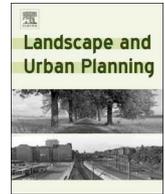




ELSEVIER

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Administrative boundaries and urban areas in Italy: A perspective from scaling laws

Massimiliano Alvioli (Massi)*

Istituto di Ricerca per la Protezione Idrogeologica, Consiglio Nazionale delle Ricerche, via Madonna Alta 126, I-06128 Perugia, Italy



ARTICLE INFO

Keywords:

Urban scaling
Power law
City delineation
Administrative boundaries
Urban planning

ABSTRACT

Delineating boundaries of urban areas is no easy task, due to the inherent complexity of the problem, heterogeneity of relevant data and little consensus on how to properly measure the results. Any such delineation must eventually be cast onto administrative boundaries, an essential requirement for real-world applications. In the effort of relating administrative and alternative boundaries, we investigated in Italy the validity of general scaling laws, such as the area-population relation, and proposed a practical application. Relying on open data for population, settlements and road networks, we showed the extent to which scaling relations hold for different boundaries for urban areas, and how they compare to each other. We considered, beside Italian municipalities, urban areas based on the idea of “natural cities”, obtained using head/tail breaks of areas related to human mobility as an explicit indicator of existence of a city. Area-population data for administrative boundaries can be reconciled with scaling relations valid for both the world’s cities data and with those obtained from natural cities, provided an effective area is adopted in place of polygon planimetric area of municipalities. We eventually proposed an aggregation of administrative units using the empirical scaling relation as an objective function for accepting or rejecting pairwise fusion of boundaries. We suggest considering such a method, along with expert considerations, as an additional tool for real-world urban planning as seen from the very general perspective of seemingly abstract scaling laws.

1. Introduction

Virtually all fields of science have investigated and identified scaling laws for many years. Studies based on very general laws, such as the Zipf’s one regarding the rank-frequency distributions in physical and social phenomena following a power-law (Zipf, 1949), and Gibrat’s law, describing rates of growth of different phenomena such as private firms and cities (Gibrat, 1931), date back to the first half of the 20th century. Similar relations were soon found to hold for a wide range of phenomena, and investigated in a substantially large body of literature. Geo-spatial, human geography and urban sciences make no exception: examples of studies of city size distributions as power laws can be found as early as in 1913 (Auerbach, 1913). Fang and Yu (2017) recently compiled a review of more than 32,000 research papers on the topic of urban agglomeration studies.

Scaling relations regarding urban systems imply that there is no special size for a city, nor a special stage of development: scaling relations tend to be universal. A scaling relation for a given set of urban systems implies that all of those cities are a scaled version of one another, which makes their shape, size and expansion predictable (Batty,

2008; Bettencourt & West, 2010). Despite the large number of studies about universal properties of urban systems, and claims of potential applications for planning, though, practical applications seems to be hard to obtain, with notable exceptions (Barthélemy, Bordin, Berestycki, & Gribaudo, 2013; Furtado, 2015). Most of the existing studies aim at showing the validity of scaling relations, through extensive data gathering from heterogeneous sources (Bettencourt, Lobo, Helbing, Kühnert, & West, 2007; Jiang & Liu, 2012; Fang & Yu, 2017; Long, Zhai, Shen, & Ye, 2018) and development of dedicated analytical and statistical methods (Warton, Wright, Falster, & Westoby, 2006; Clauset, Shalizi, & Newman, 2009; Taskinen & Warton, 2011).

A notable effort has been devoted to the very delineation of urban boundaries, which is crucial to any further speculation about the properties of urban systems. Boundaries determine not only the size but also the physical form of cities which, in turn, is deeply related to their inter-urban and intra-urban function. Social implications of city forms stemming from plans and mobility within neighborhoods were discussed by Park (1915). A modern view of the relation between form and function was clearly stated by Batty and Longley (1994), who recognized the fractal geometry of cities and identified them in

* Corresponding author.

E-mail address: massimiliano.alvioli@irpi.cnr.it.<https://doi.org/10.1016/j.landurbplan.2020.103906>

Received 7 August 2019; Received in revised form 8 July 2020; Accepted 23 July 2020

0169-2046/© 2020 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

mathematical terms with population density functions. More recently Venables (2017) argued that the relation of form and function depends on the evolution in time of cities and on the economic interactions in which cities are involved during their development. This is a view of cities as complex entities characterized by emergent phenomena, whose shape is not well defined. The metrics to estimate urban forms are also object of debate (Vanderhaegen & Canters, 2017).

A number of different approaches to city boundary delineation exist in the literature, which we will review and discuss in Section 2. Nevertheless, in order to aim at any practical application of scientifically sound models one has to face the fact that, political or administrative boundaries exist in any country of the world and at different geographical scales. Hence, existing political boundaries have to be considered for practical applications of scientific conclusions (Hamilton & Rae, 2018; Thomas, Jones, Caruso, & Gerber, 2018), no matter what method is used to establish geographical boundaries of cities.

In this paper, we assessed the extent to which planimetric area-population ($A-P$) scaling laws holds in Italy, with respect to both urban boundaries and administrative subdivisions. We further applied scaling relations to propose a fusion of a large sub set of administrative boundaries on the very general grounds provided by $A-P$ relations. We considered Italian municipalities, the national administrative subdivision of finest granularity (8,092 polygons in year 2011, see Section 3). Next we considered one of the existing definitions of urban areas, based on the idea of “natural cities”. They can be obtained from nodes of communication networks, as originally proposed by (Rozenfeld, Rybski, Gabaix, & Makse, 2009) and recently applied in a range of different study areas and data by Jiang and Jia (2011) and Jiang and Liu (2012). Such definition of cities, or urban areas, imply different measures of planimetric areas and population distributions with respect to the ones encompassed by administrative boundaries, which we describe in detail in Section 4.

To our knowledge, we performed the first delineation of modern city boundaries in Italy, based on one of the methods existing in the literature, and calculated scaling exponents for $A-P$ relations. A study of scaling relations in a few European countries, including Italy, was previously performed by Bettencourt and Lobo, 2016. Bertuglia, Bianchi, and Mela, 2012 represents a notable reference to urban studies in Italy, though not specifically focused on scaling relations but more in general on a science of cities, of the methods and tools used in their analyses, mostly within complexity and self-organization theories.

We performed a selection of pairs of neighbouring municipalities and, using distance from the empirical $A-P$ relation as an objective function, we accepted or rejected the fusion of the two municipalities, were the candidate mergers have aggregate population and area. Mergers were constrained to be within the same higher-level administrative area, namely one the twenty Italian Regions. The results of the analysis outlined above are described in detail in Section 5, and discussed in Section 6. In particular, we discuss the similarities and differences between scaling relations obtained within municipalities' boundaries and within “natural cities”, and we compare with two existing clustering of municipalities operated by various Italian agencies. Conclusions of this work are drawn in Section 7.

2. Background: Scaling of urban variables and delineation of urban areas

The complexity of phenomena occurring within cities, and contributing to their birth and expansion, have been investigated in many respects, since the early studies. Power-law scaling relations are usually studied as a function of the size of cities, taken as their population. Scaling was found to hold for city spatial extent (area), total length of networks (street, piping, wiring), economic and social indicators, and so on (Bettencourt et al., 2007). Infrastructures, such as the total road surface and the total length of electrical cables, show a scaling exponent smaller than unity. On the other hand, quantities related to wealth

creation and innovation, such as new patents, employment and gross product, are represented by power laws with an exponent larger than unity: super linear scaling signals that as the city size increases, such quantities must do so at an ever increasing rate in order to sustain system development.

To make sense of urban scaling, Bettencourt (2013) reported about a comprehensive mathematical framework based on a set of basic principles. The complex dynamics of spatial, infrastructural and social variables could be formalized defining a net urban output function. It was used as a measure of urban efficiency, and its maximization could distinguish between optimal and sub optimal cities. The approach was inspired by the explanation of universal scaling relations in plants (West, Brown, & Enquist, 1999), based on resources being distributed through the system by branching networks. That was later generalized to other living systems (West, 2004) and other domains, including urban systems (West, 2017).

Alternative explanations of urban scaling exists; for example, using cellular automata, WWhite and Engelenhite and Engelen (1993) and Rybski, Ros, and Kropp (2013) simulated city size and fractality (both widely used to explain self-organized systems) within a percolation model. Recently, Liu, Batty, Wang, and Corcoran (2019) studied several aspects of urban change and, most notably, advocated adoption of cellular automata-based models by politicians and practitioners.

So far, no standardized international criteria exist for determining the boundaries of a city and often multiple boundary definitions are available for any given city (United Nations, 2018). As a matter of fact, United Nations (2018) categorized urban areas in three nested levels, in order of increasing sizes: *City Proper*, *Urban Agglomeration* and *Metropolitan Area*, corresponding to different areal and population sizes. Other definitions exist in the literature and in practitioners work, often contrasting with the definitions above. For example, Fang and Yu (2017) consider an “urban agglomeration” as a system of “integrated cities”, and surveyed 32,231 studies published starting from the beginning of 20th century, in which the same spatial organizations has been termed either as “megalopolis”, “city group” and “city cluster”.

Correspondingly, the delineation of cities has been performed in a number of different ways, relying on very heterogeneous data sources, even if no general consensus has been reached on how to delineate a city (Masucci, Arcaute, Hatna, Stanilov, & Batty, 2015). Batty (2018) distinguished city delineation methods by three different criteria: (a) population/urbanization density, (b) interactions, described by different kinds of networks, and (c) geographical proximity/contiguity. We reviewed a sub set of them, given then large number of existing relevant studies, and adopted in this work one of the existing methods.

Differences in city delineation methods imply different levels of aggregation of variables in social, economic and spatial urban-related analyses. This is known as the modifiable areal unit problem (MAUP, Openshaw, 1984; Manley, 2014), which may show up in any study associated with the use of data aggregated within geographical areas. It has been investigated in various fields of spatial analysis, such as landscape ecology (Jelinski & Wu, 1996), natural hazards (Alvioli et al., 2016; Alvioli, Guzzetti, & Marchesini, 2020) and of course road traffic analysis (Viegas, Martinez, & Silva, 2009) and census data analysis (Flowerdew, 2011). In this work, we did not explicitly consider MAUP effects, though grouping a varying number of municipalities into one single boundary clearly falls into that class of problems.

One seemingly straightforward tool for city delineation, as an example of density-based methods above, is the use of a map of artificial surfaces, detected using remote sensing (Lu, Li, Kuang, & Moran, 2014). One can assume that any impervious (or sealed) surface is part of an urban system (Ma, Omer, Osaragi, Sandberg, & Jiang, 2018; Thomas et al., 2018). In fact impervious surfaces are both built-up and non-built-up, and include a variety of objects that can be identified with human locations or activities (Steidl, Schleicher, & Sannier, 2016). Impervious areas, though, come as raster layers, and the raster grid cells have to be clustered in order to obtain individual cities. This step

introduces the additional difficulty of delineating boundaries between areas who might actually have relations, either spatial or regarding human activities: cities are difficult to delineate, and also difficult to study in isolation (Bettencourt & West, 2010).

