Journal of Research and Didactics in Geography (J-READING), 2, 7, Dec., 2018, pp. 95-104
DOI: 10.4458/1682-09

J - READING
JOURNAL OF RESEARCH AND DIDACTICS IN
GEOGRAPHY
homepage: www.j-reading.org

New ISTAT “microzones” layer: a new way to read land cover statistics

Stefano Mugnoli, Fabio Lipizzi, Antonella Esposto[a]
[a] DIPS-DCAT-ATA, ISTAT (National Institute of Statistics), Rome, Italy
Email: mugnoli@istat.it

Received: July 2018 – Accepted: October 2018

Abstract

The aim of this paper is to describe the procedures used to integrate geographic datasets in order to produce new ISTAT “microzones” layer, an upgrade of the 2011 Census cartography.
The above is an experiment based on the management and fusion between Land Cover data concerning hinterland areas (i.e. cadastral data) and maps produced by regional or local authorities. All the activities are in progress and under assessment. Integration and elaboration are carried out using a number of ArcGIS 10.4.1 tools. The main achievements so far have been to produce integrated geographic datasets and to link microzones land cover and use legend with LUCAS (Land Use/Cover Area frame Survey) one. Other very useful pilot data are represented by four band high resolution aerial images; calculating simple radiometric indices (SAVI, ENDVI), in fact, it can be possible to improve the estimates of vegetation cover, especially in urban areas. All the information collected can be a very useful way of improving the quality of land cover/use statistical data, although the integration of data that came from different sources can involve an accuracy loss and a generalization of the final product; the activities will be extended to the entire Italian territory to enhance the value of the input data. Future work will be planned to automate all data processing and integration with other geographic database sources, in order to increase data details and to reduce the generalization of the same.

Keywords: ArcGIS 10.4.1, Census Data, Enumeration Areas, Land Cover/Use

1. Introduction

The renewal process of the Italian National Institute of Statistics (ISTAT) provides the data production through the new Integrated Statistical Registers system (SIR). One of the four SIR registers is the Base register of the site (RSLB) that will make it possible to uniquely locate all SIR information. For this reason, ISTAT has planned the implementation of the enumeration areas layer called “microzones”. Therefore, the new microzones layer constitutes the base map to contain and disseminate SIR data and information (Lipizzi, 2013).

This paper aims to briefly set out the methodology used to realize the new ISTAT microzones layer; some legend details will be provided in order to better understand the way in which each polygon is represented on the map. Its name is related to the fact that the microzones layer is a further subdivision of the ISTAT enumeration areas layer; the latter is divided into very small polygons, homogenous in their LC/LU aspects; this creates a kind of a plot made up of many micro-areas.

The ISTAT census enumeration areas vector layer, in fact, represents the cornerstone to analyse the Italian territory from a statistical point of view. All the data collected during census surveys are linked to each of the over 400,000 enumeration areas drawn on Italy. This dense plot helps us to describe the entire national territory in a very detailed way, particularly in urban areas.

Therefore, in order to improve LC/LU statistics and to better characterize each enumeration area, the ISTAT ATA (Environment Territory) unit planned to produce a sort of a microzones mosaic layer described by a land cover/use definition compatible with the LUCAS (Land Use/Cover Area frame Survey) legend. This certainly allows both to define more clearly the homogenous areas contour for the future and to optimize the geo-localization of all census variables. With regard to the above, it is important to remember that ISTAT has just planned a continuous population Census survey that should start very soon. It is thus absolutely fundamental to have a very detailed reference cartography.

Census geographical datasets are essentially used for classifying and characterizing national territory in relation to resident population, buildings, services and industry. Supplementing this information with land cover and land use data, it can be possible not only to produce comprehensive data on land cover/use, but better to calculate some statistical parameters (i.e. population density) at local and global level too. But not just that: in fact, statistical information at this level of detail can be used to evaluate other important phenomena like soil consumption, urban sprawl (European Environment Agency, 2006), accessibility to territory and the demographic change in population distribution.

