International Journal of *Geo-Information*

Technical Note

Official Statistics, Building Censuses, and OpenStreetMap Completeness in Italy

Gianluigi Salvucci [1] and Luca Salvati [2,*]

[1] Italian National Institute of Statistics (ISTAT), Piazzale G. Marconi 16, I-00144 Rome, Italy; gianluigi.salvucci@istat.it
[2] Department of Economics and Law, University of Macerata, Via Armaroli 43, I-62100 Macerata, Italy
* Correspondence: luca.salvati@unimc.it; Tel.: +39-661-5710; Fax: +39-661-571-036

Abstract: The present study provides a simplified framework verifying the degree of coverage and completeness of settlement maps derived from the OpenStreetMap (OSM) database at the national scale, with a possible use in official statistics. Measuring the completeness of the objects (i.e., buildings) derived from OpenStreetMap database supports its potential use in building/population censuses and other diachronic surveys, as well as administrative sources such as the register of building permits and land-use cadasters. A series of measurements at different scales are proposed and tested for Italy, in line with earlier studies. While recognizing the potential of the OpenStreetMap database for official statistics, the present work underlines the urgent need of an additional (spatially explicit) analysis overcoming the data heterogeneity and sub-optimal coverage of the OSM information source.

Keywords: volunteered geographic information (VGI); comparative analysis; spatial testing; sampling under-coverage

check for updates

Citation: Salvucci, G.; Salvati, L. Official Statistics, Building Censuses, and OpenStreetMap Completeness in Italy. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 29. https://doi.org/10.3390/ijgi11010029

Academic Editor: Wolfgang Kainz

Received: 21 October 2021
Accepted: 26 December 2021
Published: 31 December 2021

1. Introduction

Recent advances in Web 2.0 technologies have transformed Internet users from "content seekers" to "data producers" [1]. Notably, this process has simultaneously led to the search for free software and open data, in turn creating a collaborative consciousness where users become both data creators and data sharers [2]. Technological progress allows even less skilled stakeholders to produce significant data [3]. From this perspective, open software allows a reduction in the cost of assessment tools and solutions to repetitive operations [4]. In a context where each user becomes a sensor that acts in a global process of information sharing [5], data are no longer the monopoly of a given institution/authority. We are instead witnessing a process of "democratization" of geographic data or a sort of "new geography without geographers" [6]. On the one hand, one can arrive at the paradox that data produced by single users are more up-to-date and truthful than administrative data [7]. On the other hand, a question may arise: "How reliable are data from users?". If all these assumptions serve to bring young people closer to geographical analysis, e.g., increasing awareness of the problems related to the area where they live, then this "new geography" literature is likely to be an appropriate tool for improving official statistics [8]. However, geography is more than the mere enumeration of place names, rather claiming to provide a truly systematic analysis of the individual elements within a given territory [9].

Contributing to the shift from "data users" to "data producers", the OpenStreetMap (OSM) community was launched in July 2004. The limited access to free geographic data was a motivation at the base of its (so rapid) development [10]. In Europe, up to that time, spatially explicit data sources were mainly collected and disseminated by the National Statistical Institutes of individual countries [11]. Moreover—at least in some cases—these data were less precise and had a more restricted availability than equivalent systems, e.g., in the United States of America (the U.S. Tiger system releases extremely accurate data with regular updates). Dissemination of spatially explicit data is still difficult in European countries, and, for instance, the Eurostat system of official statistics currently provides a wide range of (aggregate) geographic information, which rarely goes below the municipal scale [12–14]. For this reason, the OSM archive—especially in Europe—may complement (and fill the lack of) spatially explicit data referring to, e.g., buildings, streets, and addresses. These elements were considered the necessary information for geo-referencing, geo-coding, and other operational procedures of interest for different disciplines and applications [15–17]. The main stimuli for digitization in an OSM frame came from the ability to acquire GPX tracks through Potlatch, and later through Yahoo, which made satellite imagery available since 2006 [9]. However, despite strong technological improvements, OSM users are still regarded as less skilled practitioners, mostly editing geographic objects for fun.

OSM includes easy technology, precise enough to be used for administrative purposes. Building and land-use maps from OSM were proposed to replace cadasters [8]. However, the use of administrative sources requires a careful analysis of spatial coverage. Which buildings have OSM users mapped? Which places received attention and careful inspection? To answer these questions, one would examine the individual building information available from OSM in representative areas. Based on empirical studies, OSM was demonstrated to provide significant improvements to Corine Land Cover maps [18], becoming essential, e.g., in emergency management. For instance, a significant boost in data production was observed as an indirect response to the Nepal earthquake [19], likely reflecting a sense of community and participation during disasters. At the same time, the "wikification" of geography [20] proved to be a problem in terms of object representation and map completeness. Who guarantees the quality of the released data? From this perspective, data quality includes both (i) accuracy of shapes and positioning of geographical elements with respect to reality, and (ii) precision of the supplementary information eventually associated with geographical data [21]. Additionally, can the considerable effort of users replace administrative sources? The geographical world is evolving, and the truth is no longer found in archives but in their sharing. In order to answer this last question, a series of statistical indicators can be proposed and calculated from the comparison of the OSM database and official sources [22]. Based on these premises, the present study evaluates the intrinsic quality of data derived from OSM with respect to official statistics (e.g., building maps). Additionally, the study suggests operational approaches to verify and improve the completeness and spatial coverage of the data collected for individual buildings within the frame of volunteered geographic information. The final objective of the study is to delineate advantages (and weaknesses) of an extensive use of OSM in official statistics.

