopolitical position of Friuli Venezia Giulia), who would offer a broad view that crossed the boundaries between each disciplinary area. That is why he was a strong supporter of *Statistics and Computer Science for Business Management* degree courses (Bachelor's and Master's degree) at the University of Udine, based, as is easily inferred from their name, on three distinct skills. These courses provided, in turn, very useful professional figures to a territory whose economic structure, similar to the national one, largely comprises small and medium-sized enterprises.

* Department of Languages and Literatures, Communication, Education and Society (DILL), University of Udine. Udine: Italy. E-mail:gianpietro.zaccomer@uniud.it

This work focuses on territorial reorganisation exercises in Friuli Venezia Giulia through the use of *Cluster Analysis*, a classical multivariate statistical analysis tool. A number of suggestions, based on various empiric experiences gained in over twenty years in the field, will be presented. These cover the main decisions a researcher has to take during his analysis and include those related to metrics, agglomeration strategies and contiguity constraints. In particular, with respect to the last point and its significant methodological criticalities, suggestions will be provided on how to use *privileged direction indicators* of the phenomenon being studied.

Key words. socioeconomic indicators; territory; cluster analysis; contiguity constraint.

1. Introduction. This work aims to put forward a few general considerations regarding the statistical problem of territorial reorganisation of a NUTS-2 region, starting with the experiences gleaned through research by the University of Udine, over the last 25 years, into Friuli Venezia Giulia (FVG), a NUTS-2 level region located in the North East corner of Italy. These experiences mainly concern: the creation of a regional energy planning model; the first implementation of the regional law on fuel prices and its subsequent amendments; the definition of industrial district boundaries and studies covering the collapse of this economic model; and, last but not least, other studies into various socioeconomic phenomena that have affected and still affect, FVG. These experiences involved Prof. Strassoldo from different points of view (as research leader in the Economic Sciences Department and then the Statistical Sciences Department, and as Chief scientist for the journal, Osservatorio Permanente dell'Economia del FVG; however, the considerations proposed here stem largely from his teachings and ideas regarding how, in statistical modelling, a territory should be adequately – not trivially – taken into consideration.

At the end of the Eighties, the statistical modelling of georeferenced data was still evolving, even though the foundations for current models had already been laid (Ripley 1981). It was, however, still too underdeveloped to have gained an important role within its field. Its evolution has led to the theoretical and instrumental *corpus* that is now generally referred to as *Spatial Statistics*, or rather, the branch of Statistics that uses space and spatial relationships explicitly in its modelling. It is applied to all those sciences, including socioeconomic ones, where space and territory play an important role (Arbia, Espa 1996). Several applied disciplines have contributed to its development, (physics, mathematics, medicine, biology, etc.), as have certain ideas that – a few decades earlier – were born within geographical thinking, especially its economic viewpoint.

In some of the regional experiences just mentioned, Prof. Strassoldo had already contributed to the emergence of the concept of a kind of *territory supremacy*, not only in the sense of a predominant role of the territory capable of significantly modifying the manifestation of socioeconomic phenomena, but also in terms of making, with its full component of relationships, certain models recommended in the literature only partially applicable to the regional case being studied. Such a point of view was uncommon among statisticians who, at the time, still largely saw the territory as an obstacle to any model. Indeed, most orthodox economists traditionally disregarded the territory in their modelling for many years. This background awareness was implicitly based on overcoming the heated academic debate between *orthodox geography* and *quantitative geography.* This debate began in Italy in the Seventies and still continues, despite the strong development of GIS (Romagnoli 2002, p. 15). Overcoming this juxtaposition means that geography must use spatial statistical tools with awareness. Such results do not mean the end of the research process, but rather its continuation in a way that sees all the qualitative and quantitative data previously left out of consideration now being integrated. Such data can enable the researcher not only to more critically interpret his results and more deeply understand the phenomenon, but also allow him to express qualitative evaluations of any violations of the hypotheses at the heart of almost every statistical model. In this way, the terminological shift from *spatial analysis* to *territorial analysis* is legitimised. The first is merely statistical and mainly focused on measuring and modelling phenomena, while the second stems from a geographical *transcalar approach* (at least in its weaker version, or *multiscalarity* as defined by Bonavero 2005, p. 5), which aims to fully understand the phenomenon and its relations with the territory that support it.