In short, our product can be considered a sort of "Land Cover/Use (LC/LU) Synthetic Layer" (from the Greek term "*σύνθεσις*" in the sense of getting together geo-statistical information derived from many different geographic datasets); its main use is to support statistical surveys since it is the result of integration and harmonization of different kinds of thematic archives such as administrative, demographic, infrastructure (road, railway, ports, airport, etc.), agricultural Census data and environmental maps. Moreover, the peculiar legend of the map is undoubtedly useful in better understanding the synthesis process.

In Italy CISIS[1] (Centro Interregionale per i Sistemi Informatici Geografici e Statistici) has contributed to harmonising geographical and statistical data. One of the most important results is the release of the database "DB Prior 10K" at national level. The database developed by CPSG (Comitato Permanente per i Sistemi Geografici), provides some layers (i.e. streets, railways, hydrography) with the same data structure.

Therefore, every geographical ISTAT data is designed to pursue the same purpose: to provide standardised information for the entire national territory.

The final geo-statistical microzones layer was developed through collaboration of many people and after the review of many different intermediate products.

In the end, the activity is the sequel of many ISTAT experimentations (Lipizzi and Mugnoli, 2010; Chiocchini and Mugnoli, 2014; Mugnoli et al., 2011; Lombardo et al., 2017).

2. Methodology

The purpose of the new ISTAT microzones layer is to synthesize and integrate different land cover/use and geo-statistical datasets in just only one statistical cartography; but it should not be considered as a cartography product in the strict sense since the vertex distance and the generalization of the polygon is not related to a specific scale; moreover there is no real Minimum Map Unit (MMU) (Marchetti, 2002).

Having said that, considering ISTAT enumeration areas as a basic layer, other many different official LC/LU datasets, at Continental level or realized by national and regional Italian

[1] For more information regarding CISIS activities: http://www.cisis.it/.

institutes (i.e. CORINE LC[2] and cadastral data distributed by AGEA: Agency for the Agricultural Supplies, local LC/LU maps and Technical Cartography) and street/railway networks are merged with the first in order to concentrate both statistics and geographic information in a unique product. In Figure 1 is shown an example of an enumeration area divided into some microzones; the old enumeration are as are in red while the features that have been extracted are in yellow (i.e. streets, railways, rivers, aquaculture facilities).

The main idea arose from an experimental PSN (National Statistical Program) project called Synthetic Statistical Land Cover map (Italian acronym CSCS), in which different geographic datasets are collected together in a unique topological correct layer. In Figure 2 the same area of Figure 1: some features have come to light (buildings in pink, roads in grey, sport facilities in brick, water areas in blue, cropland in ochre, fruit trees in light green and woodland in dark green).

Street and railway layers were converted into polygons starting from commercial networks and running differentiated buffer algorithms; i.e. for roads the buffer was on the basis of the Functional Road Class (FRC) field.

In brief, data integration is based on ArcGIS 10.4.1 algorithms that make it possible to analyze and synthesize vector layers in a very simple way; the most used tools for this experimentation belong to analysis and data management ones. An adjustment of the map topology was necessary by editing tools (i.e. to eliminate micro polygons that were created during the process).

The main problem in our case was the fact that LC/LU local maps are focused on the CORINE legend and it is not so easy, after having integrated CTR and street networks, to classify interstitial zones, especially in urban areas.

Therefore, aerial images processing made a substantial contribution, both in extra urban and urban areas covered by vegetation. In fact, by using some simple vegetation indices it is possible to isolate green areas such as woods, grasslands, urban parks and trees along the roads.

Four band aerial orthophotos at high resolution (50 cm in extra-urban areas and 20 cm in urban centres) are suitable for the extraction of vegetation indices (Xue and Boafeng, 2017) thanks to the presence of the Near Infrared band (NIR). In order to create a complementary layer, some vegetation indices were extracted. The first choice went for SAVI (Soil Adjusted Vegetation Index) (Huete, 1988) that can be automatically calculated by an ERDAS/Imagine software algorithm and ENDVI (Enhanced Normalized Difference Vegetation Index)[3] that has been performed by a batch process in python[4].