2. Materials and Methods

2.1. Study Area and Logical Framework

The present study adopts Italy (301,330 km^2) as a case study. A national coverage is appropriate when comparing the representativeness of open/volunteered information sources with official data sources. As far as the settlement maps derived from official statistics in Italy, the National Institute of Statistics (ISTAT) no longer carries out the traditional census of buildings, and the related maps produced in past surveys were—at some instances—partial [17]. Technical difficulties arose when managing these large amounts of data. Concerning the possible comparison of official information with other data sources, including OSM, an extensive and comparative analysis of the quality of building maps at enough large scales is particularly hard to realize [18]. An earlier study [23] has investigated the municipal territory at a very high spatial resolution using hexagonal grids (Figure 1). Grids overcome the intrinsic limits in the use of administrative boundaries, allowing a direct investigation of the coverage rate [10]. While documenting the relatively good coverage rate downtown in respect with that in the suburbs, the empirical results of such studies do not verify whether the differential coverage rate depends on the fact that OSM users may encounter difficulties in editing buildings (and collecting their basic characteristics) outside the boundaries of inner cities. More likely, given that most users only map their neighborhoods, differential coverage could be a consequence of a lower density of OSM users in suburban/rural areas [10] (p. 1078).

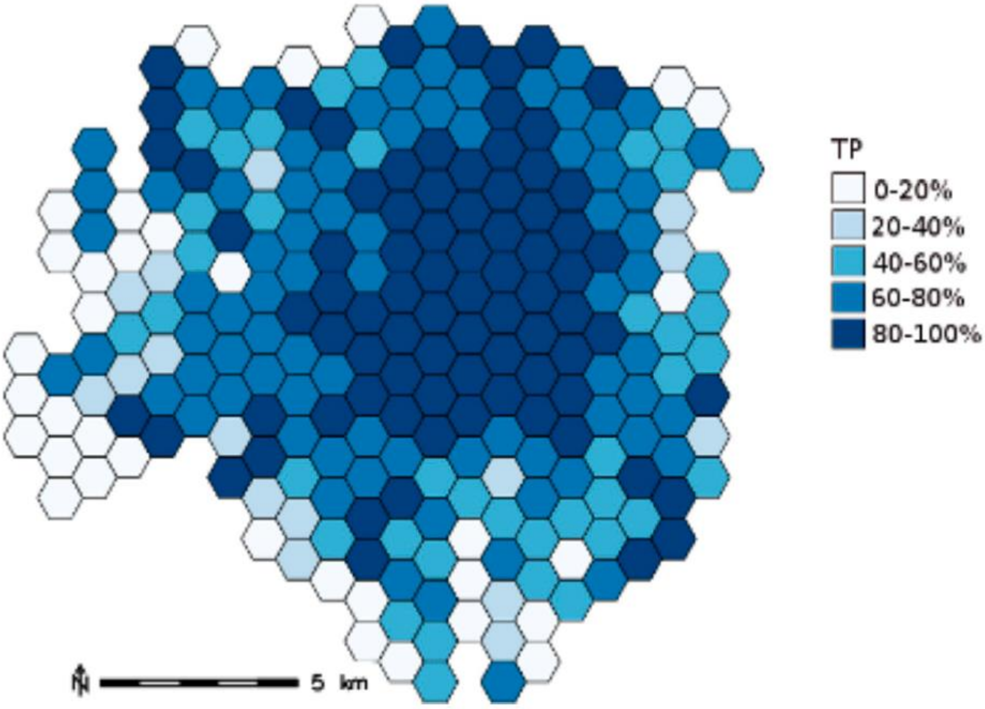

Figure 1. An example of the spatial distribution of the percent share of OSM building area in total area of settlements derived from administrative archives (TP) in Milan municipality [23].

In the present work, we tried to improve the location accuracy of any investigated geographical element, e.g., checking for the position of the shape's vertices in a representative sample of buildings [10]. The difficulty in applying this methodology to Italy lies in the partial availability of a reference geo-database whose semantic and geographical quality is fully known and certified. Considering that a total of 24,156,305 buildings were surveyed at the national level (building census), comparing homologous point features between OSM and official statistics is particularly complex. Similar works were carried out in small portions of the national territory; for instance, the full content of two databases was compared in a part of Pavia province, Northern Italy [24], performing a visual comparison between the buildings captured in the ortho-photographs made available on Google Maps and those surveyed in OSM (Figure 2). Application of such a methodology at the national scale implies editing all the individual buildings, which is time-consuming and methodologically complex. This issue is in common with other European (mostly Mediterranean) countries [25–27].

Figure 2. Identification of buildings from a reference database (overlap of Google maps (image) and OpenStreetMap (point) features): an example from Trivolzio, Pavia, Northern Italy (Franzini et al., 2020).

The choice of a reference database for computation arises as another important issue [11]. Up to few years ago, there were no open-license building archives at the national scale in Italy that can be used as a reference, both in terms of completeness and location accuracy. The Italian cadaster is considered the most authoritative source, although dissemination in digital, editable formats is still restricted to specific stakeholders. Although some cadastral data have been recently disseminated freely under the CC-BY license, only partial information is available via WMS services. Such a dissemination mode may allow the retrieval of information through relational queries, while preventing the direct use of geo-referenced geometries (i.e., shapefiles). The civil protection authority of Italy has recently made available a geo-database of buildings containing the "urban structural aggregates" for the entire national territory. These are geometries starting from the analysis of individual buildings' shape based on various (regional) databases, which have been acquired into polygon formats containing individual buildings, which were identified with a sequential identifier. This archive was created for earthquake prevention, by aggregating contiguous buildings. In fact, in case of an earthquake, it is important to estimate the behavior of the aggregates and not only of the single buildings. The archive benefited of a wide spatial coverage under a CC BY 4.0 license. This condition allows users as well as the related administrative sources to reproduce the database's content in any way. Based on this premises, this archive is taken as a key reference in the context of the present study and was extensively used as the buildings open data archive (BODA) in the following analysis.

