AN APPROACH FOR DETERMINING AREAS OF SOCIO-ECONOMIC DEPRIVATION AT SUB-MUNICIPAL LEVEL

Marco Ballin, Gianpiero Bianchi, Giancarlo Carbonetti, Paolo Lorusso

Abstract. Municipalities have a growing need for detailed and interpretable data to implement specific measures to reduce socio-economic exclusion and deprivation while improving urban quality and promoting the economic, social, and cultural development of urban areas of the resident population. Traditional administrative boundaries may not represent the true spatial distribution of socio-economic phenomena, limiting the effectiveness of such interventions.

This study proposes a parameterized procedure for aggregating small, contiguous spatial units, such as census enumeration areas, into larger, homogeneous domains. With highly detailed demographic, social and economic data, this approach provides in-depth understanding of complex urban issues. The approach is applied to a case study focusing on socio-economic deprivation at sub-municipal level, which considers multiple dimensions such as economic distress, employment instability and low educational attainment. These factors contribute in different ways to the social exclusion of individuals.

The analysis uses results from the population and housing census and data from administrative registers collected by Italian National Institute of Statistics (Istat). The different aspects are synthesized into a single metric using the composite AMPI index at the urban enumeration area level. The results highlight the potential of data-driven spatial aggregation to support effective urban planning and decision-making, revealing patterns that go beyond conventional administrative partitions.

1. Introduction

The availability of data and indicators on individuals and households referring to the minimum territorial unit, coinciding with the census Enumeration Areas (*EAs*), has in the past offered the possibility of determining critical areas within municipalities - aggregations of *EAs* - according to criteria of statistical homogeneity and spatial contiguity.

These areas were identified by the municipalities for specific local policy planning objectives or to access funding provided by the central government to contrast social exclusion, unemployment of the weakest groups (young people, women), to favour local entrepreneurship with various forms of tax exemptions.

Some examples are the "*Zone Franche Urbane*" or the "*Aree urbane degradate*" defined through specific regulatory measures[1] and defined on a predefined set of indicators based on the results of the *EA*-level census. The design of such areas was often carried out heuristically from the most critical *EAs* by defining a spatial cluster according to specific size limits in terms of resident population.

Several well-established algorithmic frameworks exist for aggregating contiguous spatial units into internally homogeneous regions. SKATER, Spatial 'K'luster Analysis by Tree Edge Removal, (Assunção *et al.*, 2006) builds a minimum spanning tree across spatial units connected by adjacency, weighted by attribute dissimilarity, and partitions the tree by pruning edges to yield contiguous, homogeneous regions. Another notable approach that can be used is the Max-p regionalization algorithm (Duque *et al.*, 2012). Max-p determines the largest possible number of regions under constraints, such as minimum size and contiguity, by optimizing internal homogeneity using heuristic or mixed integer strategies. More recently, spatially constrained spectral clustering has emerged, embedding spatial contiguity within a similarity graph, and applying spectral partitioning to balance spatial and attribute coherence. Yuan (Yuan *et al.*, 2015) propose a version using a truncated exponential kernel and recursive partitioning to produce nested coherent regions.

With this study it is proposed an intuitive and parameterized procedure for aggregating *EAs*, into larger, homogeneous domains. With highly detailed demographic, social and economic data, this approach provides in-depth understanding of complex urban issues.

The paper is organized as follows: section 2 presents the Istat project on socio-economic deprivation at the sub-municipal level; section 3 illustrates the algorithm developed for the identification of critical areas; section 4 presents some results for the municipality of Bologna.

2. The study of socio-economic deprivation at sub-municipal level

Istat, in collaboration with several municipalities, is conducting a project to study socio-economic deprivation of population at sub-municipal level (Carbonetti *et al.*, 2025). This project makes use of the data available from the permanent population census, administrative sources and thematic registers developed by Istat in recent years, as well as the ability to integrate these sources and geocode information down

[1] "*Zone Franche Urbane" – Legge 27 dicembre 2006, n. 296 (Finanziaria 2007), art. 1 commi 340-343). "Aree urbane degradate" – Decreto del Presidente del Consiglio dei Ministri del 15 ottobre 2015.*

to the census *EAs*. The main objective is to measure, represent and analyse the phenomenon of socio-economic deprivation of individuals and households at sub-municipal level. This is a multidimensional phenomenon defined as "*a condition in which households and individuals experience difficulties in adequately meeting their basic needs due to a lack of or insufficient economic, employment, educational, and social resources and opportunities*".