For the SAVI the main difficulty is to calculate a canopy background adjustment factor that varies from 0 to 1. The index that can give even more homogeneous results is the ENDVI; in fact, this index uses the red and green as the reflective channels and blue as the absorption channel. It must be remembered that a normal healthy plant will reflect both visible green light and NIR light. The formula is:

$$ENDVI = \frac{[(NIR + GREEN) - (2 * BLUE)]}{[(NIR + GREEN) + (2 * BLUE)]}$$

[2] For more information regarding CORINE LC: https://www.eea.europa.eu/publications/COR0-landcover.

[3] For more information regarding ENDVI: https://maxmax.com/endvi.htm.

[4] Regarding several vegetation indices see also: Mróz and Sobieraj, 2004.

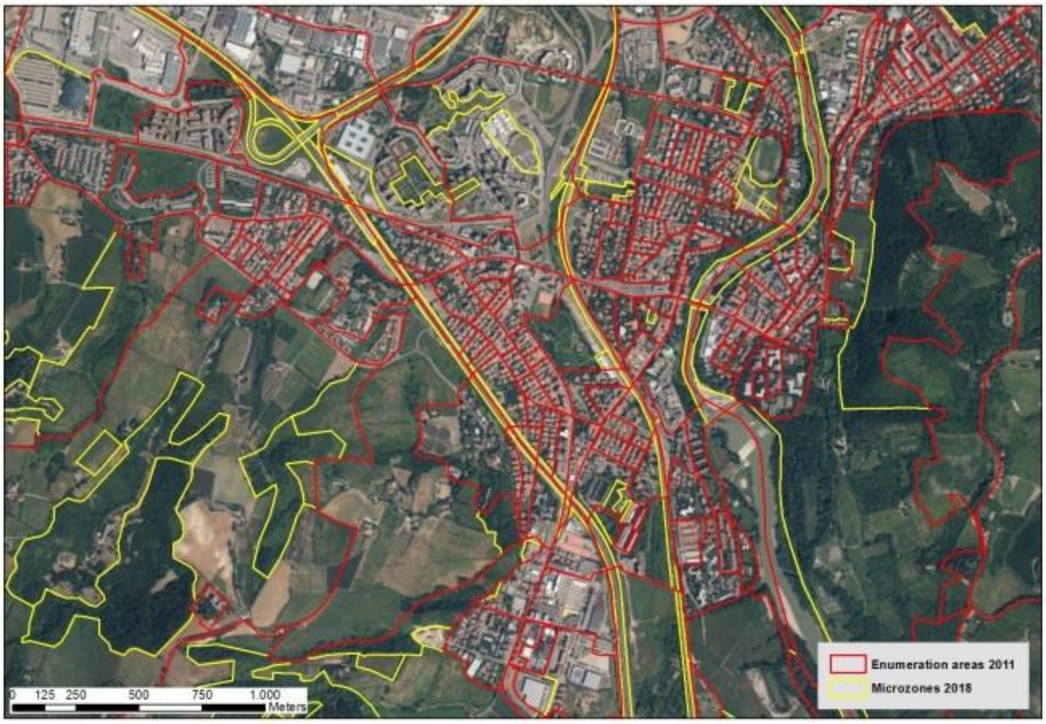

Figure 1. An example of microzones in extra urban area. Bologna municipality. Source: ISTAT.