2.2. Buildings Open Data Archive (BODA)

By paying particular attention to data resolution (understood as the ability to identify individual buildings) and data completeness, BODA integrated different cartographical sources of data, that include the following: (i) administrative sources of data derived from cartographic services and other technical offices of regions, provinces, municipalities, and other entities, most of them participating in the National Statistical System, SISTAN (Table 1); (ii) the National GeoPortal of Italy (http://www.pcn.minambiente.it/mattm/ accessed on 10 November 2021); (iii) individual data sources and geographical layers produced directly by the Italian Civil Protection service (https://rischi.protezionecivile.

gov.it/it/approfondimento/dataset-nazionale-degli-aggregati-strutturali-italiani accessed on 10 November 2021).

Table 1. List of administrative sources used for the construction of the BODA (last access on 10 November 2021.

Region	Website
Piedmont	https://www.geoportale.piemonte.it/cms/progetti/progetto-mosaicatura-catastale
Aosta Valley	https://mappe.partout.it/pub/geonavitg/geodownload.asp?carta=CTR
Lombardy	https://www.geoportale.regione.lombardia.it
Trentino Alto Adige	http://www.openkat.it/
Veneto	https://idt2.regione.veneto.it/idt/downloader/download
Friuli Venezia Giulia	http://irdat.regione.fvg.it/consultatore-dati-ambientali-territoriali/search
Liguria	https://www.regione.liguria.it/open-data/item/7099-carta-tecnica-regionale-1-5000-dal-2007-ii-edizione-3d-db-topografico.html
Emilia Romagna	https://geoportale.regione.emilia-romagna.it/download/download-data?type=dbtopo
Tuscany	http://www502.regione.toscana.it/geoscopio/
Umbria	http://www.umbriageo.regione.umbria.it/pagina/fabbricati-sistema-ecografico-catastale-regione-um
Marche	https://www.regione.marche.it/Regione-Utile/Paesaggio-Territorio-Urbanistica-Genio-Civile/Cartografia-regionale/Repertorio/Carta-tecnica-numerica-110000/opendata
Latium	http://dati.lazio.it/catalog/it/dataset/carta-tecnica-regionale-2002-2003-5k-roma
Abruzzo	http://opendata.regione.abruzzo.it/content/dbtr-regione-abruzzo-scala-15000-edizione-2007-formato-shp
Molise	Not available
Campania	http://sit.cittametropolitana.na.it/downloads.php
Apulia	http://www.sit.puglia.it/portal/portale_cartografie_tecniche_tematiche/Download
Basilicata	http://dati.regione.basilicata.it/catalog/dataset/database-topografico-tema-edificato
Calabria	https://sciamlab.com/opendatahub/dataset/regcal_ctr5k-dbt
Sicily	https://www.sitr.regione.sicilia.it/download/download-carta-tecnica-2000/
Sardinia	http://www.sardegnageoportale.it/areetematiche/databasegeotopografico/

Despite the extensive technical efforts, it was not always possible to monitor the quality of all administrative sources below the spatial level of regional authorities (e.g., provinces, municipalities). Only for Campania region (Southern Italy), the regional data were integrated with provincial data. By contrast, the Molise region did not provide administrative data for any territorial entity. For these reasons, the regional technical maps (CTR) identifying individual buildings with a relatively high spatial resolution were considered to achieve a complete national coverage. The National GeoPortal (GPN)—an application of the Italian Ministry of the Environment, Territory, and Sea (MATTM), where buildings in specific areas of the country (e.g., provincial capital towns, and formalized urban centers) were extensively mapped—was used to integrate these information sources. However, this is only a small part of the information typical of land (cadastral) registers in Italy. Considering that BODA surveyed aggregates of buildings, and not the individual buildings, this database was implemented taking account of the following issues: (i) all individual buildings surveyed from administrative and open data sources were considered in the final BODA archive; (ii) all the buildings derived from GPN that were not contiguous with any building surveyed in the archive were included in the archive and, finally, (iii) structural aggregates ("ensembles of buildings") not intersecting any building in the archive were also included in the archive. Despite these technical adjustments, the geometric information derived from the regional technical maps is the dominant layer in the final BODA. Finally, it should be noted that, because of a recent hacker attack, the availability of administrative data produced and disseminated by the Latium Region, Central Italy, has been greatly reduced. The regional authority had in fact suspended the operation of the regional GeoPortal. We, therefore, resorted to copies of these administrative data available at other sources.

2.3. The Spatial Distribution of Buildings Derived from Official Statistics in Italy

To support census operations, ISTAT partitioned the Italian territory into more than 400 thousand polygons ("enumeration districts") classified into four types of settlements, from LOC1 to LOC4 (for more details regarding the segmentation of the Italian territory into enumeration districts, see the technical notes of general censuses at www.istat.it website accessed on 30 December 2021). More precisely, the 2011 census database (the last census wave in Italy) was organized into (i) 271,229 LOC1 districts, including strictly built-up areas where buildings are located at a maximum distance of 70 m from each other (64,826 km^2); (ii) 41,306 LOC2 districts of moderately dense built-up areas with buildings at a maximum distance of 30 m from each other—including peri-urban and rural dense settlements, and (iii) 89,586 LOC3 and LOC4 districts, respectively, with productive settlements (industry and services) and sparse settlements in rural areas covering the remaining part of Italy.