Arcaute et al. (2015) studied data from UK census using different density thresholds for both population densities and commuting flow destinations. They looked at scaling laws of a large set of urban indicators, aggregated within the delineated cities. They found that most of the indicators exhibited a scaling exponent (cf. Eq. (1)) very close to unity, thus apparently violating the expected universal regime of either a sub linear scaling for infrastructure and services, a super linear scaling regime for outcomes of social interactions (Bettencourt, 2013) or a linear regime, in principle expected only for goods related to individual human needs (Arcaute et al., 2015). They explained the values of scaling exponents suggesting that two different regimes exist between cities confined within a domestic environment as opposed to those driving international dynamics.

Along a similar line, Cottineau, Hatna, Arcaute, and Batty (2017) performed a sensitivity analysis of scaling relations on the spatial aggregations of census units in France, aggregated in about 5,000 different ways by combining criteria that took into account density, commuting flows and population cutoffs. The general conclusion was that scaling estimations are subject to large variations, and that variations with respect to city delineation criteria are neither random nor universal for all the variables under study. As a consequence, even if Cottineau et al. (2017) did not propose a “good” definition of city to study urban scaling, and they clarified that the methodological choice should be operated depending on the attribute under investigation.

Oliveira, Furtado, Andrade, and Makse (2018) proposed a worldwide model to delineate city boundaries using a population density threshold and a distance threshold representing the mutual commuting distance between populated areas, singled out by a local clustering procedure. They used worldwide census data at about 1 km resolution. They found that Zipf’s law for population holds for their cities, and proposed a list of the 25 most populated cities in the world, according to their algorithm.

Street networks are undoubtedly proxies for human presence and activity. The study of networks and human mobility, for various purposes, is a field of research on its own and we will not attempt to review it here (Barbosa et al., 2018). It suffices to mention that studies of spatial networks and their topological properties are very closely related to urban studies (Barthélemy & Flammini, 2008) and network optimization is relevant for urban planning (Durand, 2007; Barthélemy & Flammini, 2008). Masucci et al. (2015) used intersections of street networks and a parameter-free clustering technique to distinguish between urban streets and rural streets in the UK. Clusters obtained from urban streets actually exhibited scaling behaviour, while rural ones did not, suggesting two distinct phases in the process of urbanization.

Non-physical networks also represent human activities and, thus, the existence of urban areas. For example, Shutter et al. (2018) considered the interrelations among different types of occupations and how they eventually define an overall urban economy in six world’s countries and showed scaling with city size. Yin, Soliman, Yin, and Wang (2017) considered citizen commute patterns obtained from online social network data to define anthropographic boundaries to delineate cities, finding good correspondence with administrative boundaries in the UK. Yin et al. (2017) also noted that, while social networks data may have limitations and biases, they also have the important advantage of allowing time dependent and, at least in principle, real-time analyses (Khodabandelou, Gauthier, & Fiore, 2018). The relevance of non-physical networks, most importantly of digital communications, was also stressed by Batty (2018), who envisaged the next revolution dictating the shape of cities as the digital one, after the industrial and the electrical revolutions.

Eventually, as an example of a method probably falling in the category of proximity-based methods described by Batty (2018) and Jiang

and Liu (2012) described a technique for delineating city boundaries using street networks to single out individual blocks, defined by the portion of space within closed street loops. The collection of the resulting block areas were found to follow a log-normal (fat-tailed) distribution, characterized by a much larger number of small areas with respect to large areas. The mean value of the distribution was used to distinguish “large” from “small” areas. Jiang and Liu (2012) noted that a high smaller-to-large (about 80%–90% to 20%–10%) area ratio qualitatively signals the presence of scaling, which was also studied by means of power-law fits. In this work, we applied a variant of this data-driven, parameter-free method in Italy.

3. Available data

All of the data utilized in this work were either open data available from governmental agencies, data from volunteered geographic information projects or data otherwise available for free on the world wide web (WWW).

We downloaded a vector layer containing 8092 polygons representing Italian municipalities from the WWW site of the Italian National Institute for Statistics (ISTAT, 2017). A database table associated with the vector layer contained data about population, aggregated at municipality level from census data surveyed in 2011, the most recent one. We did not exploit the full resolution of population data at census section level because we are interested in a municipality-level analysis. The next-level political subdivision in Italy is represented by Provinces and Regions. Area, population and number of municipalities in individual Italian Regions are listed in Table 1.

From the ISTAT WWW site we also downloaded an aggregation of municipalities operated by ISTAT itself, based on 2011 census and commuting data, which goes under the name of *Sistemi Locali del Lavoro* (local labour systems, SLL). An additional aggregation of municipalities exists in Italy, which was prepared with different methods and with different aims (that we do not discuss here), with respect to SLL. The second aggregation goes under the name *Unioni di Comuni* (UDC,

Table 1

Brief description of Italian Regions. English name is provided, where existing; “Number” refers to the number of municipalities. Three-letter codes for Regions are used in Figs. 3 and 7.

Code	Region	Population (2011)	Area [km ²]	Number (2011)	Mean Area [km ²]	Pop. Density [km ⁻²]
ABR	Abruzzo	1,307,309	10,796	305	35.4	121.1
BAS	Basilicata	578,036	9992	131	76.3	57.8
CAL	Calabria	1,959,050	15,085	409	36.9	129.9
CAM	Campania	5,766,810	13,600	551	24.7	424.0
EMR	Emilia-Romagna	4,342,135	22,188	348	63.8	195.7
FVG	Friuli-Venezia Giulia	1,218,985	7848	218	36.0	155.3
LAZ	Lazio	5,502,886	17,202	378	45.5	319.9
LIG	Liguria	1,570,694	5,414	235	23.0	290.1
LOM	Lombardy	9,704,151	23,874	1,544	15.5	406.5
MAR	Marche	1,541,319	9712	239	40.6	158.7
MOL	Molise	313,660	4441	136	32.7	70.6
PIE	Piedimont	4,356,904	25,406	1,206	21.1	171.5
PUG	Apulia	4,052,566	19,358	258	75.0	209.3
SAR	Sardinia	1,639,362	24,114	377	64.0	67.9
SIC	Sicily	5,002,904	25,718	390	65.9	194.5
TAA	Trentino-South Tyrol	1,029,475	13,603	333	40.8	75.7
TOS	Tuscany	3,672,202	22,988	287	80.1	159.7
UMB	Umbria	884,268	8455	92	91.9	104.6
VAO	Aosta Valley	126,806	3262	74	44.1	38.9
VEN	Veneto	4,857,210	18,405	581	31.7	263.9

literally associations of municipalities). We compared both SLL and UDC to the aggregation proposed in this work, in Section 6.

We downloaded a continental (European) imperviousness layer, available for various years from the Copernicus Land Monitoring Service (CLMS, 2018). We considered the layer corresponding to year 2012, the closest in time to the census data year, 2011. The raster layer has 20 m x 20 m resolution and values of the raster cells give a measure of the degree of imperviousness, in percentage. In this work we were interested in any sealed surface *i.e.*, with any imperviousness degree different from zero, as a first approximation. Impervious areas included: housing areas, parking lots, railways close to other impervious areas, industrial and commercial areas, various paved and covered soil, and others. They did not include: inactive built-up areas, temporary coverage, sand, dump areas, cultivated areas, and others.

We downloaded road network vector layers from the OpenStreetMap WWW site (OSM, 2018). From these layers we extracted the locations of street intersection to generate natural cities, as described in detail in Section 4.3. The total number of street segments in the data set is 4,063,932 (including overlapping segments across neighbouring Regions), amounting to 8,127,864 intersections.

Eventually, to visually compare scaling relations in Italy with independent area/population data for European and world cities, we downloaded data from the WWW site of the Demographia World Urban Areas, (DWUA, 2018). Data were compiled applying a consistent definition of built-up urban areas, disregarding political boundaries that are generally associated with metropolitan areas or sub-national boundaries. This data set was not crucial since our methodology (Section 4) is not dependent on world's cities data, and neither are our results or conclusions. It was used in the log-log plots in Fig. 1, for illustrative purposes. In the figure, world's cities data are grouped within ten individual countries, five in Europe and five elsewhere. We discuss the plots in Fig. 1 in the next Section.

4. Methods

The ultimate aim of this work is to propose a practical application of long-known urban scaling relations, in particular of area-population relations, for the fusion of existing administrative boundaries. A concise outline of the methods used to achieve such a goal is as follows:

- We considered municipalities, the finest administrative subdivision in Italy, and investigated area-population relations within their boundaries.
- We checked the validity of scaling relations using either the total area of municipalities, A , and a reduced (effective) area, A_{eff} .
- Then, we focused on an independent method for the delineation of city boundaries available in the literature, namely the method of natural cities, based on the idea that street intersections are locations of human activities.
- This objective (parameter-free) delineation of city boundaries was used to extract area-population relations, to be compared with the ones followed by the area and population of individual municipalities.
- Eventually, we singled out candidate municipality polygons for iterative pairwise mergers. Mergers were accepted if the overall area and population within the aggregate polygons were favoured by the empirical A - P relation and overlapped a city, rejected otherwise.

Throughout this work, we committed to the idea that deviations from area-population scaling laws exhibited by some of the points representative of urban areas in the area-population plane (P , A) measure how each city, or urban area alike, deviates from expectations based on its size (Bettencourt, 2013; West, 2017; van Raan, 2019). In fact, a scaling relation of the form:

$$A = \beta P^\alpha, \quad (1)$$

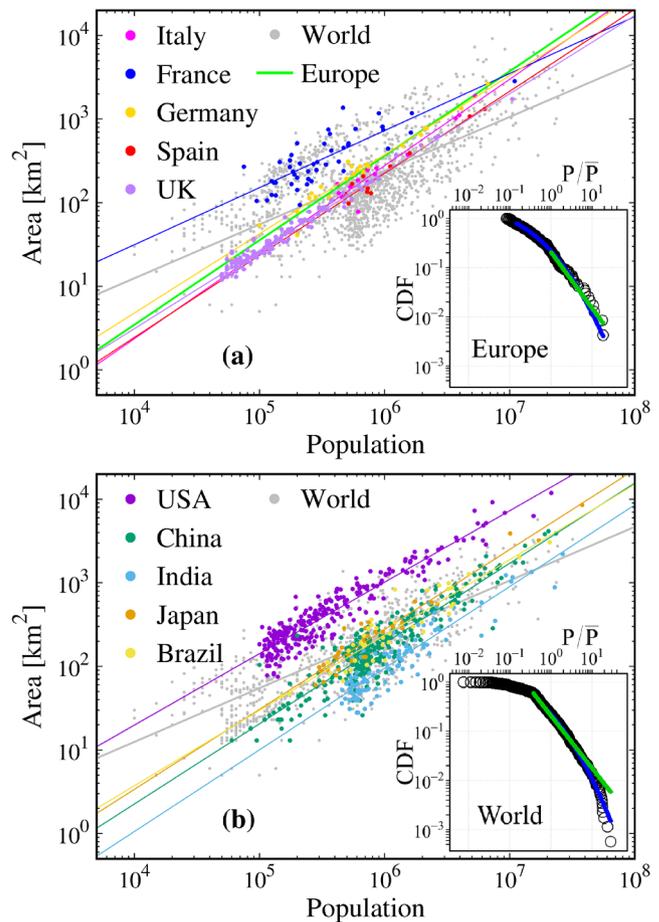


Fig. 1. (a) A - P relations of a few European countries and corresponding linear fits as in Eq. (2); “Europe” refers to the only five countries shown in this Figure. (b) same as (a), for a few world’s countries. “World” refers to the whole set of 1,750 cities. Slopes and correlation coefficients of the fitting curves are: Italy, $\alpha = 1.03 \pm 0.09$, $R^2 = 0.91$; France, $\alpha = 0.68 \pm 0.07$, $R^2 = 0.66$; Germany, $\alpha = 0.95 \pm 0.04$, $R^2 = 0.95$; Spain, $\alpha = 0.98 \pm 0.07$, $R^2 = 0.94$; UK, $\alpha = 0.94 \pm 0.01$, $R^2 = 0.97$; USA, $\alpha = 0.85 \pm 0.02$, $R^2 = 0.89$; China, $\alpha = 0.96 \pm 0.02$, $R^2 = 0.85$; India, $\alpha = 0.98 \pm 0.04$, $R^2 = 0.81$; Japan, $\alpha = 0.96 \pm 0.04$, $R^2 = 0.93$; Brazil, $\alpha = 0.90 \pm 0.06$, $R^2 = 0.84$; Europe, $\alpha = 1.01 \pm 0.03$, $R^2 = 0.78$; World, $\alpha = 0.64 \pm 0.02$, $R^2 = 0.50$. (Data source: DWUA, 2018). The range of exponents listed in Ref. Bettencourt, 2013 is [0.56, 1.04]. The insets show CDFs (Eq. (4)) of population data normalized to its average (P/\bar{P}), fitted as a log-normal (blue) and as a power-law (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