These meanings are very different from the ones Zani and Napolitano proposed in 1990 in an official statistical setting. Here, the first type of analysis deals with quantitative studies into phenomena that require geometric grids created by the researcher while the second is solely concerned with studies based on already created administrative divisions such as NUTS (Zani 1993, p. 94).

2. Reference theoretical framework. The considerations proposed here concern territorial reorganisation exercises that are undertaken when a regulation or research objective requires it. Such exercises, from a statistical point of view, are usually intended to identify a specific grouping of territorial units by means of a given set of indicators[1]. Considering the FVG region, several similar studies have been done making use of official LAU-2 (i.e. municipal) level data. In this case, the problem involved finding a different division to those of an administrative basis, one that would group municipalities considered homogeneous in terms of the analysed socioeconomic phenomenon. To find a solution to such a problem, a well-established statistical tool is usually used: *Cluster Analysis* (CLA) is a multivariate statistical method that allows one to group territorial units – municipalities in this case – into wider homogeneous areas, respecting both i*nternal cohesion* (units belonging to the same group must be as alike as possible) and e*xternal isolation* (groups must be as different as possible from each other). The emerging results do not depend solely on the indicator set chosen for that analysis, but also on certain fundamental choices made by the researcher, such as: type of analysis, strategy and metric used for the agglomeration algorithm and, possibly, the introduction of territorial constraints (Cerioli, Zani 2007, pp. 267-457; Zani 1993, pp. 93-121).

The first of these choices is which CLA to use. An essential element for making this choice is knowing a priori the number of groups, as this may already be set by legislation: in such a case, the reorganisation exercise consists solely in identifying their structure. However, in most cases, identifying the number of groups makes up part of the problem to be solved. In this case a *hierarchical agglomerative analysis* is recommended as it provides, rather than just a single division, a whole hierarchy of partitions to choose from (ranging from many groups representing the individual municipalities, through to a single group comprising all municipalities). Its graphical representation is called a *dendrogram*, which is very useful in helping one understand, on the basis of its cartograms, how the territorial units are spatially grouped, this being a very important aspect for the final interpretation of the CLA results.

Focusing on the second choice, that of the agglomeration algorithm strategy, the literature proposes different algorithms: the *single linkage (nearest neighbour*) and the c*omplete linkage (farthest neighbour)*. The first produces the well-known *chaining effect,* that is to say, a tendency to create only a few initial groups to which all other territorial units are then subsequently linked. The risk, here, is to compromise the internal cohesion; on the other hand, this algorithm is less influenced by any possible correlation between indicators in the set. The complete linkage algorithm, in contrast, does not create the chaining effect but rather pinpoints a set of well-balanced homogeneous groups. This is a good algorithm to use, especially if one is studying clear-cut and ideally dichotomous phenomena within the territory without any overlapping areas.

Otherwise, an increasing (albeit small) imbalance will emerge in the numerosity of the groups. And worse still, unlike the first algorithm, this algorithm is influenced by any correlation between indicators. In seeking to avoid these flaws different algorithms have been proposed: *average linkage* (*between* and *within groups*), the *median method* and *centroid method* provide intermediate results with respect to those obtained using the more "extreme" strategies just discussed.