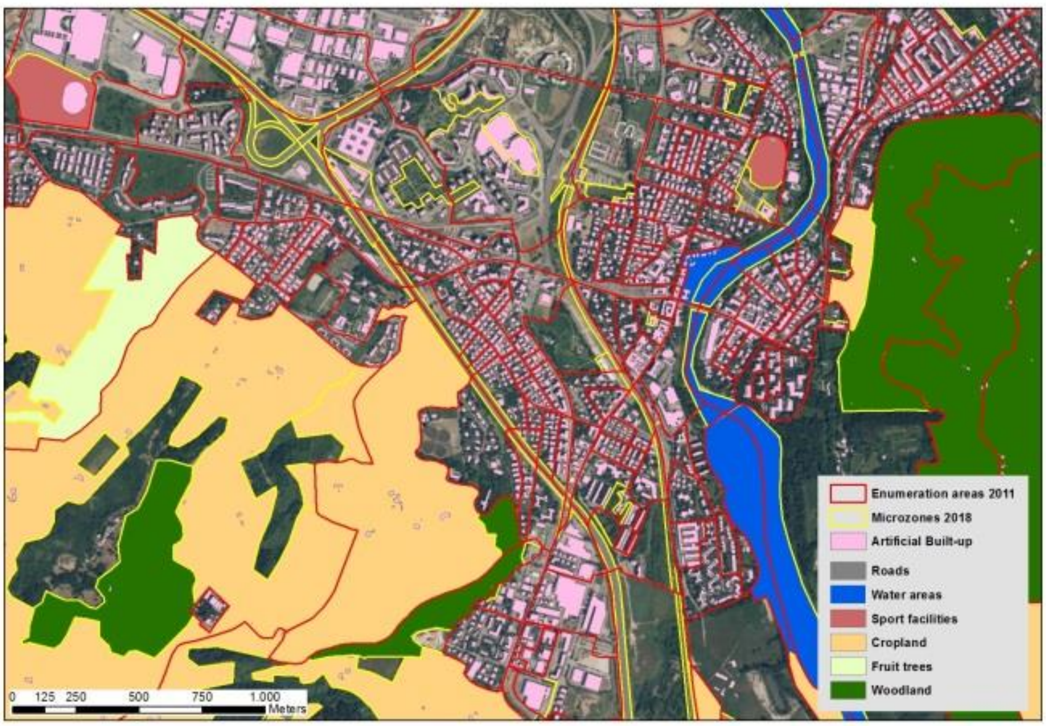

Figure 2. An example of Synthetic Statistical Land Cover map. Source: ISTAT elaboration on regional LC/LU and street network data.

This ENDVI formula sums the NIR and Green channels together for the reflective channel. The Blue channel is multiplied by two to compensate for the NIR and Green channels being summed together.

In Figure 3A is shown a result of image processing algorithm to extract ENDVI index in an urban area (Rome EUR), where white pixels represent green areas. After that, it is necessary to reclassify the image for two reasons: it is fundamental to convert the continuous file in a discrete one; it must be made sure to isolate just the pixels that represent the green areas; an example of this approach is provided in Figure 3B (in green just the pixels that had a DN>=0,2).

Therefore, with the help of two ERDAS/Imagine algorithms (Clump and Eliminate) it was possible to group together green pixels in clusters and then remove areas below 1,000 sqm (Kienast-Brown and Boettinger, 2010; Dobermann et al., 2003; Della Rocca et al., 2001). The first algorithm identifies groups of contiguous pixels in the same class, called “raster regions” or “clumps”; a clump image is the input of the second algorithm that eliminates clumps below a given area. In this way, small noise (salt and pepper) effects are removed in thematic maps.

Vegetation index images can be reclassified into a two class thematic layers: areas with and without vegetation. In this case, we have decided to identify the two classes as green urban areas (inside ISTAT urban centres) and artificial land.

Finally, raster regions are converted into a polygon vector layer.

Figure 3A. An example of ENDVI index image in urban area (Rome EUR). Source: ISTAT elaboration on data AGEA.

Figure 3B. Figure 3A reclassified (Rome EUR). Source: ISTAT elaboration on data AGEA.

3. CSCS and Microzone LC legend

ISTAT enumeration are as are described by a lot of attributes that identify each polygon from an administrative and statistical point of view. There are some codes that can be useful to frame each area in a sort of LC/LU classification.

