REBUILDING A PSEUDO POPULATION REGISTER FOR ESTIMATING PHYSICAL VULNERABILITY AT THE LOCAL LEVEL: A CASE STUDY OF SPATIAL MICRO-SIMULATION IN SONDRIO[1]

Alberto Vitalini, Simona Ballabio, Flavio Verrecchia

Abstract. A wide range of user groups, ranging from policy makers to media commentators, is increasingly seeking more detailed spatial information on health-related topics. This information is needed to gain a better understanding of their communities, more effectively allocate resources, and plan activities and interventions in a more efficient manner. However, due to the sensitivity of the topic of health, it can be challenging to obtain detailed or significant local data. To meet this need, small area estimation (SAE) methodologies are popular as a means of providing spatially detailed insights. Among the various SAE methodologies available, static spatial microsimulation has enabled the simulation of previously unknown variables, such as physical vulnerability, smoking, alcohol consumption, and obesity at the municipal level. This paper presents the initial results of application of static spatial simulation, in order to create synthetic population dataset and estimate "physical vulnerability" of elderly in the municipalities in the province of Sondrio. Physical vulnerability is measured by the prevalence of people who report suffering from chronic or long-term illnesses and who have limitations in their daily activities. A combinatorial optimization (CO) algorithm called simulated annealing, developed at the University of Leeds, is used to simulate the distributions of the "physical vulnerability". This algorithm combines public microdata from the Multiscopo Survey-Aspects of Daily Life-2021 and data from the 2021 Permanent Population Census, which are disseminated in Istat's public databases.

1. Background

With the aging population, the number of elderly individuals affected by chronic illnesses and disabilities is increasing, posing challenges for healthcare systems and policymakers in providing effective care and assistance. It is projected that the aging population will significantly increase in the coming years, putting pressure on

[1] The work is the joint responsibility of the authors. Paragraph 1 is attributed to Flavio Verrecchia, paragraph 2 to Alberto Vitalini, paragraphs 3 and 4 are attributed to Simona Ballabio.

healthcare systems and leading to rising healthcare costs. For example, in Lombardy, it is estimated that the number of individuals aged 65 and older will reach 2.9 million by 2070, surpassing the figure by over half a million compared to 2023 (Istat, 2023).

Identifying older people at risk of physical decline is important as it allows healthcare professionals and policymakers to prioritise care and support for those in greatest need. Frail elderly individuals are more likely to experience negative health outcomes such as falls, hospitalisations, and disabilities (Clegg et al., 2013). In this context, having data on physical vulnerability can help better allocate resources and develop more targeted interventions. For instance, if a particular area exhibits a higher prevalence of physical vulnerability among the elderly, policymakers can allocate more resources to address the issue in that area.

The need for health information at both individual and community levels is crucial, but unfortunately, there is a chronic lack of available data. Official statistics from ISTAT, the Italian National Institute of Statistics, are limited due to data protection regulations that require statistical units to be non-identifiable for data release. This means that information on the entire population at the municipal level is extremely limited, especially in the health domain. Even the sample surveys conducted by ISTAT as part of official statistics are not conclusive for obtaining municipal level information. These surveys are designed to provide reliable information at the national, regional, and geographic levels but may not be sufficient for local objectives. For example, if we wanted to estimate the number of physically frail elderly individuals at the municipal level, we could only calculate a value for the municipalities where there are survey respondents, and the sample size would likely be too small to provide accurate information. To overcome this problem, policymakers could invest in data collection and analysis, but resources and capacity to gather and analyse accurate and reliable data from many local structures -such as small municipalities- are limited.

2. Methods and Data

Considering the limitations of available data sources, it is necessary to explore and evaluate alternative solutions to obtain the required information. Among these solutions, methods of "small area estimation" (SAE) are popular in providing detailed information on specific areas. Small area estimation methodologies are statistical techniques used to estimate parameters in areas where the sample size of a survey is too small for reliable estimation and/or where data have not been collected (Asian Development Bank, 2020).

Traditionally there are two types of small area estimation – direct and indirect estimation. Direct small area estimation is based on survey design and includes three estimators called the Horvitz-Thompson estimator, GREG estimator and modified direct estimator (Asian Development Bank, 2020). On the other hand, indirect approaches of small area estimation can be divided into two classes – statistical (Rao, 2003) and geographic approaches (Rahman *et al.*, 2010). Within the geographic approach, spatial microsimulation models (SMMs) have been widely applied on health outcomes and behaviours in populations (Smith *et al.*, 2021).