Regarding these theoretical considerations, in most studies on Friuli Venezia Giulia three types of problems, occurring simultaneously, have been highlighted: no particular distribution of the set's indicators; the presence of a correlation between indicators; the presence of overlapping areas where the phenomenon's behaviour on a territorial level is unclear. In such cases the *Ward algorithm* has been used, with some satisfaction: by being based on a minimisation of the total variance increase at each step of the algorithm, it is not influenced by the chaining effect, so guaranteeing the groups' internal cohesion; however, unlike the complete linkage strategy, it produces well-balanced groups even in the presence of overlapping areas (see Marra 1990, pp. 171-175). The main problem if you want to apply this algorithm is that, unlike single linkage, is highly sensitive to any correlation between indicators. Using the *Principal Components Analysis* (CPA) as a preliminary to the CLA is one way to solve this problem at its root. The Principal Components (unrotated) are simply s*ynthetic indicators*, the result of linear combinations of the set's indicator that, by their nature, are uncorrelated. This not only solves the problem but also allows a legitimate choice of the *Euclidean metric*, that is to say, the third decision a researcher must take. If the CPA legitimates this choice, using its square makes choosing the proper c*ut-off point of the dendrogram* a lot easier, identifying both the optimal number and the groups' structure, so guaranteeing both the internal cohesion and the external isolation. This approach has been confirmed by another theoretical result: it is possible to apply the Mahalanobis metric directly to the set's indicators rather than the Euclidean one. The Mahalanobis metric allows one to directly take account of any correlations between indicators. It was shown that, when applied to standard indicators (a necessary step because of the different measurement units present in the set), this kind of metric achieves the same outcomes to those obtained when applying the squared Euclidean distance to the principal (non-factor rotated) components (Sadocchi 1987, p. 213-214).

The last aspect to be discussed, perhaps the most immediate if one considers the expected outcomes, is that of introducing *territorial constraints* within the agglomeration algorithm. This is usually the case when legislation calls for the creation of groups made up of neighbouring municipalities: introducing a *contiguity constraint* in the agglomeration algorithm can technically solve the matter. This allows the grouping of two municipalities if, and only if, these are neighbouring ones. Unfortunately this kind of approach upsets the subsequent idea of a hierarchical analysis because, due to the constraint effect, municipalities that are less homogeneous than non-neighbouring ones may be grouped together, meaning that the internal cohesion of such groups is no longer guaranteed. It also prevents the creation of the dendrogram and the determination of the number of groups. In other words, introducing a constraint takes one back to a non-hierarchical type of analysis and, subsequently, the need to know a priori the number of groups. Furthermore, it must be remembered that the groups obtained in this way are no longer internally cohesive. In short, from an operational point of view, using such a constraint is not recommended unless expressly requested by the end user of the research who, informed of the risks, prefers to favour operational aspects over those offering scientific accuracy.

One should now probably ask if it is possible to obtain a territorial division that, at least in basic terms, provides sufficiently contiguous groups without the need to insert a contiguity constraint in the agglomeration algorithm strategy.

3. Empirical findings on Friuli Venezia Giulia. The statistical problem that has been discussed concerns the possibility to create groups that respect both internal cohesion and external isolation, that are well-balanced in their numerosity, and made up of territorial units that border each other as much as possible. Empirical experience has highlighted how, in certain cases and at high levels of agglomeration, it is still possible to obtain sufficiently contiguous groups without having to introduce the constraint. This is because it is possible to introduce an indicator to the set that serves to determine the direction of the studied phenomenon, provided the phenomenon is not multidirectional.

In the case of FVG, in socioeconomic phenomena, average altimetry is certainly a legitimate indicator as it is substantially linked to every anthropic activity in its territory. It can be easily calculated using the Bravais-Pearson bivariate correlation analysis on every indicator in the set. This simple observation, valid in many territories, allows one to gain a practical solution to the problem because, due to its particular morphological configuration, the average municipality altimetry in FVG increases in a largely south-north direction (with the exception of a few municipalities in the south-east, particularly in the Carso area), so allowing the implicit inclusion in the analysis of a weak contiguity principle, which – although not always respected – by having no constraints, guarantees the internal cohesion of the groups. Average altimetry is not the only indicator one can use to determine the phenomenon's direction. In the study on regional versus Slovenian fuel prices, another indicator was used with a specific east-west spatial direction, namely the distance between every municipality and the nearest cross-border petrol station.