Since 2001 each enumeration area had been identified according to a key related to its main "vocation". This sort of legend was focused especially on human activities, uses or services for the citizen.

Having considered the need to define a clear and useful LC/LU legend to uniquely describe the entire national territory, the choice has fallen upon LUCAS (Land Use and Cover area frame Survey) because this is a "survey that provides harmonized and comparable statistics on land use and land cover across the whole of the EU's territory"[5]. And not just for this reason, the CSCS class legend has been based on the LUCAS one because LUCAS is an LC pure legend; moreover, all the map layers at our disposal make it possible to identify each CSCS polygon by a LUCAS class. Upon completion of the CSCS description, it is easy to transfer the classification to the microzones layer since the latter is a sort of summary of the former.

The first draft provides a 45 LC class, mostly at LUCAS level 1. But classifying each microzone is not always simple, especially in the case in which polygons can be referred to LU rather than to LC. For example, it is very difficult to characterize the "green urban areas" on the basis of the LC pure legend, as LUCAS is. Usually green areas are classified on the basis of their use (i.e. amusement parks, community gardens, etc.). Attempts have been made to separate grasslands and woodlands from "green artificial" ones. In this specific case an invaluable aid is represented by image processing, as described above.

[5] For more information regarding LUCAS survey: http://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey.

4. Topology rules and accuracy assessment

When different geographical databases are merged into a unique layer, some overlay errors inevitably occur. It is therefore essential to define very strict topology rules upstream. First of all, you have to decide the overlay order of the layer. In our case the basic layer is represented by water (river, lake, lagoons, etc.) and wetlands; above this railways, streets and buildings in this order; then, agricultural and natural area layers; finally the polygons derived from the ENDVI.

Of course, in so doing, it is necessary to deal with the overlay areas (bridge, road crossings, etc.). Using some simple ArcGIS 10.4.1 analysis algorithms (Intersect and Symmetrical difference), different layers can be merged automatically without topology errors. To give an example, Basilicata, an Italian region of about 10,073 sq.km composed of 5,107 enumeration areas, is divided into 317,928 polygons in CSCS layers.

It is only thanks to the fact that the topology is correct that it is possible to evaluate the land cover of each class. In Table 1 is shown a summary of land cover surfaces for Basilicata related to the LUCAS legend at level 0.

X,Y tolerance is set at 1m, the same as the enumeration area layers.

LUCAS Class	Definition	Surface (ha)	% Region
A00	Artificialland	27,796.86	2.76
B00	Cropland	530,902.33	52.70
C00	Woodland	219,757.59	21.82
D00	Shrubland	64,059.24	6.36
E00	Grassland	17,669.29	1.75
F00	Bare land and lichens	120,381.03	11.95
G00	Water areas	11,098.02	1.10
H00	Wetlands	174.75	0.02

Table 1. Basilicata land cover surfaces according to LUCAS legend (2011). Source: ISTAT elaboration.

An additional benefit in using the LUCAS legend is the possibility to assess the accuracy of the CSCS layer by LUCAS points themselves. The total accuracy for Basilicata referred to 597 LUCAS points (year 2012) stands at about 70% (69.23%). Class accuracy varies from 72.02% for the woodland to 33.33% for the grassland[6].

The real problem is due to the number of LUCAS points of the less represented classes. In our case, for example, we have just 5 points for the “Bare land and lichens/moss” and no LUCAS points for the “wetlands”. Moreover, it is clear from the error matrix that there are clear overlaps between natural grasslands (pastures) and agricultural ones.

The CSCS layer is completed for 8 Italian regions and it is now proceeding to transfer information to the Census 2011 enumeration area layers to create microzones layer. This process has been completed just for Emilia-Romagna that now has 65,346 microzones vs 38,603 old enumeration areas.

In Figure 4A a focus on the Census 2011 enumeration layer (Parma municipality); in Figure 4B the same zone for the microzones one. Different colors represent different LC classes.