2.4. A Comparative Analysis of OSM and BODA Coverage in Italy

The spatial partition illustrated in Section 2.3 was used in the present study as a stratification variable for comparison of OSM and BODA building map coverage in Italy. More specifically, the stratification variable distinguished strictly urban, LOC1 districts from LOC2-3-4 districts, hereafter referred to as "rural", based on the distance between buildings. Spatial stratification may provide additional information about the coverage of OSM database with respect to an official source such as BODA. Based on this stratification (urban vs. rural), a comparative analysis of OSM and BODA coverage was performed at the regional scale in Italy (20 regional authorities), considering together (i) the total number of buildings surveyed in OSM and BODA separately and (ii) the total surface area of surveyed buildings in both data sources. It is worth noting that different sources may interpret the same building differently. In particular, the editing process in OpenStreetMap may lead to the consideration of two or more adjacent buildings as a single polygon, e.g., because of an insufficient resolution of the selected ortho-photo-images. This is the case of semi-detached buildings (e.g., villas), which can be recorded as two (smaller) buildings within the administrative archive. Figure 3 provides two examples taken from the spatial overlap of OSM and BODA building shapes, documenting how buildings surveyed within administrative sources can be divided into different single entities with respect to the reality. Only a field inspection (or detailed interpretation of aerial photographs) may clarify this issue which is, at the same time, one of the possible explanations of the different coverage of building maps frequently in urban and rural districts [10].

Figure 3. (**Left**) A visual example of the comparison of building data derived from the National Geo-Portal (GPN_12_ED) and the OSM database (OSM_12) in a strictly urban location of Rome, Central Italy (red and black indicate the building polygons identified through OSM and BODA, respectively); the full coverage of OSM with respect to the BODA archive is documented here. (**Right**) Spatial overlap of the content of OSM and BODA archives: an example from Frosolone, a rural municipality in Molise, Southern Italy; the partial coverage of OSM with respect to BODA is documented here.

Based on these premises, three different approaches were adopted in our study: (i) an aggregate, count-based comparison between OSM and BODA databases at the regional scale (i.e., considering the absolute number of buildings with the same characteristics in both archives), (ii) a refined count of individual buildings present in both databases based on centroid spatial superposition between OSM and BODA databases; and, finally, (iii) an "object-oriented" approach based on the calculation of building surfaces in OSM and BODA based on polygon intersection. For case (iii), an index computing the ratio between the intersect area and the reference area was made available [10]; this ratio equals 1 when the building surface in the two databases coincides. Results of these approaches were summarized at the regional scale in Italy, considered an appropriate domain of analysis for the implementation of official statistics [28]. Italian regions were ordered from north to south, following the unique identification code provided by ISTAT and ranging from 1 (Piedmont) and 20 (Sardinia). This coding system is used at the national and European level (i.e., Eurostat) within the Nomenclature of Territorial Statistical Units (NUTS).

3. Results

A comparative analysis of the content of OSM and BODA databases (Table 2) documents differences as far as the number of buildings and the surface area of buildings are concerned, suggesting relevant disparities in the coverage rate among regions in Italy. As expected, buildings in urban areas (LOC1 according with ISTAT classification) were more abundant than buildings located in rural areas (LOC2-3-4) for both data sources. A total of 14,121,227 buildings were included in the OSM repository for a total area of 3640 km^2. A total of 25,981,541 buildings were enumerated in the BODA covering 5474 km^2.

3.1. Comparing the Coverage of Building Surveys Derived from Different Data Sources

The content of OSM and BODA databases was compared according to empirical approaches presented in an earlier study (Hecht, 2013). The first one was based on aggregate counts of the elements surveyed in the different databases, e.g., by administrative region. However, this count does not guarantee that the two archives refer to the same building entity. The second step implied counts of the building centroids in the reference database (BODA) that overlap buildings' polygons in the OSM database. This step is more detailed than the first one, but it does not consider whether the elements found in the same places are exactly the same entity. The third step included a specific analysis of the overlapping surface area of buildings surveyed in the two archives.

Table 2. Statistical distribution of buildings in OSM and BODA databases for Italy, by urban/rural district and administrative region.

Region	Total Number of Buildings					Surface Area of Buildings				
	Urban	Rural	Total	Urban (%)	Rural (%)	Urban	Rural	Total	Urban (%)	Rural (%)
	OSM archive									
Piedmont	688,877	450,615	1,139,492	60.4	39.6	194.52	106.82	301.34	64.5	35.4
Aosta Valley	37,094	34,868	71,962	51.6	48.4	7.80	5.14	12.93	60.3	39.7
Lombardy	1,239,227	273,940	1,513,167	81.9	18.1	424.04	102.20	526.24	80.6	19.4
Trentino Alto Adige	203,743	134,095	337,838	60.3	39.7	61.63	31.31	92.93	66.3	33.7
Veneto	1,664,300	793,503	2,457,803	67.7	32.3	349.98	157.48	507.46	69.0	31.0
Friuli Venezia Giulia	479,667	146,255	625,922	76.6	23.4	108.24	36.14	144.38	75.0	25.0
Liguria	350,365	171,120	521,485	67.2	32.8	65.66	15.98	81.64	80.4	19.6
Emilia Romagna	790,617	516,489	1,307,106	60.5	39.5	252.23	135.08	387.31	65.1	34.9
Tuscany	889,468	435,144	1,324,612	67.1	32.8	187.38	75.65	263.03	71.2	28.8
Umbria	68,474	41,221	109,695	62.4	37.6	27.33	12.14	39.48	69.2	30.8
Marche	92,096	66,264	158,360	58.2	41.8	32.05	18.41	50.47	63.5	36.5
Latium	491,448	183,719	675,167	72.8	27.2	171.02	69.44	240.46	71.1	28.9
Abruzzo	107,117	33,876	140,993	76.0	24.0	36.07	13.96	50.03	72.1	27.9
Molise	29,012	22,203	51,215	56.7	43.3	7.21	6.12	13.33	54.1	45.9
Campania	404,567	89,760	494,327	81.8	18.2	145.90	30.47	176.37	82.7	17.3
Apulia	698,835	896,943	1,595,778	43.8	56.2	227.63	131.55	359.18	63.4	36.6
Basilicata	72,385	85,295	157,680	45.9	54.1	17.74	15.87	33.61	52.8	47.2
Calabria	155,854	43,101	198,955	78.3	21.7	35.84	12.61	48.45	74.0	26.0
Sicily	529,782	119,042	648,824	81.6	18.3	117.77	35.70	153.46	76.7	23.3
Sardinia	317,302	273,544	590,846	53.7	46.3	102.31	55.36	157.67	64.9	35.1
Italy	9,310,230	4,810,997	14,121,227	65.9	34.1	2572.35	1067.43	3639.77	70.7	29.3