This paper focuses on the potential of the spatial microsimulation approach, demonstrating its application in estimating the number of physically vulnerable elderly individuals in the municipalities of the province of Sondrio.

There are two types of spatial microsimulation: static and dynamic. In practice, while a static microsimulation model provides an estimated population by synthesizing data at a specific point in time, a dynamic microsimulation model incorporates the ability to model changes and transitions over time, allowing the population to age and evolve within the simulation (Tanton, 2014). In this paper spatial static microsimulation is used.

The spatial microsimulation takes in two types of data:

1. Statistics on small areas, such as aggregated census tables for each municipality in a region, provided by ISTAT . These tables were constructed based on the data from the 2021 Permanent Census of the Population (Istat, 2021). Specifically, two tables were used: the first one containing the distribution of residents in two age groups (65 to 74 years old and 75 years and older) in each municipality of the province of Sondrio, by sex and citizenship (Italian or foreign/stateless), and the second one containing the distribution of residents aged 65 and older in each municipality of the province of Sondrio, by educational attainment in four categories: no education or elementary school, middle school diploma, high school diploma, and university degree or postgraduate degree.
2. Microdata from a sample survey. In this paper, publicly available microdata from the "Multipurpose Survey on Households: Aspects of Daily Life" (2021 edition), specifically related to the geographic distribution of Northwestern Italy, were used. It should be noted that these data do not include the ISTAT code for the municipality or province of residence of the interviewee, only the region code.

Spatial microsimulation utilizes the microdata from Multipurpose Survey on Households as a container of "donor records" for constructing a micro-population in each individual municipality, with gender, age, educational attainment, and citizenship characteristics that allow for the most accurate reproduction of the distribution of residents in the cells of the census tables.

Within the family of static spatial microsimulation techniques, three types of alternative algorithms dominate the literature) and allow for the creation of simulated spatial microdata: iterative proportional fitting (IPF), combinatorial optimization (CO), and generalized regression reweighting (GREGWT) (O'Donoghue *et al.*, 2014).

In this paper, the combinatorial optimization (CO) algorithm called simulated annealing, implemented in the Flexible Modelling Framework (FMF) software developed at the University of Leeds (Harland *et al*., 2012; Harland, 2013), is used for spatial microsimulation.

2.1. The Six Steps of Spatial Microsimulation

The complete process of microsimulation can be divided into six conceptual steps:

1. Definition of small areas.
2. Definition of variables to be estimated.
3. Definition of "constraining" variables.
4. Construction of the microsimulation model.
5. Calculation of the percentage of people within a municipality that fall into the category of the variable of interest.
6. Evaluation of the results.

In the first step (point 1), the municipalities of the province of Sondrio are considered as small areas, and the population aged 65 and older is the target population for the study.

In the second step (point 2), the concept of "physical vulnerability" is defined. Physical vulnerability is a complex and multifactorial concept, which makes its accurate measurement difficult. For example, physical vulnerability can be influenced by factors such as age, chronic conditions, cognitive decline, and social support, which can be challenging to measure comprehensively (Rockwood *et al.*, 2005).

In this study, based on the publicly available sample information from the "Multipurpose Survey on Households: Aspects of Daily Life," an attempt was made to create a measure of "physical vulnerability" that is easy to calculate, easy to understand, and at the same time, meets the needs of potential users (in our case, local administrators) and is comparable at the territorial level. The proposed measure allows for identifying individuals at higher risk of physical vulnerability based on their self-reported health status and limitations in activities due to health problems and was derived by combining the responses of two survey questions formulated as follows: "How is your overall health?" and "To what extent do you have limitations in your usual activities lasting at least 6 months due to health problems?" The responses to these questions are converted into a binary variable, where a value of

“1” indicates that the person is physically vulnerable, and a value of “0” indicates that the person is not physically vulnerable. Specifically, a person is considered physically vulnerable if they report their health as "poor" or "very poor" and, at the same time, report limitations in their usual activities due to health problems. In all other cases, they are considered not vulnerable.