To understand how both direction indicators work, in a single cartographic image, a hierarchical cluster analysis has been done, using the Ward algorithm, exclusively on average altimetry and on distance from cross-border petrol station. The result for five groups is shown in Figure 1: it is graphically clear that, while altimetry produced a separation of groups in the south-north direction, the distance from cross-border petrol station produced an east-west directional separation. This analysis highlights how the groups created are "naturally" contiguous with no need for constraints. It is obvious that this is not a perfect solution, as the one obtained using a constraint would have been, given the separation of some municipalities from their reference group: in fact most of these municipalities fall under the different treatment provided for by regional legislation for those municipalities lacking a petrol station in their territory (see Zaccomer 2008, p. 44-45). In this case, for solely operational reasons, it is possible to include the few "separated" municipalities with the contiguous group surrounding them, in other words, rework the groups at a later stage, violating internal cohesion for just these few municipalities. It is a good compromise between the need for scientific accuracy and the operational needs of those who commissioned the study.

Before describing a more complex case in line with the one just proposed, it is worth talking about another possibility that emerges when an indicator is known solely for the division of a less dense territory, in this case, the province (NUTS-3 level), but which could also be the Local Labour Systems or the Inter-municipal Territorial Union. This largely occurs when information is obtained from too small a sample to give reliable municipal estimates. Generally two routes are followed in such a case: one can either exclude this indicator from the set or one can estimate the municipal value using an *allocation method*. The latter option is used when the information provided by the provincial indicator is considered essential for the phenomenon's analysis.

Usually, the allocation method, no matter how complex, is always based on information contained in other more well-known variables (for example population or size of municipal area); as such, not only does it produce a purely theoretical value, that at times is not even close to the actual value, but, when the "allocated" indicator is inserted in the CPA, it also risks seeing the whole estimation process wasted, especially when other indicators in the set use the same allocation variable for their construction. If, instead, one inserts the indicator directly at a higher level, one avoids this problem and tend to favour the grouping of municipalities from the same province, in other words, those neighbouring each other. In the fuel price case already discussed, having a sample-based estimate of provincial evasion rates and combining these with the direction indicator for distance from cross-border petrol station, produced the results shown in Figure 2.

In Figure 2 one can observe that the territorial reorganisation, achieved again for five groups, substantially respects the territorial contiguity and highlights the isolation effect of the Tagliamento River that, in the South, divides the province of Udine from the province of Pordenone. This feature was completely missing in Figure 1. Clearly, this latter method is the more "forced" one and it can only be used when there is not