Based on the availability of data, nine statistical ratio indicators (variables) covering different components of deprivation (economic, occupational, and educational) were defined and calculated at *EA* level:

▪ *Economic deprivation*: % of individuals aged 70 and over living alone and not owning a home (ECO1); % of individuals in households in which no member is employed or receiving a pension from work (ECO2); % of individuals in low equivalized income households (ECO3).

▪ *Occupational deprivation*: Employment rate 25-64 years old (OCC1); % of individuals aged 0-64 living in households with very low work intensity (OCC2); % of employed persons aged 25-64 "not stable" during the year (OCC3).

▪ *Educational deprivation*: % of individuals aged 25-64 without upper secondary school education (EDU1); % of individuals aged 15-29 who are not employed and are not enrolled in any regular course of study (EDU2); % of students dropping out or repeating the year (EDU3).

Table 1 describes the numerators and denominators of the nine aggregate variables and their sources.

Subsequently the nine statistical ratio indicators (e.g. OCC1=Employed persons aged 25-64/Individuals aged 25-64) have been synthesized, through the Adjusted Mazziotta-Pareto Index (AMPI^{+}) methodology (Mazziotta e Pareto, 2016; Mazziotta e Pareto, 2017), in a composite index at the *EA*-level. Such index, the Socio-Economic Deprivation Index of population (*SED-I*), measures household and individual deprivation as a tangible condition of deprivation, as distinct from simple exposure to risk[2]. In summary, *SED-I* is a partially non-compensatory composite index based on a standardization of the individual indicators, at the reference time, that makes the indicators independent from the unit of measure (De Muro *et al.*, 2011). Details about AMPI^{+} methodology and *SED-I* computation are given in paragraph 3.1.

[2] Istat already produces an indicator which measures social and material vulnerability of population. The Social and Material Vulnerability Index measures the exposure of some population groups to situations of risk, such as uncertainty of their social and economic condition (Istat, 2020).

Table 1 – *Numerators and denominators of the nine statistical ratio indicators related to socio-economic deprivation.*

Variables	Numerator	Denominator	Sources
ECO1	Individuals aged 70 and over living alone and not owning a home	Individuals aged 70 and over	PPHC Cadaster Register
ECO2	Individuals in households in which no member is employed or receiving a pension from work	Individuals in households	PPHC Pensioners' records (INPS)
ECO3	Individuals in low equivalized income households	Individuals in households	PPHC Income Register
OCC1	Employed persons aged 25-64	Individuals aged 25-64	PPHC
OCC2	Individuals aged 0-64 living in households with very low work intensity	Individuals aged 0-64 living in households	PPHC Labour Register
OCC3	Employed persons aged 25-64 "not stable" during the year	Individuals aged 25–64 living in households with a sign of employment during the year	PPHC Labour Register
EDU1	Individuals aged 25-64 without upper secondary school education	Individuals aged 25-64	PPHC
EDU2	Individuals aged 15-29 who are not working and are not registered in any regular course of study at MIM or MUR	Individuals aged 15-29	PPHC MIM – MUR
EDU3	Students dropping out or repeating the year	Students with attendance records in the academic year t/t+1	PPHC Education Register

Sources: Istat, Permanent Population and Housing Census (PPHC); Income register, Labour register and Education register; National Social Security Institute (INPS); Ministry of Education (MIM); Ministry of University and Research (MUR); Register of dwellings and buildings (Cadastre). Reference year: 2021.

