geographical analysis

Geographical Analysis (2024) **56**, 384–403

Measuring Spatial Dispersion: An Experimental Test on the *M*-Index

Alberto Tidu, Frederick Guy, and Stefano Usai

Department of Business and Economics, University of Cagliari, Cagliari, Sardegna, Italy

Despite representing a very accurate method for assessing spatial distribution, Marcon and Puech's M *has been insufficiently exploited so far, most likely because its computation relies on pairing every point of interest (i.e., firms, plants) with every other point within the area under analysis. Such a figure rapidly grows to unmanageable levels when said area is larger than a neighborhood or when every industry is taken into account. Consequently, practical applications of* M *have been exclusively experimental and circumscribed to very limited areas or to a handful of industries. This seems much regrettable since* M *provides many advantages compared to conventional measures of spatial distribution and also to alternative distance measures. In this article, we assess the reliability of using small administrative units instead of exact postal addresses for the localization of plants, in order to reduce* M*'s computational burden. Working with a dataset that provides the location, the specific industry and the number of employees for every single plant/establishment in Italy for both manufacturing and services, we can also draw a preliminary but certainly interesting picture of Sardinia's economic geography and its development through the Great Recession toughest years between 2007 and 2012.*

Introduction

Specialized agglomerations – clusters, industrial districts, locations blessed with related variety, places benefiting of localization economies – have a central place in modern economic geography. Measurement of agglomeration, however, is not simple. Distance-based methods for measuring the localization or dispersion of particular industries, use distances between pairs of points representing the locations of individual firms or establishments (hereinafter, *plants*). These methods have clear advantages over earlier methods which were based on summary statistics for specialization by administrative unit, such as provinces or states. The most widely used distance-based methods for studying localization, most notably Duranton and Overman (2005)'s k_d, estimate the marginal likelihood of finding a relevant plant at given distances from a reference plant in the industry in question. Marcon and Puech (2003, 2010, 2017) make a case for benefits of studying the cumulative concentration of relevant plants *within*

Correspondence: Stefano Usai, Department of Business and Economics, University of Cagliari, Cagliari, Sardegna, Italy. e-mail: stefanousai@unica.it

Submitted: April 13, 2023. Revised version accepted: November 10, 2023.

doi: 10.1111/gean.12381

a given radius, rather than (or in addition to) the marginal density on the perimeter of the circle that radius defines; in their 2010 article, they propose an estimator (M) for that purpose.

Marcon and Puech's articles have been widely cited, but their M has been little used; where it has been used, it has been to study a particular sector within some relatively small area. Why might this be so? It could be that the geographical concentration of industries is of less interest over larger areas. Another explanation is the computational burden of estimating M on large datasets.

Much of the literature on spatial concentration of industries does focus on the productivity or innovation benefits of very close proximity. The classical literature on neo-Marshallian industrial districts tends to locate each district in a single small city, where close proximity facilitates both specialization in production and multiple social and civic ties which reduce transaction costs (e.g., Becattini, Pyke, and Sengenberger, 1990). Saxenian (1996) adapted this model to the Silicon Valley; the latter is built as American sprawl, but the enduring image from her account is the informal exchange of knowledge, over drinks, between the employees of different firms who repair to the same bars after work, implying again very close proximity. Writing about knowledge flows in even larger cities such as London and New York, Storper and Venables (2004) again locate "*buzz*" in face-to-face contact facilitated by close proximity. Studying six manufacturing industries in two U.S. cities, Feser and Sweeney (2002) find a scale of clustering on the order of 1–2 km. In this light, it is not surprising that Sweeney and Konty (2005, p. 142) tell us that "studies of inter-firm productivity spillovers should attempt to measure those interactions over small distances to remain consistent with economic theories of the process."

It should be noted that Feser and Sweeney (2002) do report that they are missing physical addresses for 54% of the establishments in the population they study, and that a disproportionate share of those are in outlying areas – suggesting that their results under-estimate dispersion. Duranton and Overman (2005), studying a wider range of U.K. manufacturing industries, find that localization for three-digit sectors is as important at small scales (0–50 km) as it is at a more regional level (80–140 km), finding 38.8% of them to be localized at the 30 km threshold. Note that both Feser and Sweeney (2002) and Duranton and Overman (2005), like much of the literature, are restricted to manufacturing establishments. Dealing as it does with physical products, the determinants of the scale of localization in manufacturing may be different than for the "weightless economy" of services.