Table 2. *Cont.*

Region	Total Number of Buildings					Surface Area of Buildings				
	Urban	Rural	Total	Urban (%)	Rural (%)	Urban	Rural	Total	Urban (%)	Rural (%)
					BODA archive					
Piedmont	874,336	706,198	1,580,534	55.3	44.7	307.10	184.13	491.23	62.5	
Aosta Valley	39,858	39,917	79,775	50.0	50.0	9.20	5.78	14.98	61.4	
Lombardy	2,811,587	796,441	3,608,028	77.9	22.1	605.92	179.67	785.59	77.1	
Trentino Alto Adige	273,803	215,790	489,593	55.9	44.1	134.80	96.95	231.75	58.2	
Veneto	1,636,588	759,192	2,395,780	68.3	31.7	332.25	135.17	467.42	71.1	
Friuli Venezia Giulia	2,000,977	438,751	2,439,728	82.0	18.0	297.32	79.62	376.94	78.9	
Liguria	371,287	172,686	543,973	68.2	31.7	64.83	18.29	83.12	78.0	
Emilia Romagna	1,276,852	925,642	2,202,494	58.0	42.0	254.48	157.00	411.48	61.9	
Tuscany	1,367,632	674,136	2,041,768	67.0	33.0	178.55	76.38	254.93	70.0	
Umbria	278,379	245,528	523,907	53.1	46.9	51.03	33.19	84.22	60.6	
Marche	253,176	309,649	562,825	45.0	55.0	79.77	58.18	137.95	57.8	
Latium	941,431	851,858	1,793,289	52.5	47.5	209.32	173.33	382.65	54.7	
Abruzzo	320,073	304,192	624,265	51.3	48.7	66.96	39.07	106.04	63.1	
Molise	54,375	98,734	153,109	35.5	64.5	13.57	13.66	27.23	49.8	
Campania	800,823	730,989	1,531,812	52.3	47.7	228.36	164.43	392.80	58.1	
Apulia	508,183	919,489	1,427,672	35.6	64.4	175.37	103.75	279.12	62.8	
Basilicata	133,259	259,623	392,882	33.9	66.1	26.70	27.52	54.22	49.2	
Calabria	564,671	558,530	1,123,201	50.3	49.7	115.00	53.76	168.76	68.1	
Sicily	821,598	972,595	1,794,193	45.8	54.2	256.68	320.55	577.23	44.5	
Sardinia	327,017	345,696	672,713	48.6	51.4	100.77	45.86	146.64	68.7	
Italy	15,655,905	10,325,636	25,981,541	60.3	39.7	3507.99	1966.32	5474.31	64.1	

3.1.1. Count-Based Comparison between OSM and BODA Databases

The comparison was based on two indicators: (i) number of OSM buildings to the number of buildings found in BODA within the same region; (ii) total surface (m^2) of buildings in OSM database in total surface (m^2) of buildings in BODA within the same region. Considering separately the number of buildings and the related surface area may discriminate contexts where users interpret contiguous buildings as a single building. Since administrative archives—such as BODA—are mainly based on cadaster information while OSM users distinguish buildings by their roof, discrepancies are rather inevitable and should be, at least indirectly, estimated. The indicator quantifying OSM-to-BODA coverage rate for both the number of buildings and the surface area of buildings is indicative of such condition, with the former being 54.4 and the latter reaching 66.5 for Italy (Table 3). These values indicate a better representation of buildings in the OSM database as far as surface area is concerned and can be also explained with the possibly different definition/interpretation of individual/contiguous buildings adopted in OSM and BODA sources. Comparing indicators by administrative region allows the identification of Italian regions with a larger OSM spatial coverage. A separate analysis of urban and rural settlements revealed a particularly heterogeneous context, since the OSM database enumerates a number of urban buildings equal to about 59% of the buildings enumerated in BODA, compared with 47% of the rural buildings. This may confirm an earlier assumption regarding the different OSM coverage in urban and rural areas. Results of the empirical analysis indicate Apulia and Veneto as significant cases where urban and rural buildings, respectively, were better represented in the OSM database compared with BODA source. Notably, the official building map of Apulia (BODA) is recognized as quite obsolete, and OSM elements may contribute significantly to updating this map. In such cases, building records in the OSM that are not surveyed in administrative sources cannot be considered wrong/inappropriate, since a part of the regional maps is dated, and informal buildings are still present in some areas.

Table 3. OSM-to-BODA coverage ratio (calculated separately for the total number of buildings and the total surface area of buildings) by Italian region and urban/rural district.