All measures (indicators and *SED-I*) are calculated at *EAs*, but analyses will be conducted for two levels of *EA* "aggregations" to suggest an integrated reading of the results:

- *SMA* – administrative Sub-Municipal Areas. Areas defined by municipalities for functional, administrative, or statistical purposes.
- *ADU* – Areas of socio-economic deprivation in urban contexts. Areas never previously identified, designed by Istat according to the experimental methodology proposed in this document to identify concentrations of deprivation within municipal boundaries.

The results of the deprivation study, at *SMA* and *ADU* level, represent a powerful knowledge and decision-making tool for municipal administrations. The choice of the level of analysis and intervention depends on political choices and economic availability, factors that may vary between different municipalities and between different administrations within the same municipality.

3. The *SED-I* computation and *ADU* identification procedure

The method for identification of *ADU* is based on a spatial aggregation procedure designed to identify and group contiguous territorial units with similar characteristics. The procedure is designed to be highly flexible and adaptable to different spatial or policy needs through a set of configurable parameters:

- P_{min} , P_{max}: minimum and maximum population per *ADU*. If the total population does not exceed the minimum threshold, the aggregation is discarded; when the maximum threshold is reached, the algorithm stops the aggregation procedure.
- P_{tot}: global population threshold. It defines the amount of population to be included in the whole set of *ADU*.
- Min_{EA}: minimum number of *EAs*. Minimum number of *EAs* to be included in a *ADU*.
- ε: *SED-I* homogeneity threshold. It sets the maximum deviation allowed from the area under construction and an additional territorial unit to be included.
- τ: minimum *SED-I* threshold (*SED-I* floor). Minimum composite indicator value that an aggregation must maintain to be validated. This constraint prevents the inclusion of less deprived sections from compromising the significance of the aggregations.

Since the procedure is heuristic and data-driven, different combinations of parameters can lead to different solutions, all of which are methodologically legitimate but have different operational implications, for instance:

- smaller values on homogeneity ε favor the construction of *ADU* more coherent with respect to *SED-I*, but geographically more fragmented;
- low demographic thresholds favor defining micro-areas of intervention, high thresholds lend themselves to larger-scale planning;
- a high *SED-I* threshold identifies only the most impaired areas, while a lower one allows more extensive coverage, including situations of moderate distress.

3.1. SED-I computation

Let the *EAs* be indexed by $i=1,..,N$ where *N* indicates the amount of *EA* in a given municipality. Each *EA* is characterized by *K* ($K=9$ in the case of this paper; see Table 1) statistical ratio indicators $R_i^{(k)}$ (i=1,..,N ; $k=1,..,K$). It is assumed that for each *EA* the population, $Pop(EA_i)$, and all numerator and denominator pairs ($NUM_i^{(k)}$; $DEN_i^{(k)}$) necessary for the *K* indicators computation are available.

The k^{th} indicator ratio in the i^{th} *EA* is defined as:

$$R_i^{(k)} = \frac{NUM_i^{(k)}}{DEN_i^{(k)}} \text{ assuming } \frac{0}{0} = 0 .$$

In order to compute the *SED-I* composite index, the first step consist in the "normalization" of the $R_i^{(k)}$. Normalization is computed by the following:

$$Z_i^{(k)} = 70 + 60 * \left(\frac{R_i^{(k)} - a_k}{b_k - a_k}\right) \qquad \text{(if } a_k \neq b_k\text{, 70 otherwise)}$$

where: $a_k = \mu_k - \frac{1}{2} R_{max}^{(k)}$, $b_k = \mu_k + \frac{1}{2} R_{max}^{(k)}$ (a_k and b_k are called ''goalposts''), $\mu_k = \frac{\sum_{i \in S} \mathrm{NUM}_i^{(k)}}{\sum_{i \in S} \mathrm{DEN}_i^{(k)}}$ and $R_{max}^{(k)} = \max_{i \in 1,..,N} R_i^{(k)}$.

If an indicator has negative polarity, it is transformed as $Z_i^{(k)} = 200 - Z_i^{(k)}$ (as in the case of indicator OCC1).