Moreover, there are good reasons to be interested in spatial concentrations of industry which span areas much larger than a small city, or a small knowledge-intensive district within a metropolis: on most aspects of productivity- or innovation-enhancing interaction, larger scales are relevant and the question of *how* large a scale is a live one. How large is the area relevant for Marshall's localized pool of skilled labor? A commute of 30–45 min appears to be no hindrance for skilled labor, leading many to treat the Randstad (Amsterdam-Rotterdam-Hague-Utrecht) as a single labor market (van Ham et al., 2001). This puts us, plainly, beyond the realm of a small city or district, and today many skilled jobs mix working-from-home with less frequent commutes, making the relevant area even larger. Venture capitalists tend to require their clients to locate close at hand, but for them proximity is a bit more elastic than it is for daily commuters – the usual standard is reported to be an hour's travel time, the "*one-hour rule*" (Sorenson and Stuart, 2001). Arita and McCann (2000) found that while a disproportionate share of inter-firm research collaborations in the semiconductor industry also hewed close to that rule, there was a second tier of collaborations defined by the feasibility of flying to and from a meeting with the other firm, without staying overnight. Similarly, academic research collaborations increase in number and quality when the participating universities are connected by direct low-cost flights (Catalini, Fons-Rosen, and Gaulé, 2020). Dunford (2006) argues that Italian industrial districts in luxury goods sectors should not be thought of with reference simply to particular small cities, because they are inseparable both from different stages of production taking place in other cities, and from finance, marketing, and other business services found in Milan. Consequently, the relevant spatial concentration in many sectors of industry spans much of north and central Italy. Productivity, quality management systems and "continuous improvement" are all understood as tied up with the "just-in-time" delivery of inputs; classically, the latter is facilitated by co-location (in a small industrial district, or Toyota's suppliers gathered round the mother factory near Nagoya), but with the segmentation of value chains these issues now manifest on many different geographical scales.

In short, benefits of proximity may manifest at different distances, a fact that is often addressed in the literature; it is therefore useful to have measures which can assess spatial concentration over a range of distances. Marcon and Puech's M appears to be a good tool for this but, when used with unique addresses for each firm or establishment, the computational burden can be large enough to render estimation for many industries over a large area impractical.

As an alternative approach, we show that the burden of computing M can be greatly reduced, without any substantial loss of information for proximities of 5 km or more, by using small administrative units in place of exact postal addresses. We show this for the Italian island of Sardinia. Sardinia is the second largest island in the Mediterranean, after Sicily, and has a population of about 1.5 million; it is large enough to have substantial distances between towns, and small enough that it is practical to estimate M for all establishments in all service and manufacturing sectors, using individual postal addresses. We use public data on the location, sector and employment for all plants in manufacturing and services in the years 2007 (121.071 plants) and 2012 (116.623 plants). While this omits agriculture and public institutions, it is a far more diverse set of sectors than is typical of this literature; it is also unusual in that it includes micro-firms. We estimate M for each three-digit sector. We do this first on the basis of geo-located postal addresses, which gives us M_A. We then re-estimate replacing each plant's location with a central point (centroid) of the municipality in which it is located – Sardinia has 377 municipalities – this give us M_{MP}. Using municipal centroids instead of postal addresses allows us to reduce the number of computations by several orders of magnitude; the two measures are so highly correlated that they are virtually interchangeable when the agglomeration is defined at a radius of 10 km or more. Moreover, we also estimate M_{MT} and M_{MR} where distance is refined and is based on, respectively, actual travel time and road distance between centroids: correlation with crow-flies distance between actual locations is still so high, that not only they seem useful estimates of agglomeration, but one might even speculate that whatever accuracy was lost by approximating positions is gained back by having a far more reliable measure of distance between locations.

For the avoidance of doubt, we must stress that our substitution of administrative units for more precise locations is *not* a retreat to measuring the specialization *of* the administrative unit: our measure captures the crucial differences between, for example, the proximity to plants in the same municipality, to those in a neighboring municipality, and to those on the other side of the island.