implies that a city twice as large of another city, in terms of population P , is expected to cover a planimetric area $A_2 = 2^\alpha A_1$, with A_1 and A_2 the area of the smaller and larger city, respectively. Establishing if the value of the scaling exponent α is larger, smaller or consistent with unity is crucial to understand if data implies an diseconomy of scale ($\alpha < 1$, sub linear scaling, or decreasing returns), an economy of scale ($\alpha > 1$, super linear scaling, increasing returns), or constant returns to scale ($\alpha = 1$) (the expressions increasing, decreasing and constant returns are borrowed from economics jargon; here they simply refer to proportionality to city size, and have nothing to do with monetary measures).

The basics of area-population relations are illustrated in Fig. 1, which shows how well the supposedly existing scaling holds for world’s cities data available on the WWW, not subjected to any kind of processing. The straight lines in Fig. 1 were obtained by linear regression of the log of area versus the log of population. In fact, taking the log of Eq. (1) one obtains a linear relation between the logarithms of area A and population P :

$$\log A = \beta' + \alpha \log P, \quad (2)$$

with $\beta' = \log \beta$. The values of the exponents (slopes of the linear fits in log–log representation) for cities in Europe, for Europe altogether (five countries shown in the figure) and for the whole world (the whole set of 1,750 data points) are listed in the caption of Fig. 1. The values of regression coefficients are also listed in the caption, along with the calculated uncertainties. Bettencourt, 2013 and references therein reported similar fits to the ones obtained here.

We also characterized the world's cities data set fitting the population data with power laws. To this end, we used the advanced package `powerLaw`, written by Gillespie (2014) for the R programming language. One actually fits the cumulative density function extracted from data with a power-law. Such procedure was inspired by the idea behind the Zipf's law, a relationship between the sizes S of objects in a set and their rank R , defined as follows:

$$S \propto R^{-\gamma}, \quad (3)$$

where γ is the power-law exponent. In the case of cities, size is given by population. The `powerLaw` package builds a CDF and fits that both with power-law and with log-normal functional forms. It also automatically finds the minimum value under which the data have to be discarded for the fitting procedure to be reliable.

We considered population divided by its average as main variable $x = P/\bar{P}$. The CDF function reads as follows:

$$P\left(X \geq x\right) = \left(\frac{x}{x_{min}}\right)^{1-\gamma}, \quad (4)$$

which represents the probability of finding an object with size larger than a given size in a given set (Clauset et al., 2009). Fits of the CDF of Europe's city population are shown in the inset of Fig. 1(a), and in the inset of Fig. 1(b) for world's cities. They show that data cannot be described by any of the fits in the whole P range. The power-law fit (green curves) seems to work well starting from the point where the CDFs obtained from data show a marked change in slope, while the log-normal fit (blue curves) seems to fit best the region of extremely large cities.

Oliveira et al. (2018) recently described a worldwide model to delineate city boundaries and reported about fits similar to the ones in the insets of Fig. 1. Our fits and the quoted ones share the relevant feature of a change in slope, which in our opinion makes the use power-law fits more difficult than A – P relations for the practical purpose that we aim to. Universal validity of Zipf-like power laws, and their usefulness for urban planning, was also questioned by Cottineau (2017). For these reasons, in the methodological part of this work, we did not use Zipf-like power-law fits of the CDFs to investigate administrative boundaries and urban areas in Italy, but we rather used area-population relations as in Eqs. (1) and (2).

We will show that area-population scaling relations can be deduced from data on Italian municipalities, provided area is taken into account in a proper way by selecting an effective area actually occupied by urbanization, in place of the entire area of municipalities.

4.1. Area-population relations within municipality boundaries in Italy

To investigate the existence of a scaling relation between area and population of Italian municipalities, we first utilized the whole area encompassed by each municipality's administrative boundary (polygon) and raw population data, as provided by ISTAT (see Section 3). Fig. 2 shows a log–log plot of area vs. population raw data (black dots), along with a linear fit (black curve). Fits in this figure and throughout this work were performed by the fit function provided by the Gnuplot 5.2 graphic software (Williams et al., 2013); values of slope coefficient α in Eq. (2) (the exponent of Eq. (1)), and correlation coefficients R^2 are listed in the figure's caption. We note that the slope of the fitted line is not nearly close to the slope of any urban A – P scaling

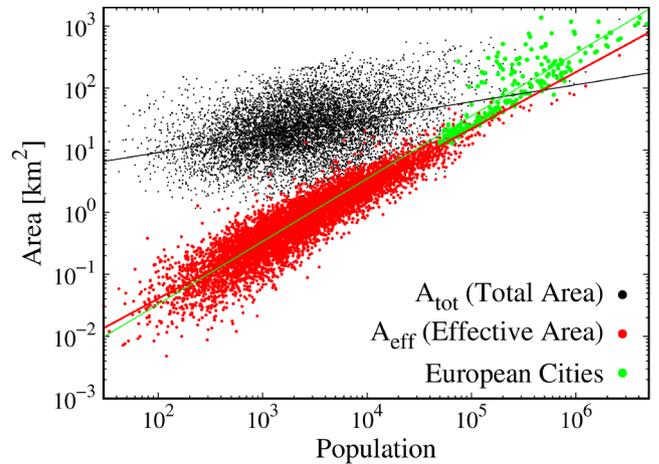


Fig. 2. A – P relations for municipalities in Italy. Black: raw data from ISTAT, 2017 ($\alpha = 0.27 \pm 0.01$, $R^2 = 0.13$). Red: area restricted to the effective area occupied by impervious area A_{eff} ($\alpha = 0.912 \pm 0.004$, $R^2 = 0.74$). Green: European cities data described in Section 3 ($\alpha = 1.01 \pm 0.03$, $R^2 = 0.63$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

relation; the fit for European cities in Fig. 1 is shown in green in the same plot for comparison. Moreover, the small value of the R^2 coefficient of the fit signals a correlation to be unlikely in this A – P data set.

4.2. Urban areas from imperviousness

An imperviousness layer is a measure of how much the soil is sealed by artifacts and, as such, it can be considered as a direct observation of continued human presence. We used the data set obtained from Copernicus WWW site (Steidl et al., 2016) to constrain the area A of municipalities to their effective area occupied by urbanization, by simply cutting the imperviousness layer within the municipalities polygons in a GIS. Counting the number of 20 m x 20 m pixels within each polygon provides an estimate of urban area. We refer to that as A_{eff} , in the following.

Beyond administrative boundaries, examples of city delineation from imperviousness maps are abundant in the literature (see e.g. Ma et al., 2018 and references therein). We used the imperviousness map only to calculate the effective area A_{eff} occupied by urbanization in individual municipalities. These urbanization clusters, though, were still selected by administrative boundaries and never labelled as “cities”. In fact, we decided to adopt a different approach, for the following reasons.

Typical approaches for city delineation from sealed surfaces are often based on numerical thresholds (number of cells, or density thresholds), to consider groups of grid cells as individual cities. Introduction of numerical parameters represents a difficulty of the method. Moreover, the raster map resolution represents an additional lower limit to the accuracy of a delineation performed from sealed surfaces. We performed preliminary city delineation (Abioli, 2020), based on the imperviousness map, showing that (a) due to the above reasons, the points representing cities on the (P, A) plane were distributed in a somewhat biased way, and (b) using such points for a linear A – P fit, we found the same slope as in the case obtained from the method of natural cities, described in the next section. Thus, we preferred the latter as a parameter-free method, adopted in the remaining of this work.

4.3. Natural cities

The definition of “natural cities” was inspired by the city clustering algorithm first proposed by Rozenfeld et al. (2009). The original

algorithm starts from a collection of point-like populated sites, and performs an iterative clustering of sites within a given radius. Iterations end when no points exist within the radius from each of the points in a cluster. The algorithm was applied by selecting streets nodes as starting points by Jiang and Jia (2011), who performed city delineation with this approach in United States, finding that the distribution of sizes (*i.e.* number of street nodes) of the resulting city set neatly follow a Zipf-like law.

The algorithm was further generalized by using either city blocks as clustering domains (Jiang & Liu, 2012), triangulated irregular network (TIN) (Jiang & Miao, 2015), or Thiessen polygons (Jiang, 2018). In all of the variants of the original algorithm, street nodes were taken as a starting point of the city delineation procedure. In this work, we used street nodes obtained from the OpenStreetMap vector layer as a starting point. We generated a TIN network from these points using the GRASS GIS module *v.delaunay*, separately for the peninsular Italy and the major islands, Sicily and Sardinia. All of the GIS operations in this work were performed within GRASS GIS, version 7.6 (Neteler, Bowman, Landa, & Metz, 2012).

We considered the area distribution of the triangles in the TIN network, calculated its average value and checked that an imbalanced ratio existed within the three sub regions of Italy for the number of areas above (smaller number) and below (larger number) the average value: namely, 0.11 for peninsular Italy, 0.14 for Sicily, and 0.21 for Sardinia. Jiang and Liu (2012) found for comparable quantities in France, Germany and UK the values of 0.05, 0.14 and 0.09 for the ratio, respectively.

An imbalanced ratio of large-to-small areas is the signature of a fat-tailed distribution, and of the possibility of splitting each of the three sets in a meaningful way using the natural breaks rule, discarding areas larger than average. Actually, we discarded all of TIN polygons with area above average, along with all of the polygons with area below average which were adjacent to a polygon with large area, following Jiang and Liu (2012). This last step is a parameter-free replacement for utilizing a clustering radius to single out cities in this work, at variance with Rozenfeld et al. (2009). The idea of natural cities was further exploited in this work to devise a fusion algorithm for administrative boundaries.