The *SED-I* index for i^{th} *EA* is defined as:

$$SED\text{-}I_i = \bar{Z}_i + \sigma_i \mathrm{CV}_i$$

Where $\bar{Z}_i = \frac{\sum_{k \in K} Z_i^{(k)}}{k}$, $\sigma_i = \sqrt{\frac{\sum_{k \in K} (Z_i^{(k)} - \bar{Z}_i)^2}{k}}$ and $\mathrm{CV}_i = \frac{\sigma_i}{\bar{Z}_i}$.

3.2. Algorithm description

The aggregation algorithm performs an iterative greedy region-growing process. At the very beginning the algorithm starts considering the whole set of *EAs*. The general iteration performed for the construction of an *ADU* can be described as follow:

1. Let *U* be the set of all territorial units not yet belonging to an *ADU*. Let $U^{(max)}$ be the unit with maximal *SED-I* in *U*;
2. indicate by $C\left(U^{(max)}\right)$ the set of units candidate to be aggregated to $U^{(max)}$. $C\left(U^{(max)}\right)$ is defined as the set all *EAs* adjacent to $U^{(max)}$ belonging to *U* and with $SED\text{-}I \geq (1 - \varepsilon) * SED\text{-}I\left(U^{(max)}\right)$. Let $C^{(u)}$ be an element of $C\left(U^{(max)}\right)$;
3. form aggregated areas ($AG^{(u)}$) by collapsing $U^{(max)}$ with each $C^{(u)}$;
4. for each $AG^{(u)}$ compute the total population, *Pop*($AG^{(u)}$), and all pairs $\left(NUM_u^{(k)}; DEN_u^{(k)}\right)$ by adding the correspondent elements of $U^{(max)}$ and $C^{(u)}$. Using such quantities compute $SED\text{-}I^{(AG)}$ for each $AG^{(u)}$ following the procedure described in paragraph 3.1;

5. carry out the following checks on each $AG^{(u)}$:
   a. $Pop(AG^{(u)}) \leq P_{max}$ (size of $AG^{(u)}$ can't be over a fixed threshold);
   b. $SED\text{-}I^{(AG)} \geq \tau$, (*SED-I* minimum threshold for a *ADU*);
6. among the $AG^{(u)}$ that meet the two conditions, choose the one with the maximum *SED-I*, let $AG^{(max)}$ be such elements;
7. create the set $C\left(AG^{(max)}\right)$ as described in step 2.

Repeat the steps 3-7, using $AG^{(max)}$ instead of $U^{(max)}$, until at least a new $AG^{(u)}$ meets the two conditions of step 5 and until $C\left(AG^{(max)}\right)$ is not empty.

If after last iteration of steps 3-7 the last $AG^{(max)}$ satisfies the following conditions:

- $Pop\left(AG^{(max)}\right) \geq P_{min}$;
- Number of *EAs* included in $AG^{(max)} \geq Min_{EA}$;

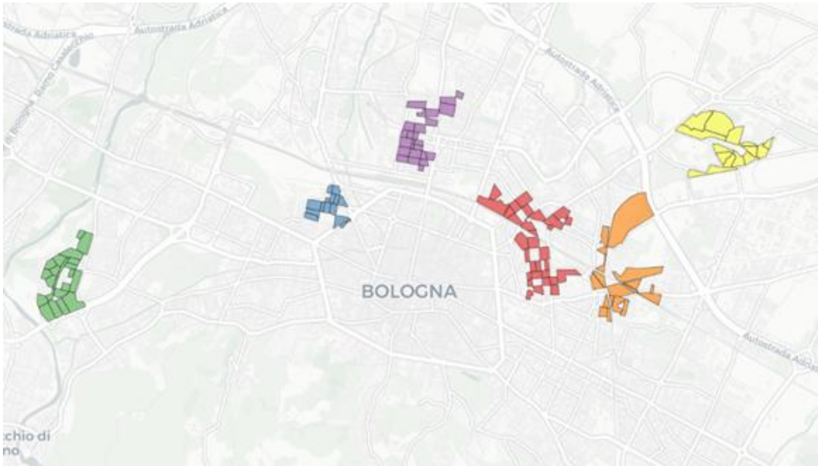
then $AG^{(max)}$ is a new *ADU*. If $AG^{(max)}$ is accepted as a new *ADU* delete all *EAs* forming such new area from *U* and repeat the algorithm from step 1.