In the second section of this article, we review the theoretical and methodological background of our problem. The third section describes our data and our method. The fourth section presents our results, first verifying the close correlation of M_A and M_{MP}, M_{MT} and M_{MR}, then using M_A to illustrate some structural changes in the industrial geography of the island, even in this rather brief span of 5 years. The fifth section concludes, with a discussion of implications and possible applications of this method.

Theoretical background

Here, we will sketch out a few aspects of alternative methods for the measurement of localization, as they are important for understanding the motivation for our contribution. For more detail on the history and comparative merits of these methods, you may consult, for example, Marcon and Puech (2003), Fratesi (2008) or – more recently – Bonneu and Thomas-Agnan (2015) and Sweeney and Gòmez-Antonio (2016).

Early measures of localization deal with the relative specialization of particular predefined geographical areas; these areas are typically administrative partitions of some larger territory of interest. This approach arose partly from limitations of data and computing, and partly from the method's usefulness in assessing specialization of politically defined places, which are often the focus of policy. Both the method and question are very similar to what we see in the empirical analysis of international trade, with sub-national units taking the place of nation states: location quotients, very close to the concept of revealed comparative advantage; employment multipliers for the net exports of particular industries.

Ellison and Glaeser (1997) add statistical sophistication to this scheme by testing a place's specialization against a null hypothesis of random spatial distribution.

These methods based on administrative areas have often been used to address localization economies and other aspects of the spatial structure of industry – see, for instance, applications of Ellison and Glaser's method by Rosenthal and Strange (2001), and by Kolko (2010), but there are serious limitations to this approach. Localization will be of interest for some reason: perhaps we are looking for a measure that will proxy or correlate with Marshallian and neo-Marshallian considerations such as improved matching in local labor markets, knowledge spillovers which may be dependent on face-to-face contact, access to and rivalry among specialized suppliers and customers, or the contributions of social networks to any or all of these. Very small administrative units will be too small to encompass the relevant interactions; very large ones extend beyond the local; actors located near the boundary of a unit may interact closely with actors on the other side of that boundary; and actual institutions often produce a mix of very large and very small units. California contains multiple local agglomerations which we may not want to lump together, while a firm in Rhode Island is in close proximity to firms just over the border in Massachusetts and Connecticut, and it may be inappropriate to ignore those interactions. In formal terms, this is the modifiable areal unit problem (Openshaw and Taylor, 1979); its impact has been demonstrated by, among others, Kopczewska (2018).

In response, and aided by improvements in both data and computing power, we now have localization indices based on distances between pairs of firms or establishments, escaping the constraints of administrative units. Where the object of study is a particular industry, the measurement of localization (or dispersion) becomes some version of the question "for any given firm (or plant), what is the likelihood of finding another firm in the same (or a related) industry within (or at) a certain distance, relative to some counterfactual?" The most commonly employed counterfactual, or null hypothesis, is that the spatial distribution would be the same as the distribution of some broader category of establishments: we might estimate the spatial concentration of employment in shoe manufacturing relative to manufacturing as a whole, or relative to employment as a whole.

Distance-based models begin with Ripley's (1976, 1977) K function, which is widely used in statistical literature but is problematic for economic applications, since it relies on a counterfactual of a completely random point distribution ignoring, for instance, the differences between city and countryside (Floch, Marcon, and Puech, 2018) and is therefore unable to account for the spatial non-homogeneity that is typical of many types of spatial distributions, such as industrial activities. During the following two decades, other functions were introduced in order to account for non-homogeneity of space: most notably D (Diggle and Chetwynd, 1991) – which is simply the difference between two Ripley's K functions and allows the comparison between distributions of two sub-populations (e.g., cases vs. controls, manufacturing firms vs. service firms) – and *Kinhom* (Baddeley, Møller, and Waagepetersen, 2000). A later addition – notably exploited by Feser and Sweeney (2002) and Sweeney and Feser (2004) – was the introduction of weights, that would allow to take into account different firms' size and relevance, on the basis of, for instance, employment or revenue.