4.4. Fusion of administrative boundaries

The aim of this paper is to propose a method for the fusion of administrative boundaries in Italy at municipality level, driven by area-population scaling relations. The proposed procedure is grounded on our view of the empirical area-population scaling relation as a “performance” assessment of each individual datum in the data set (municipalities, in this case) as compared to expectations generated by the whole data set based on the individuals’ size (population, in this case) (Bettencourt, 2013; West, 2017; van Raan, 2019; O’Clery, Curiel, & Lora, 2019).

We argue that if a robust scaling relation as a function of municipalities’ size (population) can be obtained, the vertical distance on the (P , A) plane of an individual’s representative point from the linear scaling law can be viewed as a measure of its “performance”, in absolute value (*i.e.*, considering equivalent both if it is above or below the linear fit). Under this perspective, distance of a city’s representative point from the scaling law represents a measure of how much the city is making a productive use of available resources.

Thus, we suggest that one can rank municipalities starting from the farthest from the scaling law. Starting from the highest ranking, one can select all of the adjacent (surrounding) municipalities and try and find a candidate polygon to be merged with the central one, based on two requirements: (i) the aggregate areas and populations of the candidate mergers have a representative point in the (P , A) plane which is closer to the scaling law than both the starting points, and (ii) the two candidates for the merger are connected by an urban area, as obtained by

the method natural cities. If both the requirements are fulfilled, the two polygons are merged. The procedure can be iterated until all of the existing polygons have been checked against all of the possible mergers.

The requirement (i) is the simplest one that one can devise for such a purpose; one could also adopt more involved measures for ranking performance with respect to an empirical scaling relation. The same goes for requirement (ii), because it seems reasonable that two municipalities can be considered for fusion into a single administrative unit if they share a common city, in a broad sense. The requirement is generic enough that it may be fulfilled by an urban agglomeration of any size, and the range of variation of city size is very large as it will be shown in the following. Both requirements, and the whole procedure, are suited to be implemented in a GIS with an automatic iterative procedure. The algorithm of the procedure for polygon fusion is described by the pseudo-code in the listing Algorithm 1.

Lastly, we need to account for coarser administrative boundaries besides municipalities, namely Provinces and Regions. We disregarded Provinces, which have themselves changed multiple times in Italy in recent times. We rather focused on Regions, of which Italy is subdivided into twenty, and whose properties relevant for this work are listed in Table 1. We restricted fusion between polygons of municipalities belonging to the same Region, but still used the national scaling law as a measure to accept/reject mergers.

5. Results

In this section, we describe results obtained for the following four main points:

1. validity of area-population scaling relations for Italian municipalities, described in Sections 4.1 and 4.2;
2. delineation of urban areas (or cities) in Italy, with a variation of the method of “natural cities” recently introduced in the literature, described in Section 4.3;
3. pairwise fusion of a few adjacent polygons corresponding to Italian municipalities, described in Section 4.4;
4. comparison among the proposed aggregation and existing aggregations of municipalities in Italy.

We discuss the four points separately in the following.

5.1. Area-population scaling relations for Italian municipalities

Fig. 2 shows area-population relations built with raw A and P data for Italian municipalities (black dots/curve). The figure also shows the effect of constraining the area A of municipalities to their effective area A_{eff} occupied by urbanization, in a broad sense (red dots/curve). In this work, we adopted any level of imperviousness (Steidl et al., 2016) as an indicator of urbanization. We also plotted again data for the European cities (green dots/curve), whose fit nicely compares with the A_{eff} - P result.

We explicitly investigated the effect of different definitions of A and P . First, we studied the effect of modifying P with the inclusion of daily commuting data, vs. total A ; the effect was negligible. Then, we checked the effect of adding and removing pixels overlapping with roads to modify A_{eff} . In fact, in the imperviousness layer roads may or may not be correctly identified so this step is not trivial. We found that adding missing roads reduces the slope of the scaling relation, which nevertheless still falls within the range proposed by Bettencourt (2013), while removing all of the roads again has negligible effect. Eventually, we checked that the effect of modifying P using commuting flow data vs. the A_{eff} plus roads areas in conjunction. This seemed the most accurate approximation we could devise, since people actually spend time on roads too. Again, we found no substantial change with respect to the red curve in Fig. 2, which did not account for commuting.

We may deduce from Fig. 2 and from the considerations above that

Table 2

Numerical details about polygons stemming from city delineations performed with the “natural cities” method described in Section 4.3. Results are shown in Figs. 3 and 4(d).

ID	Number of cities	Average area [km ²]	Maximum area [km ²]
Peninsular	77,103	0.244	1,311,601
Sicily	6,190	0.189	112,829
Sardinia	5,979	0.182	86,762
Italy	89,272	0.224	1,311,601

either using the only impervious area A_{eff} , or A_{eff} plus roads, produce a reasonable scaling relation against P , regardless of the use of modifications to P due to commuting data. We also note that the A_{eff} – P relation, red line in Fig. 2, compares well with European cities data. Thus, we adopted the A_{eff} – P linear fit as a baseline scaling relation in the pairwise fusion of municipalities.

5.2. Delineation of urban areas in Italy

We performed delineation of urban areas by using the idea of “natural cities” built from the bottom-up using the TIN network obtained from street nodes, as described in Section 4.3. Thus, in this work we compare results of three different approximations of A – P relations, namely: (i) plain area A of municipalities vs. P , (ii) impervious (or effective) area of municipalities A_{eff} vs. P , (iii) area from cities calculated with the natural cities method vs. the corresponding population.

Delineation of cities based on the “natural cities” method, resulted in a total of 89,272 urban agglomerations; additional numerical details are listed in Table 2. Fig. 3 shows a map of cities delineated using the method of natural cities. We operated Delaunay triangulation from street nodes independently for the three major geographically disjoint sub sets of Italy, namely peninsular Italy, Sicily and Sardinia.

We found instructive to compare the three different approximations (i)–(iii) listed above from which we built the different A – P scaling relations. Fig. 4 shows results for Sicily, to allow a larger-scale comparison. The comparison is still qualitative, but one can appreciate the differences. Transition from approximation (i), Fig. 4(a), to approximation (ii), Fig. 4(b), clearly shows the difference in area measured by the corresponding scaling relations in Fig. 2, black and red curves/dots, respectively, within municipality boundaries. When such boundaries are removed, as in approximation (iii) Fig. 4(c), areas are computed for the distinct clusters singled out by the natural city delineation method.

Fig. 5 shows a quantitative assessment of A – P relations within the approximations (ii) and (iii). The figure shows separately data for peninsular Italy, Sicily and Sardinia, where urban areas were delineated separately. The plot show a single linear fit, corresponding to aggregate data for the whole of Italy. The linear fit goes into the direction of slightly decreasing the slope obtained in case (ii). Data nicely distribute on the (P, A) plane, clearly hinting to a linear correlation. Some patterns seem to emerge, mostly as collinear structures to the overall linear fit in the Sicily and Sardinia sub sets, suggesting that some minor bias might occur in the case of delineation of natural cities in smaller areas.

We used the natural cities delineation for the polygon-fusion procedure, the final aim of this work.

5.3. Fusion of administrative boundaries

The procedure for pairwise fusion of polygons representing municipalities is iterative. We started selecting the polygon whose representative point on the A – P plane is farthest in the vertical direction from the red curve in Fig. 2, the empirical scaling relation for municipalities. Fusion is pairwise, meaning that for each polygon we checked for another candidate polygon at a time to be merged with. This means the outcome of the procedure depends on the order of selection of both

candidate polygons; nevertheless, for each checked polygon there were often only one or no acceptable mergers, making the dependence on the order of choice rather weak.

The algorithm of the procedure for polygon fusion is described in detail by the pseudo-code in the listing Algorithm 1. Polygon fusion was restricted within individual Regions in Italy, a higher-order administrative subdivision, because mergers across different Regions would be very unrealistic. The overall process was performed in parallel by 20 GIS procedures. Each procedure goes to completion *i.e.*, checks all of the existing polygons and performs the possible mergers according to the requirements described in Section 4.4, in a few hours. At the end of the run, we re-calculated the new fitting parameters α , β of Eq. (2), when the 20 individual results obtained for each Region were collected in one single map. The result for the new parameters of the national scaling law is substantially the same as for the parameters obtained from the original polygons.

To visualize the effect of the fusion procedure, Fig. 6 shows a few examples among the few hundreds mergers operated by the proposed algorithm. In the figure, examples are shown of mergers between a point above and a point below the linear fit, and of examples of both points below the fit. Examples of both points above the fit occur rarely, because the cumulative area being larger than both the original points would often cause violation of the requirement for the new point to be closer to the linear fit on the vertical (A) direction. The figure also shows three of the many mergers undergone by the largest municipality, including the city of Rome, which after applying the proposed procedure would incorporate many neighbouring municipalities, as shown in Fig. 7.

Fig. 7 shows a pictorial representation of the performance of the original subdivision, consisting of 8092 polygons, and of the subdivision after the fusion procedure, consisting of 6571 polygons. The figure shows only new polygons of the proposed delineation, for clarity of presentation. The performance is measured as the vertical distance (in powers of ten) of each representative (P, A) point from the empirical scaling law. Note that, in this acceptance, the vertical distance represents the amount by which each A_{eff} exceeds or falls behind the expectation suggested by the scaling law, given the municipality’s population.

At variance with the procedure itself, in which vertical distance was taken in absolute value, we have coloured each polygon in Fig. 7 in 9 classes: a central one, in grey, four classes below the scaling law and four classes above, to visually distinguish points falling below or above the fit with different colour ramps. The below-above classes are symmetric, and were dictated by a data-driven classification using the head/tail break method with respect to the vertical distance of each point from the scaling law, in the log–log scale. Table 3 lists the number of polygons and the total area A (not A_{eff}) in the different classes.

Since we performed the procedure for the fusion of boundaries within administrative Regions, we show in Fig. 8 the corresponding 20 individual results. The figure includes A_{eff} – P points for existing municipalities (labelled as “in”) and the points for the new delineation (“out”), along with the initial and final slopes of the fits calculated within each Region. The figure aims at illustrating differences between the various Regions, but in the fusion procedure we always used the national scaling law.