If at least one of the last two conditions is not met, $AG^{(max)}$ is not accepted as a new *ADU*. In such case the procedure re-starts from the following *EA* with the greatest *SED-I* in *U*. If adding the population of previous *ADU* and $AG^{(max)}$ is greater than P_{tot} the algorithm, stops. Furthermore, the algorithm stops when all starting point have been tested and no additional *ADU* can be created.

4. Results

To demonstrate the effectiveness of the proposed procedure and to highlight the critical role of parameter selection, we conducted a series of simulations aimed at identifying aggregations characterized by significant socio-economic disease patterns in the city of Bologna. In the first simulation, the parameters were defined as follows: the homogeneity tolerance range (ε) was set to 0.1; the global population P_{tot} threshold was established at 38,000, corresponding to approximately 10% of the total population in the census *EAs* of the inhabited center; each aggregation was required to have a population between P_{min}=3,000 and P_{max}=7,500; the *SED-I* floor was set to 100 plus the mean squared error of the *SED-I* distribution across *EAs*, resulting in a threshold value of 103.69 (Figure 1); Min_{EA} has been set to 5.

Figure 1 – *Bologna: SED-I floor=103.69; Population 3,000 – 7,500 (Istat, 2021).*



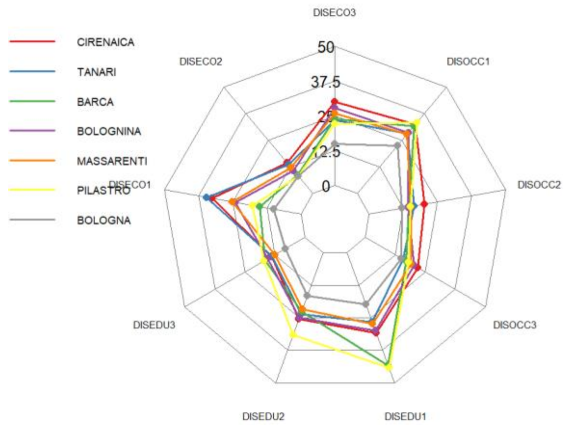
Six critical aggregations were identified, roughly corresponding to the neighborhoods of the following disctricts of Bologna: "Cirenaica", "Tanari", "Barca", "Bolognina", "Massarenti", and "Pilastro". A summary (Table 2) of these aggregations (*ADUs*) is provided: population size (POP), final *SED-I* value (*SED-I*), starting section *SED-I* (START *SED-I*), total number of sections (N.of *EAs*) included, and the number of sections with an *SED-I* value above 110 and 105, respectively.

Table 2 – *Bologna: SED-I floor=103.69; Population 3,000 – 7,500. ADU's summary (Istat, 2021).*

ADU	DISTRICT	POP	SED-I	START SED-I	N. of EAs	N. of EAs SED-I>110	N. of EAs SED-I>105
1	Cirenaica	7,323	106.40	116.09	31	8	17
2	Tanari	3,028	104.70	113.77	15	2	3
3	Barca	4,438	104.55	112.51	21	1	8
4	Bolognina	7,413	104.67	112.42	24	3	10
5	Massarenti	3,121	104.05	112.30	18	1	4
6	Pilastro	5,484	105.52	108.79	14	0	9

In the "radar chart" (Figure 2), we compare how the nine indicators of the six *ADUs* behave relative to the overall values for the municipality of Bologna.

Figure 2 – *Bologna: SED-I floor=103.69; Population 3,000 – 7,500. ADU's indicators (Istat, 2021).*



As expected, all the aggregations show higher values than the city average. The most critical issues emerge in terms of EDU1 and EDU2, where the "Pilastro" area stands out as the most critical, also recording the highest value for EDU3. The "Barca" area follows closely in terms of EDU1, while for ECO1, both "Cirenaica" and "Tanari" show values nearly four times higher than the city average.