In economic geography, the most widely used distance-based model is that of Duranton and Overman (2005). They propose five characteristics that sound distance measures should have: .

1. They should be comparable across industries;
2. They should control for overall agglomeration trends across industries;
3. They should separate spatial concentration from industrial concentration;
4. They should be unbiased with respect to the degree of spatial aggregation; and
5. They should provide an indication of the significance of the results.

Duranton and Overman's *Kd* respects the five criteria they had identified, and is a function of density of the probability of finding a neighbor at a certain distance.

Duranton and Overman's *Kd* is a marginal density estimate. That is, for each plant in some industry of interest – call it industry A – we consider a distance from that plant r and ask "how likely are we to find other plants in industry A at that distance from the first plant, compared with what the likelihood would be if plants in industry A were distributed randomly across the population of industrial sites?"

Marcon and Puech's (2010) M is, in contrast, a cumulative function: for a given distance from a plant, it answers the question "within a radius r of a representative plant in industry A, can we expect industry A's share of all industry to be greater, or less, than industry A's overall share?"

For any particular industry, either *Kd* or M can be plotted as a function of distance. If we imagine a simple case of an industry with a single cluster of plants in one small area – say for simplicity a circular area with a radius of 1 km. Each plant in that industry will be within 1 km to many similar plants, so for distances of a kilometer or less *Kd* will be large; beyond 1 km, *Kd* will drop abruptly to zero. M, on the other hand, will taper off gradually. Marcon and Puech (2010) note that, in such a case, *Kd* gives a more precise picture of the localization pattern than M does; however, with more complex (and realistic) spatial patterns – either multiple, randomly scattered clusters of localized firms; or repulsion, with widely spaced firms avoiding proximity to competitors – *Kd* becomes difficult to interpret, and indeed has a problem distinguishing between localization and repulsion. For this reason, they recommend treating the two approaches as complementary, and further recommend starting with M to get an overall picture of the situation, before moving to *Kd* to understand how tight the localization is within clusters. If one were to follow such an approach, the impossibility to map M correctly for very short ranges with our proposed approximation (since median distance between municipalities hovers around

5 km for Sardinia) would be even less of a problem, because the second phase would be in charge of revealing precise geographic aspects, whereas *M* would provide a general ranking of the agglomeration/dispersion level of each industry at the country level.

Applications of distance-based indices

Duranton and Overman (2005) demonstrated their approach in a study of the location patterns in manufacturing sectors in the United Kingdom, controlling for the overall tendency to agglomerate (i.e., the null hypothesis is not a random distribution all over the United Kingdom, but only the locations currently used by a manufacturing establishment). Defining dispersion "as having fewer establishments at distance d than randomness would predict" and thus featuring a distribution that appears "too regular" (Duranton and Overman, 2005, p. 1086), they found that 52% of industries exhibited localization at a 5% confidence level, with 24% of them showing dispersion at the same confidence level, highlighting a nonrandom distribution across space. Duranton and Overman's method has been applied in many subsequent studies to compare agglomeration levels across industries, and also to assess the determinants of agglomeration For Japan, Nakajima, Saito, and Uesugi (2012) find that about half of the 561 four-digit manufacturing industries they studied can be classified as localized, in contrast with a lower figure of only about 35% for service industries, also concluding that "industries are becoming neither more concentrated nor more dispersed and the location patterns are stable over time" (Nakajima, Saito, and Uesugi, 2012, p. 18). Barlet, Briant, and Crusson (2013) study the location patterns of business-oriented service and manufacturing industries in France; they use an improved version of the *Kd* index, which takes into account the number of plants in each industry. They show that concentration is more present among service industries (61%) than manufacturing industries (42%), especially at short distance. For Germany, Koh and Riedel (2014) assess the agglomeration patterns of four-digit industries using the *Kd* index. They find that 71% of manufacturing industries are localized while this ratio reaches 97% for the service industries. In line with the results above, Behrens and Bougna (2015, p. 48) find that "depending on industry definitions and years, 40% to 60% of manufacturing industries are clustered" and that localization in Canada has generally decreased during recent years. Cainelli, Ganau, and Jiang (2020) found that most Italian manufacturing industries experienced spatial dispersion processes during the period of the Great Recession. Moreover, their results indicate that space–time dispersion processes occurred within small spatial distances and a short time horizon, although space–time interactions do not seem statistically significant. Brakman, Garretsen, and Zhao (2017) examined the location of manufacturing in China and found that around 80% of industries at four-digit in China are significantly localized. Moreover, they found that localization increased rapidly in the period between 2002 and 2008, especially as a consequence of new entrants. Aleksandrova, Behrens, and Kuznetsova (2020) analyzed the agglomeration and co-agglomeration patterns of manufacturing industries in Russia and found that 80% of three-digit industries are both agglomerated and co-agglomerated. Almeida, Neto, and Rocha (2020) found that almost 90% of Brazilian manufacturing showed statistically significant localization for 2006 and 2015.