5.4. Comparison with existing aggregations

It is interesting to compare the proposed aggregation of municipalities with two existing ones, namely the SLL and UDC introduced in Section 3. Fig. 9(a) shows the A – P relation of SLL (cyan) in which we used effective area for all of the data sets included in the plot. Local labour systems, consisting of 611 polygons, are substantially larger than municipalities, 8,092 polygons, and the fit to SLL (cyan, dashed curve) has a less steep slope than both the fit to existing municipalities (grey, solid curve) and polygons after the mergers (brown, dotted line)

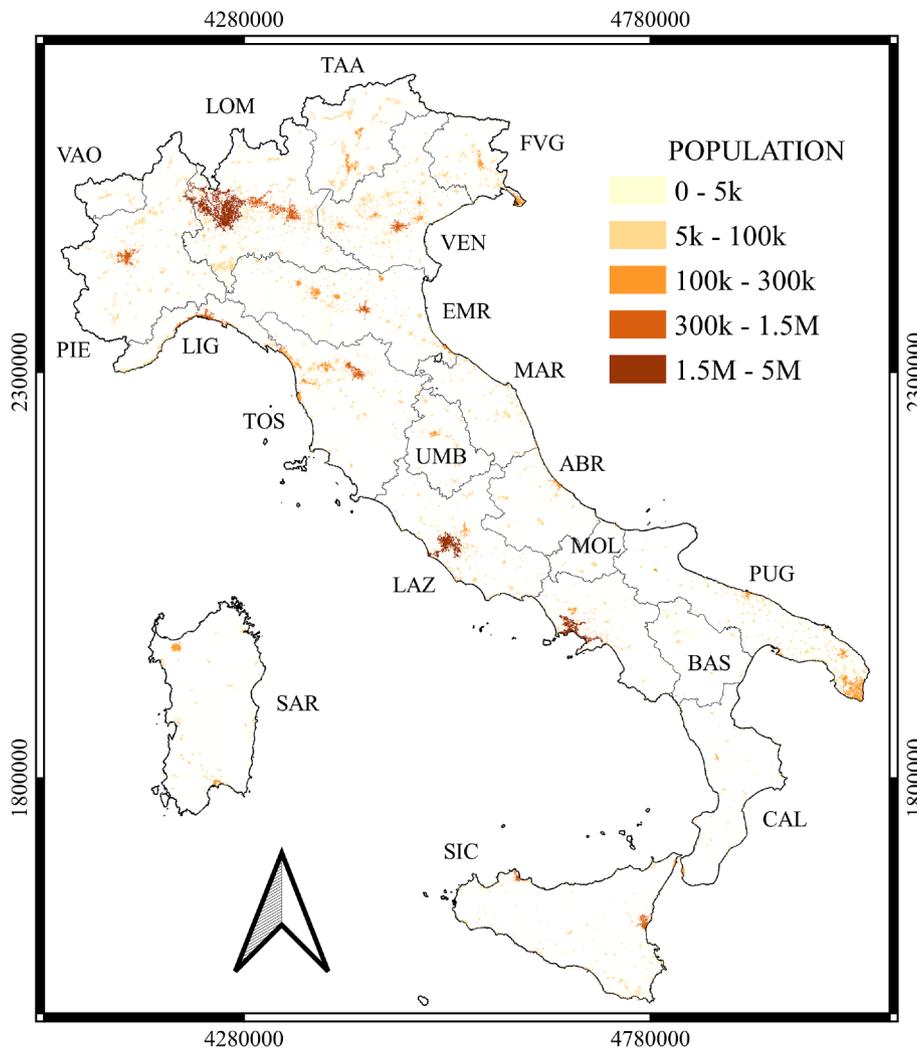


Fig. 3. Delineation of natural cities, as described in Section 4.3. See Section 5.2 and Table 2 for numerical evaluation of the results, Fig. 4 for comparison with administrative boundaries and Fig. 5 for A_{eff} - P relations. Statistics for the 20 administrative regions labelled with three-letter codes are listed in Table 1. Map is in Lambert Azimuthal Equal Area (LAEA) projection, EPSG:3035.

suggested in this work. The UDC layer contains 531 modified polygons, while our new layer contains 1031 new polygons with respect to the existing municipalities layer. The total area of the “new” polygons in UDC is 111,699 km² (3128 “old” municipalities), which compares to the 91,731 km² (2555 “old” municipalities) of the new subdivision proposed in this work. Fig. 9(a) shows only the new points contained in UDC (blue) and in the new delineation (brown), clearly highlighting that the former contains a bias towards larger areas with respect to the existing administrative boundaries (grey), failing to generate a corresponding increase in population within the merged polygons. Similar considerations holds for the SLL described above. In this work, by construction, each “new” polygon has to follow the general trend of the empirical area-population scaling law better than both the two original polygons.

A chance for direct comparison is provided by official data in Italy (ISTAT, 2017), which reveal that the number of municipalities in Italy in the last decades have decreased from 8100, in 1990, to 7904, in 2020. We applied the polygon fusion procedure to 2011 data, 8092 municipalities, because census data was available to us only for that year. Fig. 9(b) shows that the points representative of the only new (aggregated) municipalities on the (P , A) plane are distributed in a rather different way with respect to the points representative of mergers proposed in this work, deviating from the national scaling law. We stress that in Fig. 9(b) the population used for the 2020 boundaries was

from 2011, because the 2020 data were not available to us at the time of writing.

6. Discussion

An area-population relation built with raw A and P data for Italian municipalities, shown in Fig. 2 (black curve), is far from any scaling law defined in the literature for world’s cities (cf. Fig. 1). A meaningful scaling law is found instead by restricting the area A of municipalities to their effective area A_{eff} occupied by urbanization, in a broad sense. A similar quantity (namely, paved areas) was also investigated by Bettencourt (2013), therein referred to as A_n . It was reported by the reviewed literature to scale with population with exponents in the range [0.74, 0.92], which includes the range obtained in this work for municipalities’ effective area, $\alpha = 0.912 \pm 0.004$.

Bettencourt (2013) also mentioned (in the paper’s supplementary material) that up-to-date estimates of cities’ paved area from remote sensing is desirable. We believe that the data sets used in this work, available from the Copernicus programme homogeneously for the whole of Europe at high resolution and with constant updates, should be considered for similar studies at continental level and for time-dependent analyses (CLMS, 2018). The definition of criteria for the study of time evolution of urban systems was considered by different Authors (Makse, Havlin, & Stanley, 1995; Barthélemy et al., 2013; Fang & Yu,

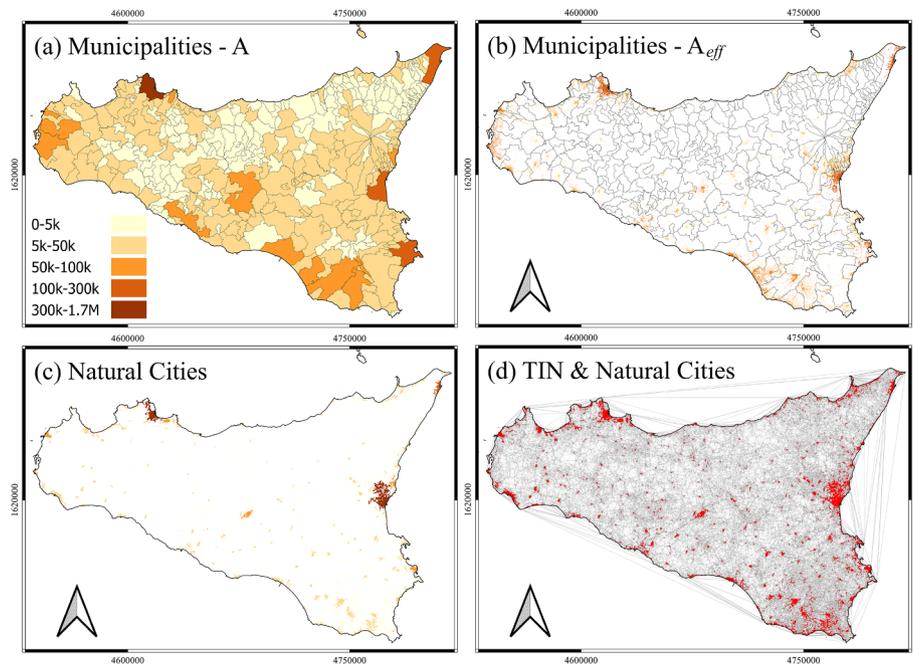


Fig. 4. Different representations of the geographical distribution of the population in Sicily. (a) total area A of municipalities, coloured according to population data (ISTAT, 2017). (b) effective area A_{eff} occupied by the imperviousness layer for each municipality (CLMS, 2018). (c) cities delineated using the definition of “natural cities” (Jiang & Liu, 2012; Jiang & Miao, 2015). (d) as in (c), but cities are colored in red regardless of their population and are superimposed to the triangulated irregular network used to delineate them. Maps are in LAEA projection, EPSG:3035. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

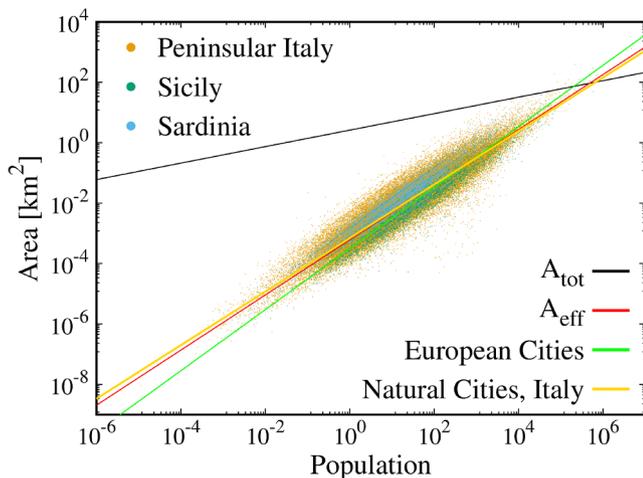


Fig. 5. A - P relations from the method of natural cities (yellow curve, $\alpha = 0.886 \pm 0.001$), considered in this work and shown in Fig. 3; dots show the (P, A) points for the three major disjoint geographical domains in Italy. We compared to results for municipalities (black and red curves; $\alpha = 0.27 \pm 0.01$ and $\alpha = 0.912 \pm 0.004$, respectively). In green the A - P fit for European cities, $\alpha = 1.01 \pm 0.03$. A comparison of the geographical distributions of the results, only for Sicily, is in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2017), and projections for the future until year 2030 also exist by authoritative organizations (United Nations, 2018). This is a relevant point, since the whole world’s population is expected to live in cities of some form in the near future, though city size distributions are expected to experience little change (Batty, 2018).

Delineation of urban areas is relevant to considerations about the 2020 pandemic event, as the diffusion of an infection is clearly dependent on the spatial distribution of citizens at the national level and dynamics of social interactions within urban boundaries and across neighbouring cities. Very recently, Stier, Berman, and Bettencourt

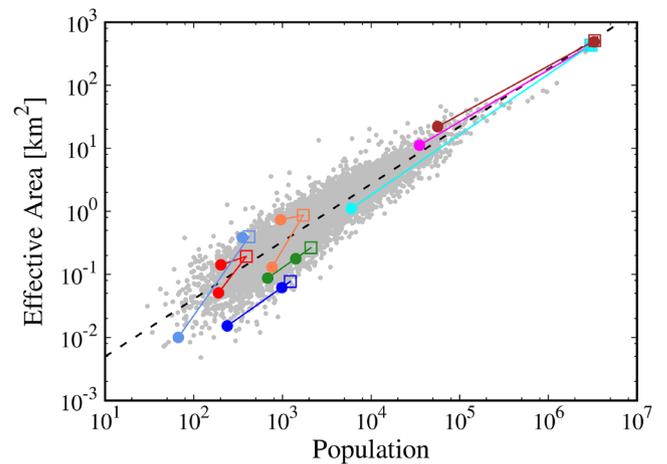


Fig. 6. Pictorial representation of pairwise fusion of polygons, each represented by a grey dot. Each colour represents a sample merger, among the few hundreds operated by Algorithm 1. Full circles correspond to the two original municipalities and empty squares to the polygon after the merger. The red, cerulean blue and orange groups show examples of mergers between a point above and a point below the linear fit (dashed line); the green and blue groups are example of both points below the fit. The brown, pink and cyan groups shows three of the many mergers undergone by the largest municipality (Rome) which after the procedure would incorporate many neighbouring municipalities (see Fig. 7). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2020) observed that growth rates of COVID-19 in US cities revealed a power-law scaling relationship to city population. Given that Italy was among the most affected countries by the virus, delineation of urban areas in Italy may acquire additional relevance in the present context and, most importantly, in the near future.