To assess the flexibility of the procedure in terms of its applicability to different contexts, the analysis has been repeated using as population thresholds P_{min}=1,000 and P_{max}=5,000 residents (results are shown in Figure 3 and Table 3).

Figure 3 – *Bologna: SED-I floor=103.69; Population 1,000 – 5,000 (Istat, 2021).*

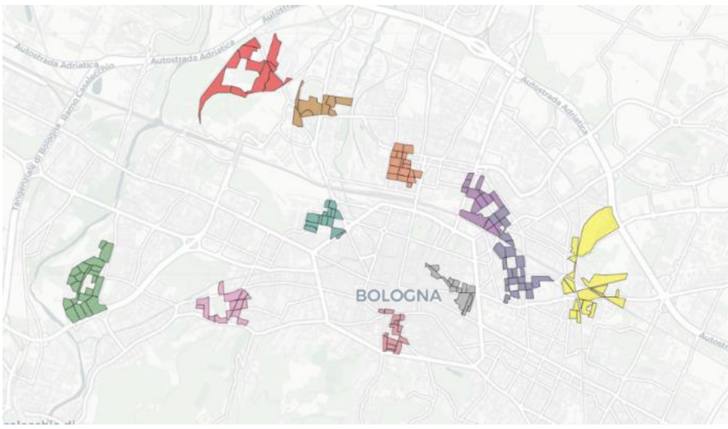
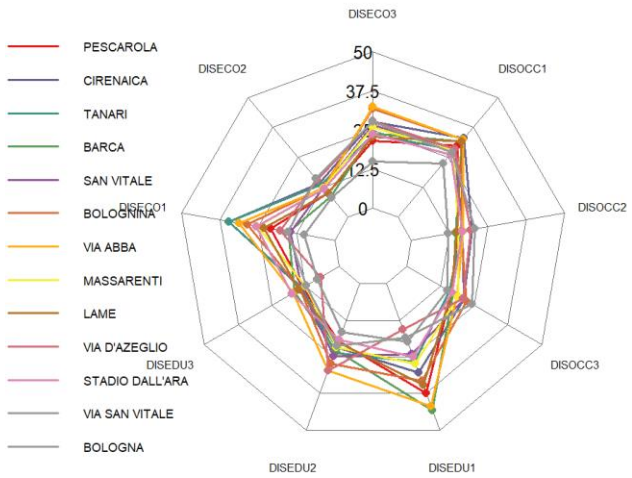


Table 3 – *Bologna: SED-I floor=103.69; Population 1,000 – 5,000. ADU's summary (Istat, 2021).*

ADU	DISTRICT	POP	SED-I	START SED-I	N. of EAs	N. of EAs SED-I>110	N. of EAs SED-I>105
1	Pescarola	2,671	103.89	118.55	10	1	3
2	Cirenaica	4,637	106.06	116.09	22	7	10
3	Tanari	3,028	104.70	113.77	15	2	3
4	Barca	4,438	104.55	112.51	21	1	8
5	San Vitale	3,168	103.70	112.46	14	2	5
6	Bolognina	4,772	106.19	112.42	17	3	10
7	Via Abba	1,690	107.17	112.34	10	2	7
8	Massarenti	3,121	104.05	112.30	18	1	4
9	Lame	1,618	104.18	111.74	7	1	2
10	Via d'Azeglio	1,558	103.75	110.65	15	1	4
11	Stadio Dall'Ara	2,722	103.75	110.17	14	2	6
12	Via San Vitale	2,344	103.75	110.05	13	2	4

In the "radar chart" (Figure 4), the values of the nine indicators for each of the twelve *ADUs* are compared to the overall values for Bologna. Interestingly, the aggregation "Via d'Azeglio" shows lower values than the city average for indicators EDU1 and EDU3, despite exhibiting significantly higher-than-average values for the other seven indicators. It records the highest value for indicator EDU2, highlighting a specific critical issue among the 15–29 age group.