Marcon and Puech's, 2010 article has been widely cited, but in contrast with Duranton and Overaman's *Kd*, few applications using *M* have been published. Those which have all deal with small areas and/or narrow selections of sectors: Jensen and Michel (2011) who use *M* to infer the spatial pattern of stores in Lyon (France) and Marcon and Puech (2014) themselves when describing the distribution of pharmacies in Lyon weighted against the distribution of nonfood retail stores. Méndez-Ortega and Arauzo-Carod (2019) compute M for creative industries and for software-developing industries in Barcelona metropolitan area, underlining how such measures provide the great advantage of being *relative* and not *absolute* (such as Duranton and Overman's Kd), thus comparable between industries and years. Moreno-Monroy and Garcìa-Cruz (2016) use M to assess the degree of spatial agglomeration and co-agglomeration of *formal* versus *informal* manufacturing activity within Cali metropolitan area in Colombia. Finally, Zhang et al. (2021), use M to explore the geographic concentration of five manufacturing industries in the Chinese urban region of Jiangsu, relying on firm-level data.

Notice that all applications of M have concerned either individual industries and/or individual urban areas: it has not been used for comparative study of sectors across a whole country, or even a large region. A possible reason for this is that M is much more computationally intensive than Kd. Computation of either M or Kd for a particular industry starts by dividing the population of plants into A (the industry in question) and Not A (all other plants or sites within some universe of sites). Computing Kd for an industry requires the computation of all plant-to-plant distances within that industry, for a total of $n_A(n_A - 1)/2$ distances, where n_A is the number of plants in industry A. The counterfactual of random distribution is obtained by drawing repeated samples of the same size – n_A – from the Not A population, and computing distances. M, on the other hand, requires computing distances not only between firms in industry A, but also between each firm in A and all firms in Not A. Unlike Kd, M has a counterfactual built in ($M > 1$ is more localized than we would expect from a random distribution, $M < 1$ is more dispersed), but if we wanted confidence intervals to test localization or dispersion we would need further computations.

Methods and data

We stressed above a key difference between Kd and M in the number of distance computations required. We need to note two other factors – in addition to the choice of index – which affect the necessary number of such computations. These are the larger population in which an industry of interest sits, and the precision with which the location of a plant is specified.

The definition of the larger population itself may have two aspects: it provides a sampling frame from which to draw the random-location counterfactual, and it may also consist of other industries for which localization indices are to be estimated. Duranton and Overman study localization in U.K. manufacturing industries and they use, for their larger population, all manufacturing plants in the United Kingdom. They note that, for sampling frame purposes, they could have used many things – postcodes, perhaps. Of greater interest for our purposes is that they could have used plants in a broader range of sectors. Duranton and Overman's article is primarily a demonstration of the use of Kd, and does not pretend to be a comprehensive comparison of localization in different sectors of the U.K. economy; if we want to move on to make such a comparison, however, we should surely add services. For the practicality of estimating Kd, adding services to the sampling frame would actually make little difference, because for each industry the number of distances computed will continue to be governed by the number of plants in that industry. For the computation of M, however, it further expands the number of firm-firm distances which must be calculated.