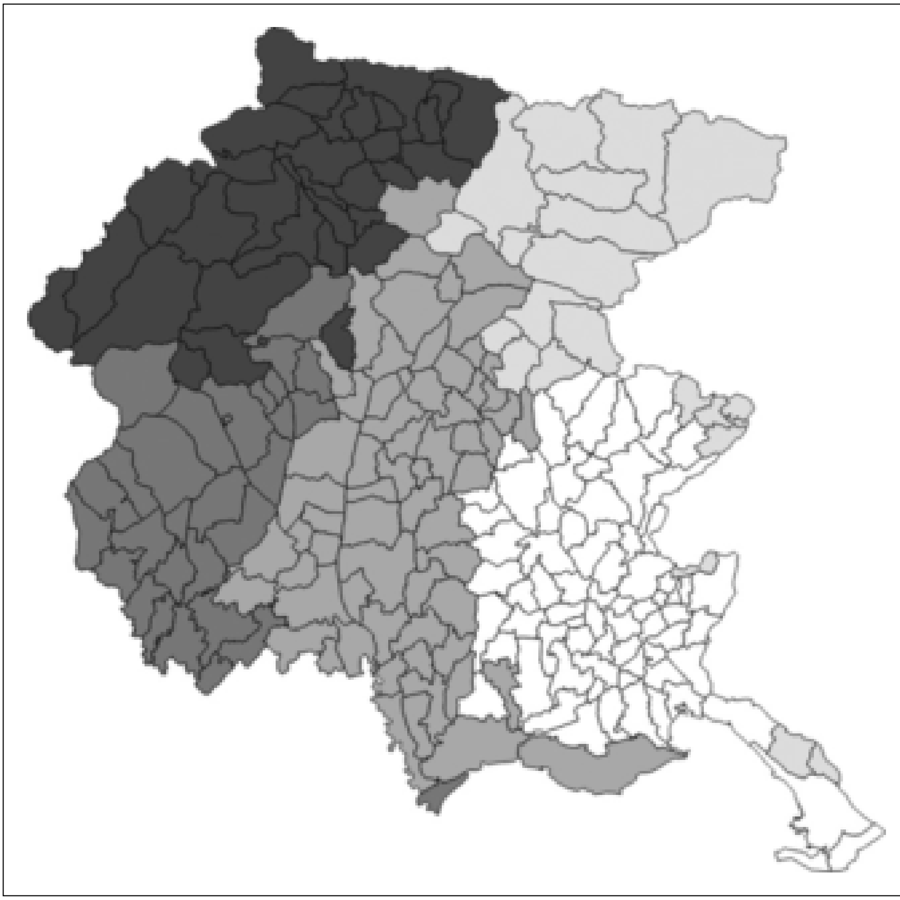

Figure 1. Territorial reorganisation using favoured directions. FVG cartogram for cluster analysis (five groups) based on distance from cross-border petrol station and on altimetry. (Source: own processing of FVG Regional Authorities and ISTAT data).

the same detailed territorial knowledge for each of the set's indicator; on the other hand, the method based on a favoured direction for the phenomenon is certainly the more elegant and formally exemplary one.

The question that of course arises, is: if one increases the number of indicators, can a single direction indicator still be useful? The answer is yes, especially when many of the indicators are strongly influenced by it and the external isolation criterion produces a rather low number of groups. To prove what has just been said, below is shown a study that was based on a

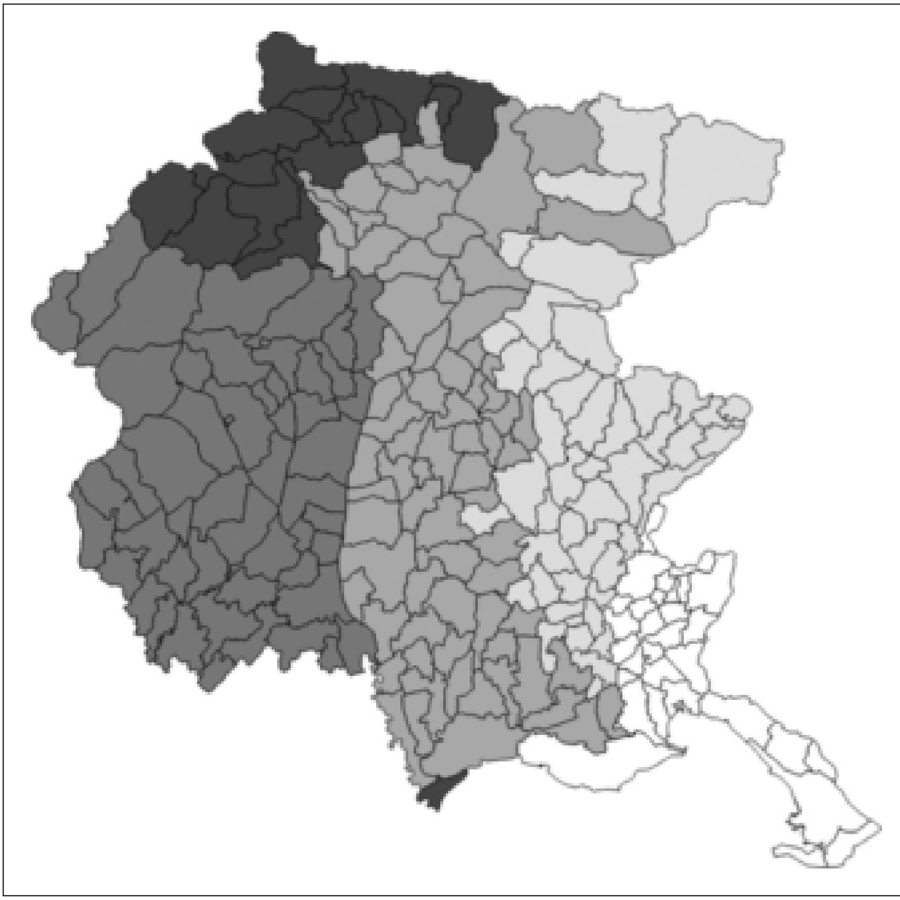

Figure 2. Territorial reorganisation using favoured directions. FVG cartogram for cluster analysis (five groups) based on the distance from cross-border petrol station and on provincial evasion rate. (Source: own processing of FVG Regional Authorities and SWG data).