Figure 4 – *Bologna: SED-I floor=103.69; Population 1,000 – 5,000. ADU's indicators (Istat, 2021).*



5. Conclusions

The paper describes a methodology and proposes a procedure for its implementation to identify sub-municipal areas that complement the minimum territorial units coinciding with the census enumeration areas (*EAs*). These new areas are created by aggregating contiguous and homogeneous *EAs*.

In the case analysed in this study, homogeneity is assessed with respect to a composite index of socio-economic deprivation (*SED-I*) developed by Istat, based on nine elementary indicators, and reflecting the level of economic, employment, educational, and social deprivation of the population residing in a specific area.

The area construction methodology is sequential. The procedure begins by selecting the *EA* with the highest observed socio-economic deprivation and continues by aggregating neighbouring homogeneous *EAs* around it. The area construction proceeds if there are contiguous *EAs* sufficiently homogeneous with the area already formed, or if it does not exceed predetermined size thresholds. Once the first area is completed, the process resumes by excluding already assigned *EAs* from further analysis. The process stops when there are no more *EAs* that meet the eligibility criteria for aggregation, or when the aggregate of constructed areas exceeds a predetermined fraction of the municipal population.

The objective is to identify the areas of the municipal territory where the highest concentrations of socio-economic deprivation of the population are observed.

One of the strengths of the procedure lies in the intuitiveness of the required parameterisations: the minimum degree of homogeneity; the minimum level of overall deprivation of each area; the minimum and maximum size of each area; the maximum fraction of municipal population that can be included in the set of critical areas. In addition to describing the procedure in detail, the document presents some examples for the city of Bologna.

Future developments will focus on the identification of a set of parameterisations applicable to groups of municipalities homogeneous with respect to their demographic size, as well as on the exploration of potential alternative forms of parameterisation of the procedure.

An important aspect will be the definition of evaluation criteria for the resulting areas. Currently, the validation of the results is conducted together with the municipalities, which are the final users of the results. The identification of these areas will in fact offer the possibility to define targeted policy actions on these areas, also considering the socio-demographic characteristics of the resident individuals and households.

References

ASSUNCÃO R. M., NEVES M. C., CÂMARA G., DA COSTA FREITAS C. 2006. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. I*nternational Journal of Geographical Information Science*, Vol. 20, No. 7, pp. 797-811.

CARBONETTI G., BIASCIUCCI F., CUTILLO A., MAZZIOTTA M., QUONDAMSTEFANO V., TAMBURRANO M. T., TRONU. D. 2025. Measuring socio-economic deprivation at sub-municipal level through the integration of census and administrative data. *Italian Journal of Economic, Demographic and Statistical Studies – RIEDS*, Vol. LXXIX No.1 Gennaio-Marzo 2025.

DE MURO P., MAZZIOTTA M., PARETO A. 2011. Composite Indices of Development and Poverty: An Application to MDGs. *Social Indicators Research*, Vol. 104, No. 1, pp.1-18.

DUQUE J. C., ANSELIN L., REY S. J. 2012. The max-p-regions problem. *Journal of Regional Science*, Vol.52, No. 3, pp. 397-419.

MAZZIOTTA M., PARETO A. 2017. Synthesis of indicators: the composite indicators approach. In: Maggino F. (Ed.) *Complexity in Society: From Indicators Construction to their Synthesis*. *Social Indicators Research Series*, Springer, pp.159-191.

MAZZIOTTA M., PARETO A. 2016. On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena. *Social Indicators Research*, Vol. 127, No. 3, pp. 983-1003.

YUAN S., TAN P. N., CHERUVELIL K. S., COLLINS S. M., SORANNO P. A. 2015. Constrained spectral clustering for regionalization: Exploring the trade-off between spatial contiguity and landscape homogeneity. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1-10). IEEE.

Marco BALLIN, Italian National Institute of Statistics, ballin@istat.it
Gianpiero BIANCHI, Italian National Institute of Statistics, gianbia@istat.it
Giancarlo CARBONETTI, Italian National Institute of Statistics, carbonet@istat.it
Paolo LORUSSO, Italian National Institute of Statistics, paolo.lorusso@istat.it