5. Conclusions

The need to have a homogeneous statistical cartography for the entire national territory is a priority not only for ISTAT but for national and local administrative authorities too. Enumeration areas layer have played a crucial role until now in describing statistical indicators in their territorial and environmental aspects.

However, enumeration areas are not suitable to describe LC and get territorial parameters to some important ISTAT surveys (i.e. agricultural census, transport and services surveys, etc.). So the new ISTAT microzones layer has to be seen as the base map phenomena.

[6] Regarding the status of land cover classification accuracy assessment see: Foody, 2002.

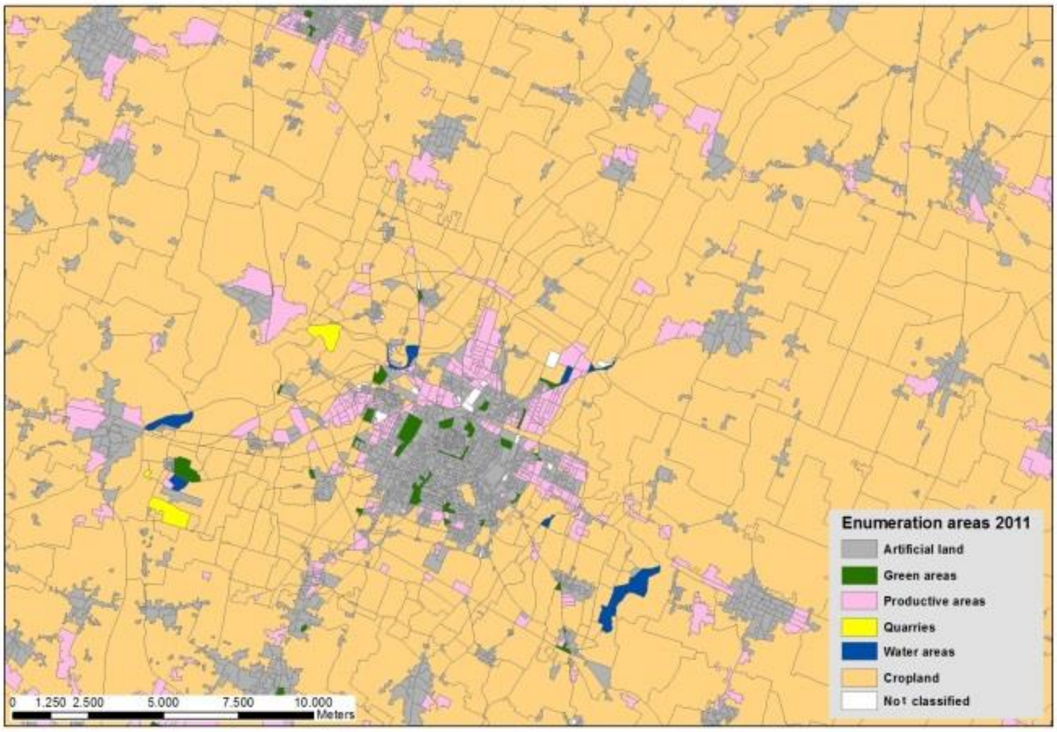

Figure 4A. Census 2011 enumeration areas classification (Parma municipality). Source: elaboration on data ISTAT.