variegated set of social indicators[2], for the benchmark statistical period 2010-11, to which was added the municipal average altimetry variable, taken from the *Digital Terrain Model* of ISTAT, providing a total amount of 15 indicators. The highest gap dendrogram cutting method identified two large groups, as can be observed in Figure 3, which were essentially contiguous without the need to introduce a constraint. From a spatial point of view, the same figure shows how the greatest difference exists between the "socially less advantaged" mountain municipalities and the rest of FVG[3].

Figure 3. FVG cartogram for cluster analysis (optimal solution for two groups) based on 14 social indicators and average altimetry. (Source: own processing of FVG Regional Authorities, ISTAT, MIUR and ACI data).

It is clear that when one wishes to increase the number of groups, so violating the hypothesis of maximum difference between groups, it may prove insufficient to insert just one phenomenon direction indicator. For example, cutting the dendrogram at a lower level, so as to have three groups, allows the social diversities across the main urban areas of FVG and of Lignano Sabbiadoro (which represents a very particular reality) to emerge, as can be seen in Figure 4. From a scientific point of view this result is plausible, indeed even desirable, since it corresponds with the reality measured

Figure 4. FVG cartogram for cluster analysis (suboptimal solution for three groups) based on 14 social indicators and on average altimetry. (Source: our processing of FVG Regional Authorities, ISTAT, MIUR and ACI data).

by the indicator set. From an operational point of view, however, it is only the final user who can evaluate the necessity for a contiguity constraint.

4. Conclusions. This work concentrates on certain basic considerations regarding statistical exercises for territorial reorganisation that, working with a set of indicators, proceed to use the classical tools for multivariate statistical analysis. In particular, it was observed how, in the case of Friuli Venezia Giulia, using the Ward algorithm on principal (unrotated) components is a valid method, which allows one to obtain good territorial divisions, based on dimensionally homogeneous and well-balanced groups. Introducing a contiguity constraint to the clustering algorithm is a decision that should always be taken with the final user, as it implies introducing critical aspects to the analysis not directly considered by the indicator set. Should there be no wish to introduce such a constraint, it has been shown from an empiric point of view how the existence of favoured directions for a socioeconomic phenomenon can help give rise to sufficiently contiguous groups. This solution, however, is best adopted when the morphological configuration of the observed region is very unique, like that of FVG.

In the spirit of Prof. Strassoldo's thinking, such results are only a starting point, which – together with other qualitative information (including that obtained via *ad hoc* sampling surveys) – provide a solid basis for a more complete analysis of the observed phenomenon. Only in this way, may an analysis be properly considered a territorial one.

[1] In this work, consideration was not given to the construction of the indicators and indices. Please see Prof. Strassoldo's studies for further detail on this aspect (Strassoldo, Mattioli, Schifini 1996).

[2] The selected indicators referred to demographic aspects (including population density, birth rate, elderly to child rate, elderly index, presence of foreigners), unemployment, education (from nursery and primary school to university and non-university tertiary education), healthcare expenditure, beds for tourist accommodation, socio-political participation (from volunteering for non-profit organisations to participation in the 2011 referendum on water), and, finally, general mobility (considering the resident population that moves around every day for work and study, and the ACI (Italian Automobile Association) number of circulating cars).

[3] For the sake of completeness it should be mentioned that such a result is obtained by introducing, together with the 15 indicators already mentioned, a further 8 economic indicators, such as agricultural area utilised, those employed in agriculture, farms, livestock, those employed in industrial activities, industrial density, those employed in tertiary activities and those employed in bank branches. In this case, the mountain municipalities that are more disadvantaged from a socioeconomic point of view were identified (see Zaccomer 2017).

Bibliografie/ References

Arbia G., Espa G. (1996). *Statistica economica territoriale.* Padova: CEDAM.