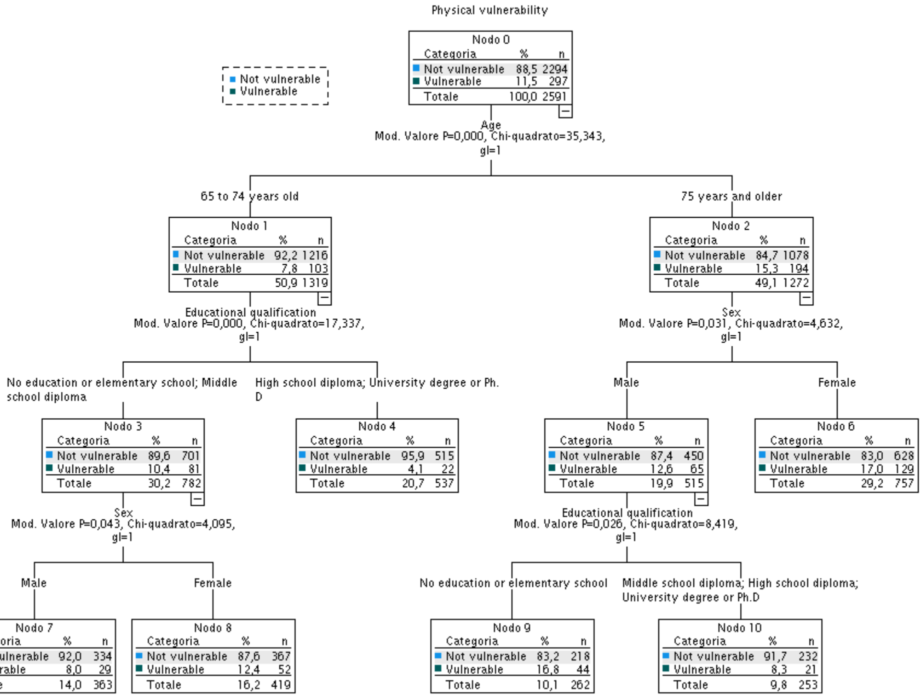
In the third step (point 3), the "constraining" variables considered are age, gender, educational attainment, and citizenship.

In the context of spatial microsimulation, from a technical standpoint, the categories of the "constraining" variables must be the same in both the tables derived from census data and the sample microdata. From a substantive standpoint, the selection of "constraining" variables should be guided by their capacity to accurately explain variability in physical vulnerability. In other words, when choosing these variables, it is important to pick ones that are strongly related with physical vulnerability. This ensures that the spatial microsimulation model will be effective in capturing and representing the real-world variations in vulnerability.

The choice of age, gender, and educational attainment variables fully satisfies this requirement as they provide valuable insights into the physical vulnerability of the elderly and are well supported by the literature (Galluzzo *et al*., 2018; European Institute for Gender Equality, 2021; Petrelli *et al.*, 2019). Age is an established predictive factor, as physical decline and chronic health conditions increase with advancing age. Gender is relevant as women tend to live longer but are more susceptible to chronic health conditions and disabilities. Higher levels of education are associated with better health outcomes and reduced risk of disability and chronic health conditions.

Preliminary analyses, based on classification trees, of the microdata from the Multipurpose Survey reveal a significant predictive capacity of these variables for the physical vulnerability status of individuals (Figure 1) consistent with the literature.

Figure 1 - *Decision tree with the target variable "Physical Vulnerability" and predictor variables "Age," "Educational Attainment," and "Gender".*



Source: Multipurpose Survey on Households: Aspects of Daily Life, ed. 2021 - Northwestern Italy.

Simulated annealing (point 4) is stochastic computational technique for finding near globally-minimum-cost solutions to large optimisation problems: in our case, to select a configuration of microdata in each municipality that closely reproduces the census tables constructed from the Istat census data used in the process. For a mathematical description of the algorithm, please refer to Appendix A and D in Harland et al. (2012).

The functioning of the algorithm can be, however, clearly explained in a more intuitive way. The algorithm starts by randomly selecting a certain number of individuals, which we can refer to as "donors," from the sample survey data and placing them in a specific municipality until reaching the correct number of residents for that municipality (provided by the Census). The procedure is repeated for all municipalities. At this point, the tables obtained for each municipality will not be identical to the Istat source tables, except for the total number of people.

Simultaneously, an error measure is created that compares the values in the cells of the Istat census tables with those obtained from the simulated data. The error measure used in this study is the Total Absolute Error (TAE), which is the sum of the absolute differences between the simulated and actual values in each cell. The TAE[2] calculates the number of people in the population that have been misclassified.