Region	Total Number of Buildings			Surface Area of Buildings		
	Urban	Rural	Total	Urban	Rural	Total
Piedmont	78.79	63.81	72.10	63.34	58.01	61.34
Aosta Valley	93.07	87.35	90.21	84.71	88.90	86.33
Lombardy	44.08	34.40	41.94	69.98	56.88	66.99
Trentino Alto Adige	74.41	62.14	69.00	45.72	32.29	40.10
Veneto	101.69	104.52	102.59	105.34	116.50	108.56
Friuli Venezia Giulia	23.97	33.33	25.66	36.41	45.39	38.30
Liguria	94.37	99.09	95.87	101.28	87.39	98.22
Emilia Romagna	61.92	55.80	59.35	99.11	86.04	94.13
Tuscany	65.04	64.55	64.88	104.95	99.04	103.18
Umbria	24.60	16.79	20.94	53.56	36.58	46.87
Marche	36.38	21.40	28.14	40.19	31.65	36.58
Latium	52.20	21.57	37.65	81.70	40.06	62.84
Abruzzo	33.47	11.14	22.59	53.87	35.73	47.18
Molise	53.36	22.49	33.45	53.11	44.79	48.93
Campania	50.52	12.28	32.27	63.89	18.53	44.90
Apulia	137.52	97.55	111.77	129.80	126.80	128.68
Basilicata	54.32	32.85	40.13	66.45	57.65	61.99
Calabria	27.60	7.72	17.71	31.16	23.45	28.71
Sicily	64.48	12.24	36.16	45.88	11.14	26.59
Sardinia	97.03	79.13	87.83	101.53	120.71	107.53
Italy	59.47	46.59	54.35	73.33	54.29	66.49

3.1.2. Building Centroid Superposition between OSM and BODA Databases

Since the two databases may indicate different (building) objects, it is appropriate to investigate further which types of OSM cartographic objects are found in BODA. For this purpose, we calculated how many centroids of BODA fell within OSM buildings. From this perspective, we calculated the "overlap share" [10] in accordance with local spatial patterns. The methodology consists in counting the buildings of the reference database that are located within the centroid of a building of the examined database, the OSM in this case (Figure 4).

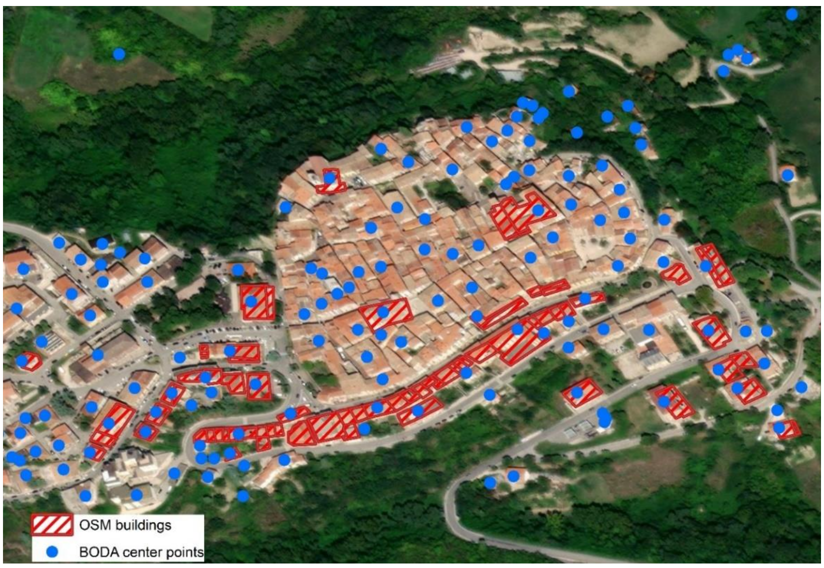

Figure 4. An example of BODA building centroid overlay on OSM building database for the municipality of Frosolone, Molise (Southern Italy).

Apulia and Veneto regions ranked in the first positions together with Liguria (93.2%) and Friuli–Venezia–Giulia (88.2%). These results also document the regional heterogeneity in the building representation typical of OSM and BODA databases (Table 4).

Table 4. (**Left**) Percent share of BODA buildings found in the OSM archive in total BODA coverage; (**right**) percent share of OSM building area that overlaps with BODA building area in Italy in total OSM building area, by administrative region and urban/rural district.

Region	BODA Buildings in OSM Archive			BODA Building Overlap with OSM		
	Urban	Rural	Total	Urban	Rural	Total
Piemonte	47.2	38.5	43.4	45.9	38.3	43.0
Aosta Valley	64.1	51.4	58.0	60.2	47.3	55.2
Lombardy	41.7	30.5	39.4	55.3	35.5	50.8
Trentino Alto Adige	64.8	47.2	57.2	35.5	19.8	29.0
Veneto	80.7	74.1	78.6	77.3	70.3	75.2
Friuli Venezia Giulia	88.9	84.9	88.2	58.0	50.3	56.4
Liguria	93.5	92.7	93.2	82.3	65.9	78.7
Emilia Romagna	77.3	62.6	71.2	80.8	67.2	75.6
Tuscany	74.8	59.6	69.9	80.4	63.8	75.4
Umbria	26.7	13.3	20.5	37.7	20.3	30.8
Marche	28.5	18.4	23.0	27.7	19.7	24.3
Latium	43.4	16.7	30.8	54.9	26.0	41.8
Abruzzo	24.0	8.1	16.4	33.4	20.2	28.5
Molise	29.5	13.4	19.1	33.7	24.1	28.9
Campania	35.2	8.2	22.4	43.7	11.4	30.1
Apulia	93.8	78.7	83.9	96.0	85.1	91.9
Basilicata	36.0	0.0	36.0	55.0	51.8	53.4
Calabria	16.6	5.7	11.1	19.8	13.3	17.7
Sicily	28.8	8.8	17.8	33.1	8.1	19.2
Sardinia	83.4	71.1	77.0	79.7	97.9	85.3
Italy	58.3	40.7	51.5	57.4	37.2	50.2

3.1.3. An "Object-Oriented" Approach Based on Polygon Intersection

The third approach was more refined in summarizing the level of completeness and correct positioning of the buildings in the OSM archive with respect to the administrative data (Figure 5). The largest intersection surface area was observed in Apulia (92%). The main advantage of this approach is that it does not take account of the issue of the semantic definition of "building" as a cartographical object, as it is substantially different among databases.