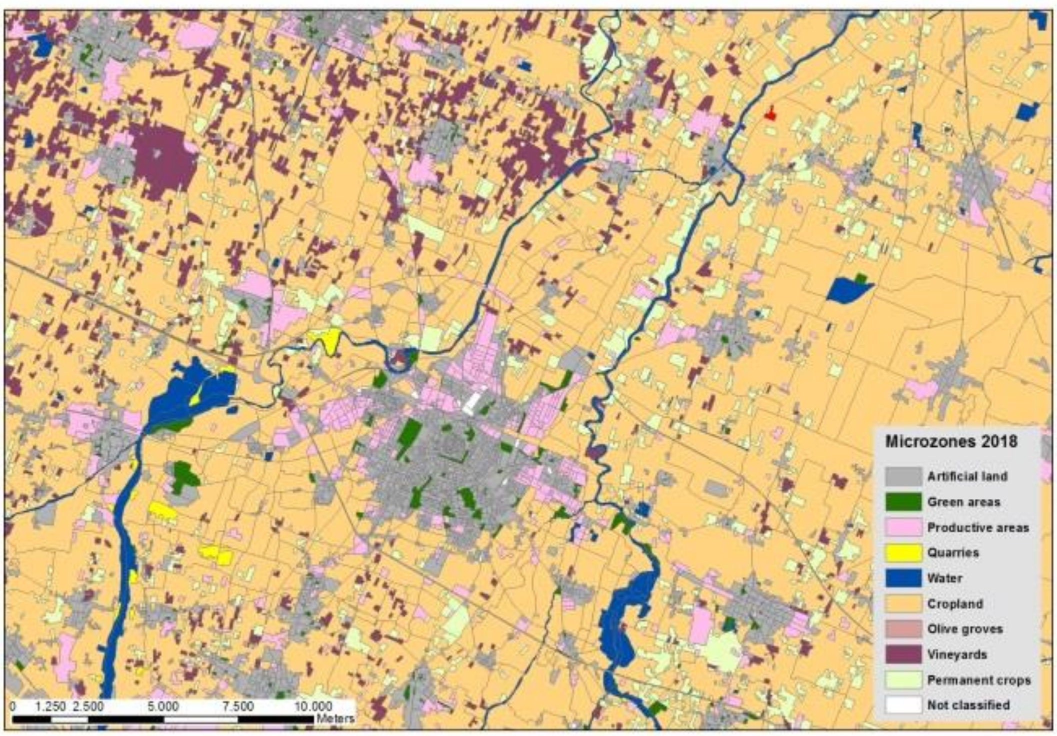

Figure 4B. Microzones 2018 classification for the same area as Figure 4A. Source: elaboration on data ISTAT.

Another important aspect is the fact that ISTAT releases the data at least at municipality level. It is important to say that the Italian municipality surfaces vary greatly: from 12 Ha (Atrani) to 1,287 sq. km (Rome). CORINE and other National cartographic production (i.e. National Map of Soil Consumption produced by ISPRA)[7] are not suitable because of their spatial resolution.

From the above, new microzones layer are really a new way to read LC/LU statistics. Just having data about LC linked to statistical ones enables one to classify any area of the National territory according to criteria which may be decided on the basis of specific needs. Moreover, the storage of the data in a geodatabase allows data extraction by very simple SQL queries. Furthermore, all the geographical data are correct in relation to topology and this fact is fundamental for the quantification of the LC/LU classes.

All the information collected so far can be a very useful instrument to improve the quality of statistical land cover/use data although the integration of data that came from different sources can involve an accuracy loss and a generalization of the final product. The future steps will be planned to automate all data processing and integration with other geographic databases so as to increase the data details and reduce the data generalization.

During 2018 the project foresees the involvement of 80% of the National territory.

Acknowledgements

- Even if the paper was devised together by the Authors, F. Lipizzi wrote paragraphs 1 and 5, S. Mugnoli wrote paragraphs 2, 3 and 4, A. Esposto worked on Figures and Tables.
- The Authors are extremely grateful to Luisa Franconi for the revision of the text.
- The Authors thank all the people of the ISTAT ATA unit who daily works to implement ISTAT geodatabases.

[7] The National Map of Soil Consumption can be downloaded from: http://www.sinanet.isprambiente.it/it/sia-ispra/download-mais/consumo-di-suolo.