The simulated annealing algorithm works by exchanging individuals between the simulated population and the sample of individuals and checking if this exchange improves the error measure. If so, the algorithm keeps the exchange; otherwise, it cancels it and looks for another person in the sample to perform the exchange. The process continues until successive modifications no longer improve the error measure or until the number of cycles set by the analyst is exhausted.

When comparing a single variable, such as gender, the process is relatively simple. However, when there are multiple variables to compare, replacing one individual may improve the distribution fit of one variable to the census data but worsen it for another. Additionally, there may be no suitable individual in the sample for replacement.

Spatial microsimulation allows estimating variables in the simulated micropopulation that are not present in the Istat census tables but are present in the sample of individuals (point 5). By taking an individual from the sample and including them in the simulated population, the algorithm also "carries over" the associated non-constraining variables, referred to as additive variables. For example, whether the individual is physically vulnerable or not. At the end of the simulation, it is sufficient to count how many records exist in each municipality with the additional variable "physically vulnerable" equal to 1. This operation allows calculating the number of vulnerable people and the rates of physically vulnerable elderly individuals for the two age classes: 65-74 and 75 years and above. The absolute estimates are not definitive since, implicitly using the data of the resident population from the census, we do not consider the elderly individuals living in care facilities such as nursing homes and therefore overestimate the number of physically vulnerable residents. To control for this source of distortion, we use census information on the "number of disabled adults and elderly residents in care facilities - nursing homes" (Istat, 2021), Resident population in cohabitation by type of cohabitation and sex - Lombardy). We subtract the number of disabled adults and elderly residents in care facilities - nursing homes from the total number of elderly individuals and apply the previously calculated rates to obtain the definitive estimates of the absolute number of physically vulnerable elderly individuals living in households. For this purpose, we calculate the vulnerability rates within each municipality.

[2] $TAE = \sum_i \sum_j |O_{ij} - E_{ij}|$. Where O_{ij} and E_{ij} are the observed and expected counts for the i,j-th cell, respectively.

To conclude the section on the method, it is necessary to evaluate the simulation results (point 6).

Firstly, the estimation of the variable under study will be more accurate the stronger the correlations between the constraining variables and the additional variables in the microdata of the Multiscopo sample survey (and if these correlations represent the variations in each simulated area considered). It should be emphasized that this requirement applies not only to spatial microsimulation techniques but also to estimation techniques for small areas based on regression models. To be valid, these techniques require a high predictive capacity of the independent variables in the chosen regression model underlying the simulation. In our study, this requirement is satisfied by the results of the analyses on the sample data, as stated in point three.

Secondly, the spatial microsimulation method ends with point estimates and does not provide uncertainty intervals around the point estimates. Currently, despite some promising attempts to solve this problem (Whitworth et al., 2017; Moretti and Whitworth, 2021), the accuracy of the estimates is generally assessed through a combination of internal validation based on the Total Absolute Error (TAE) and external validation with respect to some related results whose distribution is known (Smith et al., 2011, Edwards et al., 2012).

In this paper, internal validation of the model was performed using TAE, which, as mentioned earlier, is a measure of statistical fit that compares the observed values in the initial tables provided by the Census with the tables calculated from the estimated microdata. In our case, the final TAE is equal to 86: in other words, considering the distribution of over 43,000 units, the produced tables deviate from the census tables by only 86 units, i.e., 0.2%. This result confirms the excellent fit of the model to the initial data.

3. Results and implications for policymakers

The main result of applying spatial microsimulation is to quantify the studied problem, which in this case is the physical vulnerability of the elderly at the municipal level for the two age groups 65-74 and 75 years and above (Table 1).

Table 1 – *Extract from the table of estimates for elderly residents in nursing homes and vulnerable elderly individuals living at home by municipality code in province of Sondrio.*

Municipality Code	Residents aged 65-74, living at home, "physically vulnerable" "	Residents aged 75 and above, living at home, "physically vulnerable""	Residents in nursing homes, care homes for disabled adults and the elderly	Total residents aged 65 and above
14001	6	9	0	99
14002	38	61	0	879
14003	5	13	0	145
…	…	…	…	…
14076	1	3	0	49
14077	7	18	0	205
14078	24	53	52	748
Tot. Prov.	1657	3310	1152	43788