Very recently van Raan (2019) studied the scaling of gross urban product (GUP) versus population in Denmark, Germany and the

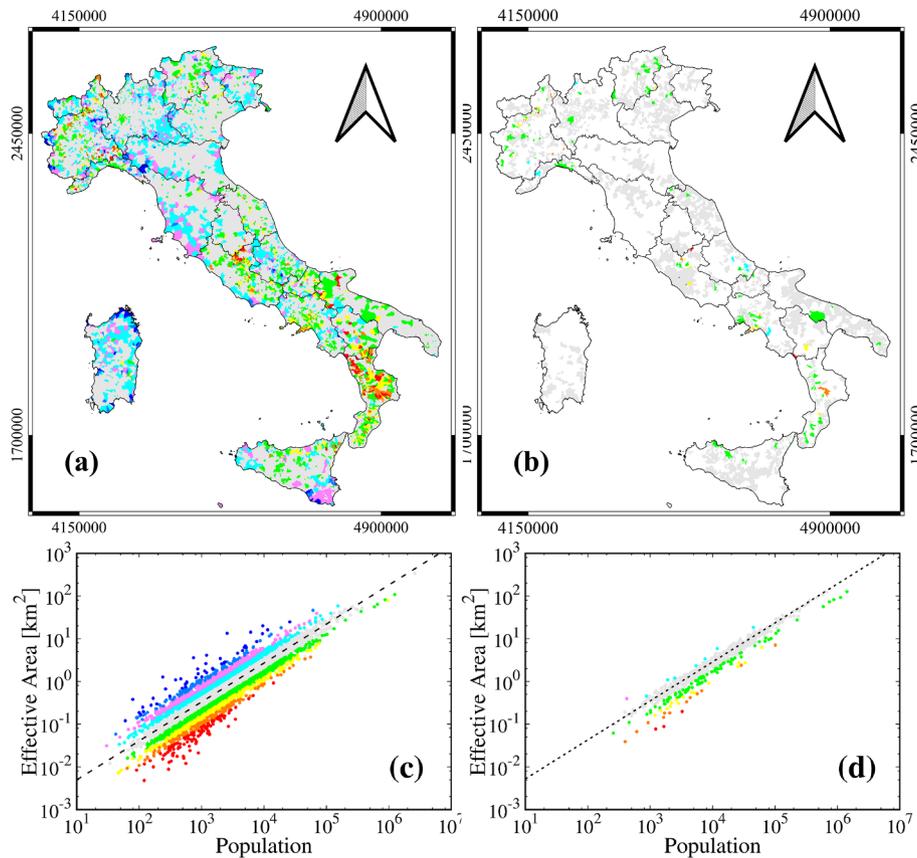


Fig. 7. Top: a pictorial representation of the (vertical) deviation of each municipality from the linear fit of Fig. 2 for the existing subdivision (a) and the (only) new 1031 polygons of the proposed subdivision (b); 5323 unmodified polygons are not shown in (b). Bottom: projection on the (P, A) plane of the corresponding municipalities depicted in the top panels. The linear fit in (c) is the same as the red curve in Fig. 2 (slope 0.912), obtained from the original 8092 points, while (d) shows the actual fit of the total (new and unmodified) 6571 (P, A) points corresponding to the proposed subdivision (slope 0.913). See Table 3 for results in terms of number of polygons and area in each class. Maps are in LAEA projection, EPSG:3035.

Table 3

The number of polygons and total area corresponding to the existing subdivision in municipalities in Italy (*old*), and the proposed one (*new*). Class 1 corresponds to grey in Fig. 7; class 2, green; class 7, cyan; class 3, yellow; class 8, magenta; class 4, orange; class 9, cerulean blue; class 5, red; class 10, blue.

Class	Number (<i>old</i>)	Area [km^2]	Number (<i>new</i>)	Area [km^2]
1	4919	181,757	4042	197,204
2, 7	1990	78,013	1580	69,625
3, 8	750	27,876	586	22,794
4, 9	272	8321	227	7525
5, 10	161	5430	136	4245

Netherlands. Like in this work, van Raan (2019) considered urban areas, cities and municipalities. In particular, they distinguished urban and rural municipalities, finding substantially different scaling exponents, and concluded that urban regions consisting of a single municipality perform better than those including many municipalities. Even if the scaling relation is different from the one considered here, *i.e.* GUP versus P instead of A_{eff} versus P , the two studies are clearly related, and the conclusions of van Raan (2019) indirectly support the motivation and methods of the present work. In fact, both studies look at the performance of larger, coarser municipality boundaries with respect to smaller, finer boundaries, and both studies consider scaling relations, actually deviations from the empirical scaling relation, as performance indices.

A sound assessment of area-population scaling for municipalities allowed to propose a criterion to merge municipalities in Italy, driven scaling relations. The exercise of practical interest, in that official data

in Italy (ISTAT, 2017) reveal that the number of municipalities changed by about 200 units in the last thirty years; unofficial sources available on the WWW reveal that fusion of municipalities (*comuni*) started in 1862. Moreover, additional variation of such boundaries are planned for the near future. Suppression or fusion of two or more municipalities into a new one were dictated by a number of different criteria, under the general aim of reducing the total number of administrative centers.

We believe that the perspective from scaling laws proposed in this work can be useful to complete the criteria currently adopted for fusion of administrative boundaries, to obtain a comprehensive point of view. Adoption of multiple criteria could avoid negative effects of individual criteria, when applied separately. For example, the only prescription of following an area-population scaling law could promote the practice of indiscriminate land use or urban sprawl, given that an administrative unit would be favoured if a large population corresponds to a large urban area. This could be avoided by the joint application of multiple criteria.

The procedure for the fusion of administrative boundaries produced a new delineation consisting in 6571 units instead of the existing 8092 units. Considering the empirical scaling relation of Fig. 2 as a metric for “performance” of each unit, the procedure reduced the total area of units classified as “less performing” from 5430 km^2 to 4245 km^2 (*cf.* Table 3), reduced the total area in intermediate classes from 114,210 km^2 to 99,944 km^2 , and increased the total area in the “most performing” class from 181,757 km^2 to 197,204 km^2 .

We stress that we used the empirical $A_{\text{eff}}-P$ scaling law in the most straightforward way we could devise, to show the feasibility of the approach. An improved procedure could be implemented by means of additional criteria, for example applying considerations of economic

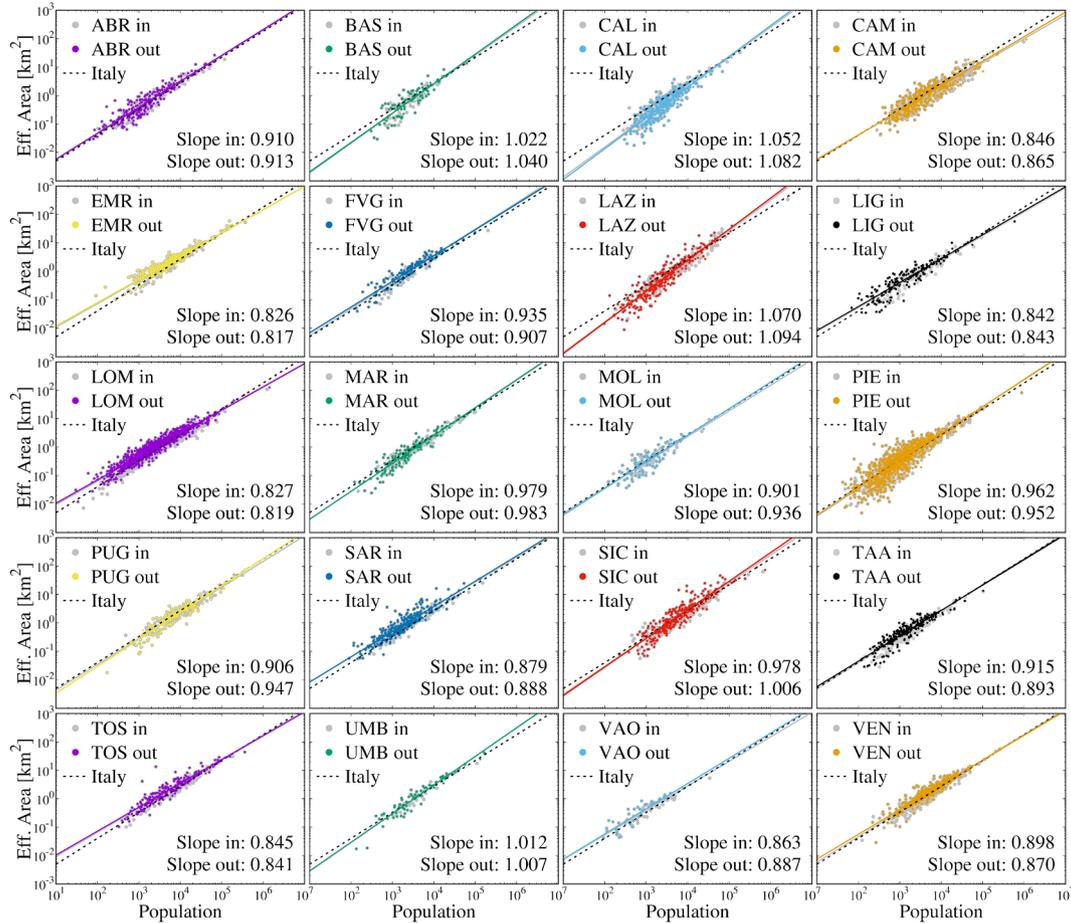


Fig. 8. A_{eff} - P relations for the individual administrative regions. In each panel, the three-letter code corresponds to Table 1 and Fig. 3; grey points and curves refer to existing municipalities' data and the corresponding fit (*in*), while coloured points and curves refer to the new delineation (*out*). The dotted curve is the A_{eff} - P national fit, first shown in Fig. 2 as a red curve and referred to throughout this work. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

nature, or considering actual commuting flows in the commuting matrix described in Section 3 to single out candidate polygons for mergers. Preliminary calculations aimed at checking the effects of simply relocating population from/to municipalities according to daily commuting flows showed to be of little relevance, as far as scaling laws are concerned (not shown here).

The SLL subdivision based on commuting flows, defined in Italy by ISTAT, provides a more sophisticated use of such information, independent on scaling considerations. Fig. 9(a) shows a comparison of the A - P relations from our aggregation method and aggregations provided by SLL and UDC, highlighting rather substantial differences. Regarding SLL, Fig. 9(a) clearly shows that the distribution of points in the (P, A) plane is biased towards larger areas than the empirical scaling law, which was expected, given the different rationale and purpose of the SLL subdivision. Regarding UDC, the differences are mainly due to the fact that both the considered existing methods do not take into account area and population simultaneously, the proposed delineation. The substantial difference is that the total population interested by UDC amounts to 11,528,890 citizens, while the corresponding figure for our new delineation is 32,551,304 citizens (SLL was delineated using different input data, so we do not report corresponding figures, here). This is clear indication that fusion of administrative boundaries performed in different ways may generate very different scenarios, population-wise. Similar considerations holds for the current subdivision into municipal units, Fig. 9(b), where the new units involved 10,570,208 citizens. A proper discussion should take into account both geographical boundaries and population data from 2020,

thought, while data in the figure correspond to population in 2011.