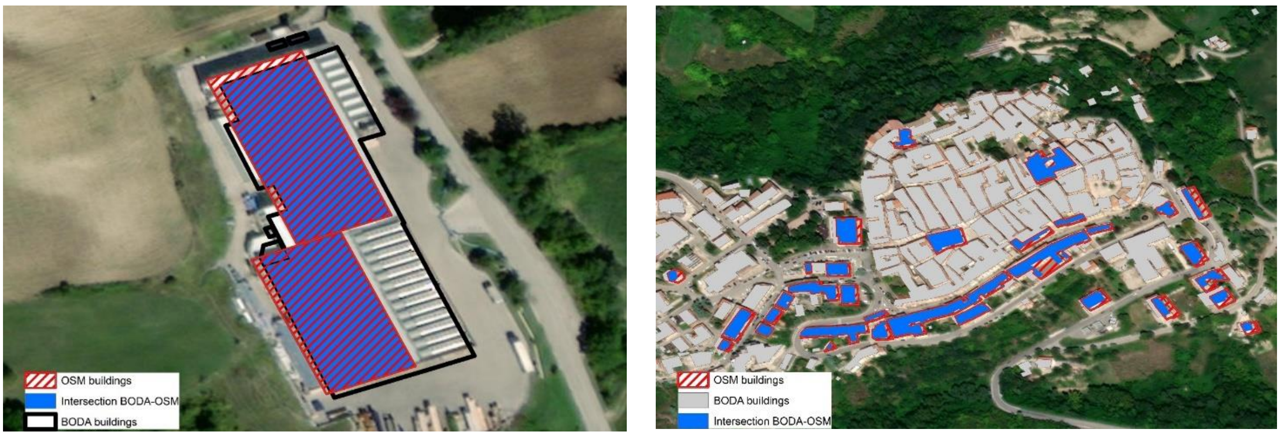

Figure 5. Examples of building intersection and the calculation of the related built-up surface area in Frosolone, Molise.

4. Conclusions

The analysis carried out in this study provides indirect support to the continuous improvement of OSM. At the same time, a comparative analysis of OSM coverage across Italian regions outlines the uncertainty of the reference data and the difficulty of identifying heterogeneity sources, both in the case of underestimations or in the case of overestimations of building numbers. One of the possible sources of heterogeneity lies in the fact that data entry into the OSM takes place in different ways, including data collection in the field, remote mapping, and direct import from independent sources. Direct import can be sometimes generate some overlaps between OSM and reference databases, generating an additional cause of data instability. Monitoring the progressive consolidation of OpenStreetMap contents (i.e., scrutinizing the stratification of input data sources forming the actual map coverage)—possibly distinguishing between individual imputation of primary (e.g., field) data, remote mapping, and direct import from external, secondary sources—is a particularly hard but challenging task. In Italian regions identified to have more accurate OSM databases, the different sources of heterogeneity could have acted even more intensively than in other socioeconomic contexts. The Apulia region may be considered a representative example of such dynamics. Further investigations on the impact of data source heterogeneity on OSM completeness for statistical use are particularly required in such contexts.

Despite the latent uncertainty in the autonomous use of OSM archives as a direct input to official statistics, our study justified further efforts exploring the role of open databases in large-scale cartographic assessments of individual buildings with enough spatial and semantic accuracy. While the empirical results of this study seem to discourage the use of OSM as a substitute of BODA, we demonstrate that, at least in certain areas of Italy, OSM may represent an authoritative source of information for the spatial distribution of different types and sizes of buildings—contributing to the verification of the quality and precision of official sources especially in specific contexts. Assuming a less intense coverage (with respect to that of official sources) of OSM in rural areas, the empirical results of our study are also of some interest for users who want to start mapping buildings in under-cover areas. A work of this type, however, requires special attention because not all the regional authorities in Italy, as well as in other European contexts, grant data licenses compatible with OSM data. Future research should finally concentrate on the implementation of new indicators testing the completeness of OSM data, considering the intrinsic characteristics of buildings as supplementary, but not secondary, information. The final objective of such studies is clearly the improvement of the geo-coding mechanisms' precision, so that the OSM archive will become more reliable for both administrative and statistical purposes.