References

1. Chiocchini R. and Mugnoli S., "Land Cover and Census integration geographic datasets to realize a statistics synthetic map", *Proceedings of European Forum for Geography and Statistics* (Krakow, 22-24 October 2014), European Forum for Geography and Statistics Conference, 2014, https://www.efgs.info/conferences/efgs/2014-krakow/.
2. Della Rocca A.B., Pignatti S., Mugnoli S. and Bianco P.M., "La Carta della Vegetazione della Tenuta di Castelporziano", in Accademia Nazionale delle Scienze detta dei XL (dei Quaranta) (Ed.), *Il Sistema Ambientale della Tenuta di Castelporziano*, Scritti e Documenti XXVI, Rome, 2001, pp. 709-747.
3. Dobermann A., Ping J.L., Simbahan G.C. and Adamchuk V.I., "Processing of yield map data for delineating yield zones", in Stafford J. and Werner A. (Eds.), *Precision Agriculture*, Wageningen, Wageningen Academic Publisher, 2003, pp. 177-186.
4. European Environment Agency, *Urban Sprawl in Europe – the ignored challenge*, EEAreport n. 10/2006.
5. Foody G.M., "Status of land cover classification accuracy assessment", *Remote sensing of environment*, 80, 1, 2002, pp. 185-201.
6. Gallego F.J., "Remote sensing and land cover area estimation", *International Journal of Remote Sensing*, 25, 15, 2004, pp. 3019-3047.
7. Huete A.R., "A soil adjusted vegetation index", *Remote sensing of environment*, 25, 3, 1988, pp. 295-309.
8. Huete A., Didan K., Miura T., Rodriguez E.P., Gao X. and Ferreira L.G., "Overview of the radiometric and biophysical performance of the MODIS vegetation indices", *Remote Sensing of Environment*, 83, 2002, pp. 195-213.
9. Kemp K., *Encyclopedia of Geographic Information Science*, SAGE Publications, 2008.
10. Kienast-Brown S. and Boettinger J.L., "App-

lying the Optimum Index Factor to multiple data types in soil survey", in Moore A.C., Kienast-Brown S., Boettinger J.L. and Hartemink A. (Eds.), *Digital Soil mapping: Bridging research, Environmental Application and Operation*, Berlin, Springer 2010, pp. 385-398.

11. Lipizzi F., *Innovazioni di processo e di prodotto nelle fasi di aggiornamento delle basi territoriali 2010-2011*, ISTAT Working Papers, n. 2/2013.
12. Lipizzi F. and Mugnoli S., "Le statistiche agricole verso il Censimento del 2010: valutazione e prospettive", *Proceedings of the conference "Le Statistiche agricole verso il Censimento del 2010: valutazioni e prospettive"* (Cassino, Università di Cassino, 26-27 October 2006), Cassino, 2010, pp. 381-394.
13. Lombardo G., Esposto A., Minguzzi R. and Mugnoli S., "La CSS ISTAT un nuovo strumento per le statistiche territoriali", *Geomedia*, XXI, 2, 2017, pp. 26-30.
14. Marchetti M., "Metodologie per una cartografia di uso del suolo multilivello e multiscala: analisi e sperimentazione applicativa", *Documenti del territorio*, 49, 2002, pp. 33-51.
15. McRoberts R.E., Wendt D.G., Nelson M.D. and Hansen M.H., "Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates", *Remote Sensing of Environment*, 81, 1, 2002, pp. 36-44.
16. Mróz M. and Sobieraj A., "Comparison of several vegetation indices calculated on the basis of a seasonal spot xs time series, and their suitability for land cover and agricultural crop identification", *Technical Sciences*, 7, 2004, pp. 39-66.
17. Mugnoli S., Chiocchini R., Cruciani S., Esposto A. and Lipizzi F., "Integrazione di dataset geografici di copertura del Suolo e Censuari per la realizzazione di una mappa statistica sintetica", *Proceedings of the XV National Conference ASITA 2011* (Colorno, 15-18 November 2011), Parma, 2011, pp. 1633-1640.
18. Xue J. and Boafeng S., "Significant Remote sensing Vegetation Indices: a Review of Developments and Applications", *Journal of Sensors*, 2017, pp. 1-17.