Recently, an additional aggregation of municipalities was proposed by the Italian Civil Protection Department (DPC, 2020), going under the name of *contesti territoriali* (literally territorial contexts CT). This particular aggregation, while bearing a certain relationship with the above-mentioned SLL and UDC, were developed for emergency planning and risk management. Thus, the purpose of the aggregation proposed here is probably not directly comparable with the CT aggregation and was not explicitly considered in this work. A meaningful comparison with a proper discussion will be presented elsewhere.

7. Conclusions

Scaling laws for urban systems are deemed near-universal, because scaling of a specific indicator on city size slightly depends on geographical location and/or the criteria imposed for the very delineation of cities. Nevertheless, once an empirical A - P law is established for a given country or otherwise homogeneous study area, it may be viewed as a general and robust measure of the "performance" of a city of a given size, in terms of expectations about its area (Bettencourt, 2013; West, 2017).

We presented an analysis of Italian urban areas in relation to existing administrative boundaries. We believe that the present work contributes to filling the existing gap in the literature, in which delineation of urban areas were not found for Italy, allowing to distinguish the behaviour of different urban systems within Europe. Our results can be summarized in the following points:

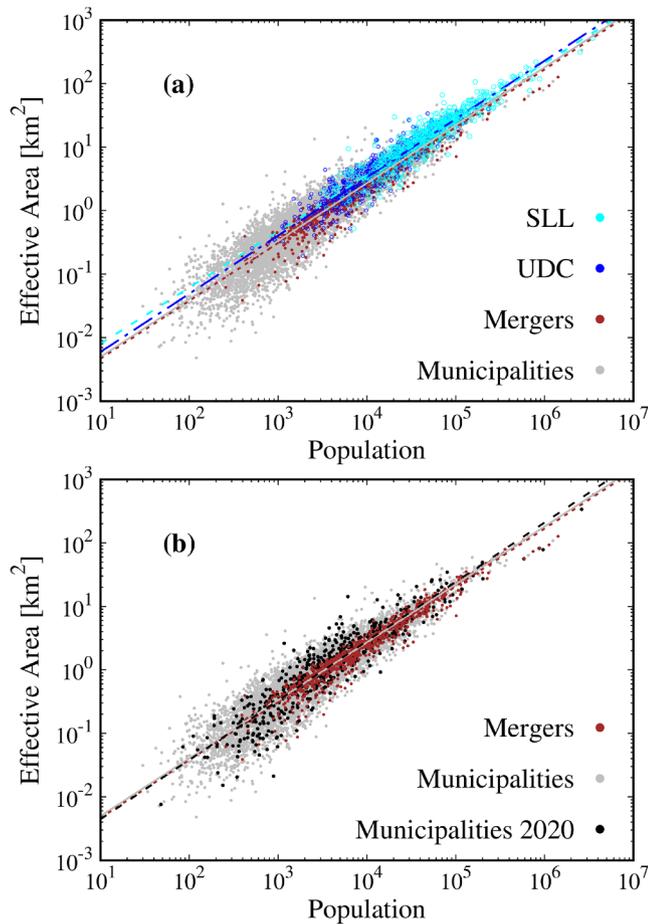


Fig. 9. Comparison of A_{eff} - P relations with existing aggregations of municipalities in Italy. (a) the existing (grey) and proposed (brown) units; we show only new mergers proposed in this work, and compared to alternative delineations, namely the “Unione di Comuni” (UDC, blue) and the “Sistemi Locali del Lavoro” (SLL, cyan). For UDC, we show the only aggregations differing for one or more units (as for the mergers), while we show all of the SLL points, which have much larger areas than all of the other delineations. (b) grey and brown as in (a), compared to the actual new municipalities in Italy, in 2020 (black points); population data were from 2011 for all the data sets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Area A for Italian municipalities scales sub linearly with population P , with a slope smaller than the slope obtained for Italian and European cities from data available on the WWF, and within the range reported for world’s cities by Bettencourt (2013). Scaling is not obtained for the total area, though, but for the effective built-up (actually, impervious – or sealed) area A_{eff} (Fig. 2);
- We obtained “natural cities” using natural breaks selection of geographical zones built from the nodes of the national street network (Figs. 3 and 4; (Jiang & Liu, 2012; Jiang & Miao, 2015)). Such a delineation of cities in Italy provides very similar A - P relations with respect to the one valid for effective area of municipalities (Fig. 5);
- A reduction of the number of administrative units, at municipality level, could be performed using the A_{eff} - P scaling law as a general criterion for pairwise fusion of neighbouring polygons (Fig. 6);
- Iterative application of the merging procedure produced a new aggregation, containing 6,571 units, as compared to the existing 8,092 units (Figs. 7 and 8). Units are larger than the existing ones, but their population still follows the empirical A - P scaling law, at variance with existing aggregations of municipalities in Italy (UDC, SLL and the current administrative boundaries; Fig. 9).

We stress that a data-driven method for delineating cities such as the definition of natural cities adopted in this work, in conjunction with the many different methods considered by e.g., Cottineau et al. (2017), may contribute in finding a comprehensive and objective delineation method.

The proposed real-world application of scaling laws is an example of the possibilities offered by such universal relations, suggesting their use in conjunction with other (expert, political, social, economic) criteria to delineate boundaries that simultaneously comply with, or are obtained from reasoned aggregation of, existing administrative subdivisions. One of the advantages of pursuing a path like the one outlined here is that it also represents a bottom-up approach, based on the ideas of emergent phenomena from self-organization, which may reveal more effective than top-down, centralized approaches to planning (Barthélemy et al., 2013).

Inclusion of general principles such an empirical A - P scaling law into the delineation of administrative boundaries goes into the direction of pursuing a comprehensive science of cities, envisaged by Batty (2013, 2018). A science of cities which takes into account complex behaviour and emergent phenomena, which form the basis of scaling laws, may lead to cities and administrative boundaries which are more resilient against environmental, socioeconomic, and political issues (Meerow, Newell, & Stults, 2016), and hopefully against natural disasters (United Nations, 2018).

We further highlight the importance of exploiting open data, including satellite and volunteered data which, in conjunction with modern open source GIS software with powerful processing capabilities, and with a multidisciplinary approach, may be relevant for urban and administrative planning. The methods presented in this work can be readily extended to any country and in particular to the whole of Europe, where data from the same sources used here is already available.

Algorithm 1. Pseudo-code structure of the algorithm, implemented in GRASS GIS, for pairwise fusion of municipality polygons driven by the empirical A_{eff} - P scaling relation as in Eq. (2) and represented by the red curve as Fig. 2. The flag chk is set to 0 at the beginning of the procedure for all of the polygons. Cities referred to in this pseudo-code are natural cities described in Sections 4.3 and 5.2.

```

1: loop: ( $A$ - $P$  points ranked by performance)
2: polygon  $\leftarrow$  worse performance
3: candidates  $\leftarrow$  adjacent polygons with  $chk \neq 1$ 
4: cities  $\leftarrow$  cities overlapping polygon and candidates
5: if [#cities > 0]
6:   loop (cities ranked by areal size)
7:     city  $\leftarrow$  largest of cities
8:     candidate  $\leftarrow$  largest of candidates over city &  $chk \neq 1$ 
9:      $D_1 \leftarrow$  residual of polygon w.r.t.  $\log A = \beta' + \alpha \log P$ 
10:     $D_2 \leftarrow$  residual of candidate w.r.t.  $\log A = \beta' + \alpha \log P$ 
11:     $D \leftarrow$  residual of merger w.r.t.  $\log A = \beta' + \alpha \log P$ 
12:    if [ $D < D_1$  &  $D < D_2$ ]
13:      merger is accepted: set  $chk = 1$  for polygon
14:      replace old polygon map to include merger
15:      optionally recalculate fit  $\rightarrow$  new  $\alpha, \beta$ 
16:    else
17:      if [#cities > 1]
18:        merger is rejected: set  $chk = 1$  for candidate
19:        cities  $\leftarrow$  cities overlapping polygon and candidates
20:        candidates  $\leftarrow$  adjacent polygons with  $chk \neq 1$ 
21:      else
22:        merger is rejected: set  $chk = 1$  for polygon
23:        continue to outer loop
24:    end if
25:  end if
26: end loop
27: end inner loop
28: else
29:   merger is rejected: set  $chk = 1$  for polygon
30:   continue to inner loop
31: end if
32: end loop
33: end outer loop

```

Acknowledgments

I am indebted with the whole Geomorphology group of CNR IRPI, Perugia, especially with Dr. Ivan Marchesini for introducing me to the world of GIS processing. I also thank Dr. Bin Jiang, Faculty of Engineering and Sustainable Development, University of Gävle, Sweden, for thorough discussions about the definition of natural cities, and Dr. Luca Fondacci, Department of Engineering, University of Perugia, Italy, for interesting discussions about urban planning. This paper also benefited from constructive comments of three anonymous reviewers.