Author Contributions: Conceptualization, Gianluigi Salvucci; methodology, Gianluigi Salvucci; software, Gianluigi Salvucci; validation, Gianluigi Salvucci; formal analysis, Gianluigi Salvucci; investigation, Gianluigi Salvucci; resources, Luca Salvati; data curation, Gianluigi Salvucci; writing—original draft preparation, Gianluigi Salvucci and Luca Salvati; writing—review and editing, Gianluigi Salvucci and Luca Salvati; visualization, Gianluigi Salvucci; supervision, Luca Salvati; project administration, Luca Salvati; funding acquisition, Luca Salvati. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Official statistics released by the Italian National Statistical Institute (ISTAT) were used in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Girres, J.F.; Touya, G. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* **2010**, *14*, 435–459. [CrossRef]
2. Mondzech, J.; Sester, M. Quality analysis of OpenStreetMap data based on application needs. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **2011**, *46*, 115–125. [CrossRef]
3. Fan, H.; Zipf, A.; Fu, Q.; Neis, P. Quality assessment for building footprints data on OpenStreetMap. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 700–719. [CrossRef]
4. Ma, D.; Sandberg, M.; Jiang, B. Characterizing the heterogeneity of the OpenStreetMap data and community. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 535–550. [CrossRef]
5. Goodchild, M. Citizens as Sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [CrossRef]
6. Turner, A. Introduction to Neogeography. 2006. Available online: https://proquestcombo.safaribooksonline.com/0596529953 (accessed on 30 November 2021).
7. Salvati, L. Towards a Polycentric Region? The Socio-economic Trajectory of Rome, an 'Eternally Mediterranean' City. *Tijdschr. Voor Econ. En Soc. Geogr.* **2014**, *105*, 268–284. [CrossRef]
8. Basiouka, S.; Potsiou, C.; Bakogiannis, E. OpenStreetMap for cadastral purposes: An application using VGI for official processes in urban areas. *Surv. Rev.* **2015**, *47*, 1752270615Y.000. [CrossRef]
9. Haklay, M.; Weber, P. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Comput.* **2008**, *7*, 12–18. [CrossRef]
10. Hecht, R. Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 1066–1091. [CrossRef]
11. Zambon, I.; Colantoni, A.; Salvati, L. Horizontal vs vertical growth: Understanding latent patterns of urban expansion in large metropolitan regions. *Sci. Total Environ.* **2019**, *654*, 778–785. [CrossRef]
12. Duvernoy, I.; Zambon, I.; Sateriano, A.; Salvati, L. Pictures from the other side of the fringe: Urban growth and peri-urban agriculture in a post-industrial city (Toulouse, France). *J. Rural Stud.* **2018**, *57*, 25–35. [CrossRef]
13. Zambon, I.; Benedetti, A.; Ferrara, C.; Salvati, L. Soil matters? A multivariate analysis of socioeconomic constraints to urban expansion in Mediterranean Europe. *Ecol. Econ.* **2018**, *146*, 173–183. [CrossRef]
14. Salvati, L.; Ciommi, M.T.; Serra, P.; Chelli, F.M. Exploring the spatial structure of housing prices under economic expansion and stagnation: The role of socio-demographic factors in metropolitan Rome, Italy. *Land Use Policy* **2019**, *81*, 143–152. [CrossRef]
15. Pili, S.; Grigoriadis, E.; Carlucci, M.; Clemente, M.; Salvati, L. Towards sustainable growth? A multi-criteria assessment of (changing) urban forms. *Ecol. Indic.* **2017**, *76*, 71–80. [CrossRef]
16. Kazemzadeh-Zow, A.; Zanganeh Shahraki, S.; Salvati, L.; Samani, N.N. A spatial zoning approach to calibrate and validate urban growth models. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 763–782. [CrossRef]
17. Carlucci, M.; Chelli, F.M.; Salvati, L. Toward a new cycle: Short-term population dynamics, gentrification, and re-urbanization of Milan (Italy). *Sustainability* **2018**, *10*, 3014. [CrossRef]
18. Fonte, C.C.; Patriarca, J.A.; Minghini, M.; Antoniou, V.; See, L.; Brovelli, M.A. Using OpenStreetMap to Create Land Use and Land Cover Maps Development of an Application. In *Volunteered Geographic Information and the Future of Geospatial Data, edited by Cláudio Elízio Calazans Campelo, Michela Bertolotto, Padraig Corcoran*; IGI Global: Amsterdam, The Netherlands, 2019; pp. 1100–1123.
19. Ahmouda, A.; Hochmair, H.H.; Cvetojevic, S. Analyzing the effect of earthquakes on OpenStreetMap contribution patterns and tweeting activities. *Geo-Spat. Inf. Sci.* **2018**, *21*, 195–212. [CrossRef]
20. Sui, D.Z. The Wwikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Comput. Environ. Urban Syst.* **2008**, *32*, 1–5. [CrossRef]
21. Bajocco, S.; Ceccarelli, T.; Smiraglia, D.; Salvati, L.; Ricotta, C. Modeling the ecological niche of long-term land use changes: The role of biophysical factors. *Ecol. Indic.* **2016**, *60*, 231–236. [CrossRef]
22. Salvati, L.; Zitti, M.; Sateriano, A. Changes in city vertical profile as an indicator of sprawl: Evidence from a Mediterranean urban region. *Habitat Int.* **2013**, *38*, 119–125. [CrossRef]
23. Brovelli, M.A.; Minghini, M.; Molinari, M.E.; Zamboni, G. Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection: The method and a case study. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *49*, 615–620. [CrossRef]
24. Franzini, M.; Annovazzi-Lodi, L.; Casella, V. Assessment of the Completeness of OpenStreetMap and Google Maps for the Province of Pavia (Italy). In Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management, Prague, Czech Republic, 7–9 May 2020; pp. 270–277. [CrossRef]
25. Ciommi, M.; Chelli, F.M.; Carlucci, M.; Salvati, L. Urban growth and demographic dynamics in southern Europe: Toward a new statistical approach to regional science. *Sustainability* **2018**, *10*, 2765. [CrossRef]
26. Ciommi, M.; Chelli, F.M.; Salvati, L. Integrating parametric and non-parametric multivariate analysis of urban growth and commuting patterns in a European metropolitan area. *Qual. Quant.* **2019**, *53*, 957–979. [CrossRef]
27. Bajocco, S.; Dragoz, E.; Gitas, I.; Smiraglia, D.; Salvati, L.; Ricotta, C. Mapping forest fuels through vegetation phenology: The role of coarse-resolution satellite time-series. *PLoS ONE* **2015**, *10*, e0119811. [CrossRef] [PubMed]
28. Salvati, L.; Guandalini, A.; Carlucci, M.; Chelli, F.M. An empirical assessment of human development through remote sensing: Evidences from Italy. *Ecol. Indic.* **2017**, *78*, 167–172. [CrossRef]