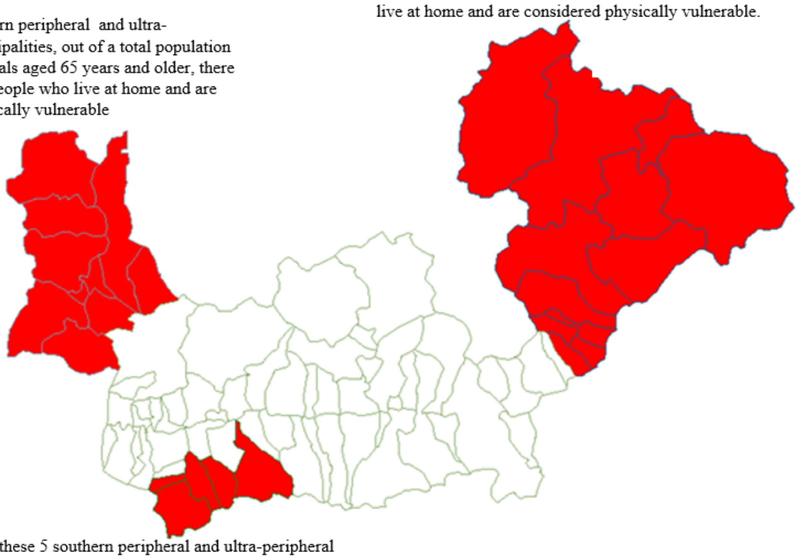
Note: tables is only for illustrative purposes and therefore does not present all values.

Synthetic data from spatial microsimulation is valuable for social policy planning, specifically in estimating the number of physically vulnerable elderly individuals in regions facing challenges related to healthcare, connectivity, transportation, population density, and social services.

Using the typology developed for internal areas (Istat, 2022), peripheral and ultra-peripheral areas in Sondrio were analysed to determine the number of vulnerable elderly people living in these areas with limited amenities. In particular peripheral and ultra-peripheral areas are characterised by their distance of over 40 minutes from a "hub," which refers to municipalities providing secondary education options (comprising at least one high school, be it scientific or classical, as well as at least one technical or vocational institute), a DEA-level hospital, and a railway station meeting at least the silver standard.

The choropleth map (Figure 2) visually illustrates the peripheral and ultra-peripheral areas in red, along with the estimated count of physically vulnerable elderly individuals residing within them.

Figure 2 – *Choropleth map with peripheral and ultra-peripheral municipalities highlighted in red and the estimated total number of vulnerable elderly individuals, living at home.*



In these 12 eastern peripheral and ultra-peripheral municipalities, out of a total population of 7,039 individuals aged 65 years and older, there are 788 elderly people who live at home and are considered physically vulnerable.

In these 10 western peripheral and ultra-peripheral municipalities, out of a total population of 5,018 individuals aged 65 years and older, there are 616 elderly people who live at home and are considered physically vulnerable

In these 5 southern peripheral and ultra-peripheral municipalities, out of a total population of 292 individuals aged 65 years and older, there are 44 elderly people who live at home and are considered physically vulnerable.

4. Conclusion and limitations of the study

This study has demonstrated that spatial microsimulation can be used to estimate the number of vulnerable elderly individuals in the province of Sondrio. Using simulated data generated through this technique, local administrators can identify areas with the highest number of vulnerable elderly individuals and concentrate resources and services to support these individuals, such as planning the construction of elderly care facilities, enhancing transportation services for people with physical limitations, or providing home support for those who have difficulty moving.

However, the technique has some limitations, particularly in its ability to provide uncertainty intervals around the central estimates. Nevertheless, this does not mean that spatial microsimulation approaches cannot have potential advantages once the issue of estimation accuracy is resolved. Therefore, in our opinion, the potential opportunities offered by spatial microsimulation approaches should not be ignored in further development.

References

ASIAN DEVELOPMENT BANK. 2020. *Introduction to small area estimation techniques. A Practical Guide for National Statistics Offices*. Manila: ADB.

CLEGG, A., YOUNG, J., ILIFFE, S., RIKKERT, M. O., ROCKWOOD, K. 2013. Frailty in elderly people, *Lancet* (London, England), Vol. 381, No. 9868, pp. 752–762.

EDWARDS, K.L., TANTON, R. 2012. Validation of Spatial Microsimulation Models. In TANTON, R., EDWARDS, K. (Eds.) *Spatial Microsimulation: A Reference Guide for Users. Understanding Population Trends and Processes*, Vol. 6, Dordrecht, Springer.