References

- Alvioli, M. (2020). Comparative study of delineation of urban areas using imperviousness products and open data. In M. Alvioli, I. Marchesini, L. Melelli, & P. Guth (Eds.). *Proceedings of the Geomorphometry 2020 Conference* (pp. 266). Perugia, Italy: CNR Edizioni. https://doi.org/10.30437/GEOMORPHOMETRY2020_1.
- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, 9, 3975–3991. <https://doi.org/10.5194/gmd-9-3975-2016>.
- Alvioli, Massimiliano, Guzzetti, Fausto, & Marchesini, Ivan (2020). Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology*, 358, Article 107124. <https://doi.org/10.1016/j.geomorph.2020.107124>.
- Arcaute, E., Hatna, E., Ferguson, P., Youn, H., Johansson, A., & Batty, M. (2015). Constructing cities, deconstructing scaling laws. *Journal of The Royal Society Interface*, 12. <https://doi.org/10.1098/rsif.2014.0745>.
- Auerbach, F. (1913). Das gesetz der bevölkerungskonzentration. *Petermann's Geographische Mitteilungen*, 59, 74–76.
- Barbosa, H., Barthélemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., et al. (2018). Human mobility: Models and applications. *Physics Reports*, 734, 1–74. <https://doi.org/10.1016/j.physrep.2018.01.001>.
- Barthélemy, M., Bordin, P., Berestycki, H., & Gribaudi, M. (2013). Self-organization versus top-down planning in the evolution of a city. *Scientific Reports*, 3, 2153. <https://doi.org/10.1038/srep02153>.
- Barthélemy, M., & Flammini, A. (2008). Modeling urban street patterns. *Physical Review Letters*, 100, Article 138702. <https://doi.org/10.1103/PhysRevLett.100.138702>.
- Batty, M. (2008). The size, scale, and shape of cities. *Science*, 319, 769–771. <https://doi.org/10.1126/science.1151419>.
- Batty, M. (2013). *The new science of cities*. The MIT Press.
- Batty, M. (2018). *Inventing future cities*. The MIT Press.
- Batty, M., & Longley, M. (1994). *Fractal cities – A geometry of form and function*. London: Academic Press 448.
- Bertuglia, C. S., Bianchi, G., & Mela, A. (2012). The city and its sciences. *Physica-Verlag*. <https://doi.org/10.1007/978-3-642-95929-5>.
- Bettencourt, L., & West, G. (2010). A unified theory of urban living. *Nature*, 467, 912–913. <https://doi.org/10.1038/467912a>.
- Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science*, 340, 1438–1441. <https://doi.org/10.1126/science.1235823>.
- Bettencourt, L. M. A., & Lobo, J. (2016). Urban scaling in Europe. *Journal of The Royal Society Interface*, 13, 20160005. <https://doi.org/10.1098/rsif.2016.0005>.
- Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104, 7301–7306. <https://doi.org/10.1073/pnas.0610172104>.
- Clauset, A., Shalizi, C., & Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, 51, 661–703. <https://doi.org/10.1137/070710111>.
- CLMS. (2018). URL: <https://land.copernicus.eu> (accessed: March 2018).
- Cottineau, C. (2017). Metazipf. A dynamic meta-analysis of city size distributions. *PLoS One*, 12, Article e0183919. <https://doi.org/10.1371/journal.pone.0183919>.
- Cottineau, C., Hatna, E., Arcaute, E., & Batty, M. (2017). Diverse cities or the systematic paradox of urban scaling laws. *Computers, Environment and Urban Systems*, 63, 80–94. <https://doi.org/10.1016/j.compenvurbys.2016.04.006>.
- DPC. (2020). URL: <http://governancerischio.protezionecivile.gov.it/rischio-idrico/progetti-generalisti> (In Italian, accessed: January 2020).
- Durand, M. (2007). Structure of optimal transport networks subject to a global constraint. *Physical Review Letters*, 98, Article 088701. <https://doi.org/10.1103/PhysRevLett.98.088701>.
- DWUA. (2018). URL: <http://www.demographia.com> (accessed: February 2018).
- Fang, C., & Yu, D. (2017). Urban agglomeration: An evolving concept of an emerging phenomenon. *Landscape and Urban Planning*, 162, 126–136. <https://doi.org/10.1016/j.landurbplan.2017.02.014>.
- Flowerdew, R. (2011). How serious is the modifiable areal unit problem for analysis of English census data? *Population Trends*, 145, 106–118. <https://doi.org/10.1057/pt.2011.20>.
- Furtado, A., Sakovski, P. A. M., & Tóvolli, M. H. (2015). *Modeling complex systems for public policies*. Brasília: IPEA.
- Gibrat, R. (1931). *Les Inégalités économiques*. Paris, France: Librairie du Recueil Sirey.
- Gillespie, C. S. (2014). *Fitting heavy tailed distributions: The powerlaw package*. arXiv e-prints, arXiv:1407.3492.
- Hamilton, R., & Rae, A. (2018). Regions from the ground up: A network partitioning approach to regional delineation. *Environment and Planning B: Urban Analytics and City Science*, 1–15. <https://doi.org/10.1177/2399808318804226>.
- ISTAT. (2017). URL: <https://www.istat.it> (accessed: June 2017).
- Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, 11, 129–140. <https://doi.org/10.1007/BF02447512>.
- Jiang, B. (2018). A topological representation for taking cities as a coherent whole. *Geographical Analysis*, 50, 298–313. <https://doi.org/10.1111/gean.12145>.
- Jiang, B., & Jia, T. (2011). Zipf's law for all the natural cities in the united states: A geospatial perspective. *International Journal of Geographical Information Science*, 25, 1269–1281. <https://doi.org/10.1080/13658816.2010.510801>.
- Jiang, B., & Liu, X. (2012). Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *International Journal of Geographical Information Science*, 26, 215–229. <https://doi.org/10.1080/13658816.2011.575074>.
- Jiang, B., & Miao, Y. (2015). The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer*, 67, 295–306. <https://doi.org/10.1080/00330124.2014.968886>.
- Khodabandelou, G., Gauthier, V., Fiore, M., A. El Yacoubi, M. (2018). Estimation of static and dynamic urban populations with mobile network metadata. *IEEE Transactions on Mobile Computing*, 1–1. doi:10.1109/TMC.2018.2871156.
- Liu, Y., Batty, M., Wang, S., & Corcoran, J. (2019). Modelling urban change with cellular automata: Contemporary issues and future research directions. *Progress in Human Geography*, 1–22. <https://doi.org/10.1177/0309132519895305>.
- Long, Y., Zhai, W., Shen, Y., & Ye, X. (2018). Understanding uneven urban expansion with natural cities using open data. *Landscape and Urban Planning*, 177, 281–293. <https://doi.org/10.1016/j.landurbplan.2017.05.008>.
- Lu, D., Li, G., Kuang, W., & Moran, E. (2014). Methods to extract impervious surface areas from satellite images. *International Journal of Digital Earth*, 7, 93–112. <https://doi.org/10.1080/17538947.2013.866173>.
- Ma, D., Omer, I., Osaragi, T., Sandberg, M., & Jiang, B. (2018). Why topology matters in predicting human activities. *Environment and Planning B: Urban Analytics and City Science*, 1–17. <https://doi.org/10.1177/2399808318792268>.
- Makse, H. A., Havlin, S., & Stanley, H. E. (1995). Modelling urban growth patterns. *Nature*, 377, 608–612. <https://doi.org/10.1038/377608a0>.
- Manley, D. (2014). Scale, aggregation, and the modifiable areal unit problem. In M. M. Fischer, & P. E. Nijkamp (Eds.). *Handbook of regional science* (pp. 1157–1171). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-23430-9_69.
- Masucci, A. P., Arcaute, E., Hatna, E., Stanilov, K., & Batty, M. (2015). On the problem of boundaries and scaling for urban street networks. *Journal of The Royal Society Interface*, 12, 20150763. <https://doi.org/10.1098/rsif.2015.0763>.
- Meerow, S., Newell, J. P., & Stults, M. (2016). Defining urban resilience: A review. *Landscape and Urban Planning*, 147, 38–49. <https://doi.org/10.1016/j.landurbplan.2015.11.011>.
- United Nations. (2018). *The World's Cities in 2018*. Department of Economic and Social Affairs, Population Division. Data booklet, ST/ESA/SER.A/417.
- Neteler, M., Bowman, M. H., Landa, M., & Metz, M. (2012). GRASS GIS: A multi-purpose open source GIS. *Environmental Modelling & Software*, 31, 124–130. <https://doi.org/10.1016/j.envsoft.2011.11.014>.
- Oliveira, E. A., Furtado, V., Andrade, J. S., & Makse, H. A. (2018). A worldwide model for boundaries of urban settlements. *Royal Society Open Science*, 5, Article 180468. <https://doi.org/10.1098/rsos.180468>.
- Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography* N. 38, Geo Books – Norwich.
- OSM. (2018). URL: <https://www.openstreetmap.org> (accessed: November 2018).
- O'Clery, N., Curiel, R., & Lora, E. (2019). Commuting times and the mobilisation of skills in emergent cities. *Applied Network Science*, 4, 118. <https://doi.org/10.1007/s41109-019-0235-z>.
- Park, R. E. (1915). The city: Suggestions for the investigation of human behavior in the city environment. *American Journal of Sociology*, 20, 577–612 URL: <http://www.jstor.org/stable/2763406>.
- Rozenfeld, H. D., Rybski, D., Gabaix, X., Makse, H. A. (2009). *The area and population of cities: New insights from a different perspective on cities*. Working Paper 15409. National Bureau of Economic Research. doi: 10.3386/w15409.
- Rybski, D., García Cantú Ros, A., Kropp, J. P. (2013). Distance-weighted city growth. *Physical Review E*, 87, 042114. doi: 10.1103/PhysRevE.87.042114.
- Shutters, S. T., Lobo, J., Munepeeraikul, R., Strumsky, D., Mellander, C., Brachert, M., Farinha, T., & Bettencourt, L. M. A. (2018). Urban occupational structures as information networks: The effect on network density of increasing number of occupations. *PLoS One*, 13. <https://doi.org/10.1371/journal.pone.0196915>.
- Steidl, M., Schleicher, C., Sannier, C. (2016). *Copernicus land monitoring service – high resolution layer imperviousness*. Technical Report version 1 of 2018-12-21. European Environment Agency.
- Stier, A. J., Berman, M. G., Bettencourt, L. M. A. (2020). *COVID-19 attack rate increases with city size*. arXiv e-prints arXiv:2003.10376.
- Taskinen, S., & Warton, D. I. (2011). Robust estimation and inference for bivariate line-fitting in allometry. *Biometrical Journal*, 53, 652–672. <https://doi.org/10.1002/bimj.201000018>.
- Thomas, I., Jones, J., Caruso, G., & Gerber, P. (2018). City delineation in European applications of LUTI models: Review and tests. *Transport Reviews*, 38, 6–32. <https://doi.org/10.1080/01441647.2017.1295112>.
- van Raan, A. F. J. (2019). Urban scaling in Denmark, Germany, and the Netherlands. *Relation with Governance Structures* arXiv e-prints arXiv:1903.03004.
- Vanderhaegen, S., & Canters, F. (2017). Mapping urban form and function at city block level using spatial metrics. *Landscape and Urban Planning*, 167, 399–409. <https://doi.org/10.1016/j.landurbplan.2017.02.014>.

- [org/10.1016/j.landurbplan.2017.05.023](https://doi.org/10.1016/j.landurbplan.2017.05.023).
- Venables, A. J. (2017). Breaking into tradables: Urban form and urban function in a developing city. *Journal of Urban Economics*, 98, 88–97. doi:10.1016/j.jue.2017.01.002 (Urbanization in Developing Countries: Past and Present).
- Viegas, J. M., Martinez, L. M., & Silva, E. A. (2009). Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. *Environment and Planning B: Planning and Design*, 36, 625–643. <https://doi.org/10.1068/b34033>.
- Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews*, 81, 259–291. <https://doi.org/10.1017/S1464793106007007>.
- West, G. B. (2017). *Scale*. Orion Publishing Group, Limited.
- West, G. B., Brown, J. H., & Enquist, B. J. (1999). A general model for the structure and allometry of plant vascular systems. *Nature*, 400, 664–667. <https://doi.org/10.1038/23251>.
- West, J. (2004). Life's universal scaling laws. *Physics Today*, 57, 36–42. <https://doi.org/10.1063/1.1809090>.
- White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A: Economy and Space*, 25, 1175–1199. <https://doi.org/10.1068/a251175>.
- Williams, T., Kelley, C., et al. (2013). *Gnuplot 5.2: an interactive plotting program*. <http://gnuplot.sourceforge.net/>.
- Yin, J., Soliman, A., Yin, D., & Wang, S. (2017). Depicting urban boundaries from a mobility network of spatial interactions: A case study of great britain with geo-located twitter data. *International Journal of Geographical Information Science*, 31, 1293–1313. <https://doi.org/10.1080/13658816.2017.1282615>.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.