Bonavero P. (2005). *L'approccio transcalare come prospettiva di analisi. Il contributo della geografia alla ricerca economica e sociale*. Quaderni dell'Istituto di Studi su Popolazione e Territorio, Serie metodologica, 3. Milano: I.S.U. Università Cattolica.

Cerioli A., Zani S. (2007). *Analisi dei dati e data mining per le decisioni aziendali*. Milano: Giuffrè.

Marra E. (1990). *Metodi per la riorganizzazione territoriale e cluster analysis*. In Guala C., Marra E. (Eds.) *Indicatori sociali e territorio*. Genova: Sagep, pp. 135-188.

Ripley B.D. (1981). *Spatial Statistics*. New York: J. Wiley and sons.

Romagnoli L. (2002). *Metodi statistici elementari per la geografia. Strumenti per l'analisi dei dati, il confronto quantitativo, lo studio e l'interpretazione delle relazioni spaziali con l'ausilio del calcolatore elettronico*. Bologna: Pàtron.

Sadocchi S. (1987). *Manuale di analisi statistica multivariata per le scienze sociali*. Milano: Angeli.

Strassoldo M., Mattioli E., Schifini S. (1996). *Teoria dei numeri indici dei prezzi e degli indicatori economici, finanziari e sociali*. Padova: CEDAM.

Zaccomer G.P. (1999). *Considerazioni metodologiche e caratteristiche strutturali dell'indagine*. In R. Grandinetti (Ed.) *Il seggiolaio e l'economia globale. La transizione evolutiva del distretto friulano della sedia attraverso i risultati di una indagine sul campo*. Padova: CEDAM, pp. 221-232.

Zaccomer G.P. (2006). *Shift-Share Analysis with Spatial Structure: an Application to Italian Industrial Districts*. Transition Studies Review, 42: 213-227.

Zaccomer G.P. (2008). *Economia, Statistica e Territorio. Informazione e metodologia statistica per la conoscenza dell'economia del Friuli Venezia Giulia*. Udine: Forum.

Zaccomer G.P. (2008). Spatial structure in a shift-share decomposition: new results for the Italian industrial districts case. *Transition Studies Review*, 48: 111-123.

Zaccomer G.P. (Ed.) (2011). *Carburanti, Statistiche e Prezzi. Esperienze di ricerca legate alla manovra di riduzione dei prezzi delle benzine e del gasolio per autotrazione in Friuli Venezia Giulia*. Udine: Forum.

Zaccomer G.P. (2017). *L'analisi territoriale socio-economica a supporto della predisposizione di un Piano Paesaggistico Regionale: il caso del Friuli Venezia Giulia*. In XXXII Congresso Geografico Italiano, *L'apporto della geografia tra rivoluzioni e riforme*, Programma e Abstract, Roma, p. 358. Gli interventi sono in corso di pubblicazione.

Zaccomer G.P., Grassetti L. (2014). La cartografia come strumento di interpretazione dei risultati di un modello di scomposizione spaziale: nuove proposte con applicazione al caso dell'occupazione in Friuli Venezia Giulia. *Bollettino dell'Associazione Italiana di Cartografia*, 152: 56-72.

Zaccomer G.P., Mason P. (2008). A definition of neighborhoods based on local labor systems: a regional application on employment data. *Statistica & Applicazioni*, 6: 215-236.

Zaccomer G.P., Mason P. (2011). A new spatial shift-share decomposition for the regional growth analysis: a local study of the employment based on Italian business statistical register. *Statistical Methods & Applications*, 3: 329-356.

Zaccomer G.P. Troiano S., Pagani L., Marangon F. (2013). *L'analisi dell'archivio ufficiale delle imprese agrituristiche presenti nel territorio del Friuli Venezia Giulia*. In Marangon F. , Troiano S. (Eds.) *Sviluppo economico locale e turismo sostenibile in Friuli Venezia Giulia*. Udine: Forum, pp. 113-127.

Zani S. (Ed.) (1993). *Metodi statistici per le analisi territoriali*. Milano: Angeli.