EUROPEAN INSTITUTE FOR GENDER EQUALITY. 2021. *Gender Equality Index 2021: Health.* Luxembourg: Office of the European Union.

GALLUZZO L., O'CAOIMH R., RODRÍGUEZ-LASO Á., BELTZER N., RANHOFF AH., VAN DER HEYDEN J. 2018. Incidence of frailty: a systematic review of scientific literature from a public health perspective, *Annali Istituto Superiore Sanità*, Vol. 54, No. 3, pp. 239-245.

HARLAND, K., HEPPENSTALL, A.J., SMITH, D., BIRKIN, M. 2012. Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques, *Journal of Artificial Societies and Social Simulation*, Vol. 15, No, 1.

HARLAND, K. 2013. Microsimulation model user guide ver. 1.0 (Flexible Modelling Framework). *Woking paper 6/13,* School of Geography, University of Leeds, United Kingdom.

ISTAT. 2021. Censimenti permanenti. Roma: Istituto Nazionale di Statistica.

ISTAT. 2022. La geografia delle aree interne nel 2020: vasti territori tra potenzialità e debolezze, *Statistiche Focus*. Roma: Istituto Nazionale di Statistica.

ISTAT. 2023. Previsioni della popolazione - anni 2021-2070. Roma: Istituto Nazionale di Statistica.

MORETTI, A., WHITWORTH, A. 2021. Estimating the Uncertainty of a Small Area Estimator Based on a Microsimulation Approach, *Sociological Methods & Research*, pp. 1-31.

O'DONOGHUE, C., MORRISSEY, K., LENNON, J. 2014. Spatial Microsimulation Modelling: A Review of Applications and Methodological Choices, *International Journal of Microsimulation*, Vol. 7, No. 1, pp. 26–75.

PETRELLI, A., DI NAPOLI, A., SEBASTIANI, G., ROSSI, A., GIORGI ROSSI, P., DEMURU, E., COSTA, G., ZENGARINI, N., ALICANDRO, G., MARCHETTI, S., MARMOT, M., FROVA, L. 2019. Atlante Italiano delle Disuguaglianze di Mortalità per Livello di Istruzione, *Epidemiologia e prevenzione*, Vol. 43, No. 1S1, pp. 1-120.

RAHMAN, A., HARDING, R., TANTON, S. L. 2010. Methodological Issues in Spatial Microsimulation Modelling for Small Area Estimation, *International Journal of Microsimulation*, Vol. 3, No. 2, pp. 3-22.

RAO J.N.K. 2003 *Small area estimation*. Hoboken, New Jersey: John Wiley & Sons.

ROCKWOOD, K., SONG, X., MACKNIGHT, C., BERGMAN, H., HOGAN, D. B., MCDOWELL, I., MITNITSKI, A. 2005. A global clinical measure of fitness and frailty in elderly people, *CMAJ: Canadian Medical Association journal*, Vol. 173, No. 5, pp. 489-495.

SMITH, D. M., HEPPENSTALL A., CAMPBELL, M. 2021. Estimating Health over Space and Time: A Review of Spatial Microsimulation Applied to Public Health, *J-Multidisciplinary Scientific Journal*, Vol. 4, No. 2, pp. 182-192.

SMITH, D.M., PEARCE, J.R., HARLAND, K. 2011. Can a Deterministic Spatial Microsimulation Model Provide Reliable Small-Area Estimates of Health Behaviours? An Example of Smoking Prevalence in New Zealand, *Health Place*, Vol. 17, No. 2, pp. 618–624.

TANTON, R. 2014. A Review of Spatial Microsimulation Methods, *International Journal of Microsimulation*, Vol. 7, No. 1, pp. 4-25.

WHITWORTH, A., CARTER, E., BALLAS, D., MOON, G. 2017. Estimating Uncertainty in Spatial Microsimulation Approaches to Small Area Estimation: A New Approach to Solving an Old Problem, *Computers, Environment and Urban Systems*, Vol. 63, pp. 50-57.

Alberto VITALINI, Istat, Sede della Lombardia, vitalini@istat.it
Simona BALLABIO, Istat, Sede della Lombardia, ballabio@istat.it
Flavio VERRECCHIA, Istat, Sede della Lombardia, verrecchia@istat.it