As for the precision with which plant location is specified: using distance methods, it is perhaps natural to want to use the most precise locations available. If a plant's street address is available, the plant can be geo-located. For many plants, this yields a unique location; for others – such as offices in a large building – it may put several at the same location. Duranton and Overman use, instead of street addresses, full U.K. postcodes, which are not quite as precise. The United Kingdom has about 1.7 million full postcodes; the average full U.K. postcode covers 15 properties or "deliverable endpoints," and a postcode can cover as many as one hundred. Since a postcode usually contains multiple street addresses, and may include more than one business establishment, the use of postcodes rather than street addresses offers the prospect of economizing on computations somewhat. But one could go further. Brakman, Garretsen, and Zhao (2017) calculate Duranton and Overman's *Kd* for Chinese manufacturing firms: in their case they were not concerned with computational feasibility, but simply lacked data on the actual addresses of firms, so they use the county in which the firm was located. Not knowing the precise location, they were unable to assess the effect of using this approximation, but they note that the mean value of intra-county distances (19 km) is very small compared with the median value of all pair-wise distances between manufacturing firms in China (around 900 km).

In this article, we are comparing estimates of M based on great-circle distances between street addresses (M_A) with estimates based on great-circle distances between the central points of Sardinia's 377 municipalities (M_{MP}). Moreover, we are also providing a comparison with estimates of M based on – respectively – travel time (M_{MT}) and road distance (M_{MR}) between municipalities' centroids. Our different estimates of M are listed in Table 1.

Our data provides the address, sector and number of employees for approximately 115,000 nonagricultural business establishments on the Italian island of Sardinia. Treating each address as unique gives us 6,612,500,000 great-circle distances to compute for M_A (this ignores a small computational economy which might be achieved by taking into account the fact that multiple establishments may share an address – see above). On the other hand, only 142,129 great-circle distances need to be computed for pairing each of the 377 municipalities in Sardinia, in order to estimate M_{MP}. This produces, by several orders of magnitude, a greater reduction in computing resources than the efficient algorithm of Scholl and Brenner (2015), although of course the latter does retain the advantage of using a unique address for each establishment.

The estimation of M_{MT} and M_{MR} does not even require the computation of any distance, because reliable estimates of travel time and road distance between each pair of municipalities are provided by ISTAT, which puts them into an origin–destination matrix and distributes them to the public as a free-access spreadsheet. Our data are an extract from ISTAT's ASIA-UL datasets for 2007 and 2012. The datasets cover all of Italy; we use only the observations for the region of Sardinia. We limit ourselves to a single region because it would not be practical to compute M_A for all nonagricultural business establishments in Italy as a whole: we are using the region simply to test the difference between M_A and M_{MP}, M_{MT} and M_{MR}. If the latter are acceptable substitutes, then they could be used for a larger population of establishments, such as those in Italy as a whole.

Dealing with an island avoids edge effects that would make establishments just across the regional border disappear. Table 2 shows that Sicily is the largest and by far the most populous Italian (and Mediterranean) island, with over 5 million inhabitants, more than three times the

Table 1. Our Estimates of M

	Crow-flies	Road distance	Travel time
Actual locations	M_A	/	/
Approximated locations	M_{MP}	M_{MR}	M_{MT}

Source: Compiled by the authors.

Table 2. Plant and Municipalities in Elba, Sardinia, and Sicily in 2012

	Plants	Plant pairs	Municipalities	Municipality pairs
Elba	3,627	6,575,751	8	28
Sardinia	116,623	6,800,403,753	377	70,876
Sicily	293,212	42,986,491,866	390	75,855
Italy	4,826,882	11,649,392,507,521	8,056	32,445,540

Source: Compiled by the authors.

population of Sardinia: it has enough establishments to make calculation of M_A daunting; moreover, edge effects across the narrow Messina Strait would probably be relevant due to proximity between the opposite metro areas of Messina and Reggio Calabria. Other Italian islands are too small, particularly in terms of distances, but also number of establishments, and number of sectors: the next largest after Sardinia, Elba, has only seven municipalities and slightly over 30,000 inhabitants. For Sardinia, computation of M_A is tractable, yet the island offers sufficient inter-establishment distances, and sufficient number and diversity of business establishments, to make measures of localization meaningful.

ASIA (*Archivio Statistico delle Imprese Attive*) is a register established in 1996 in accordance with the provisions of European Council Regulation No. 2816/93 on Community coordination in drawing up business registers for statistical purposes, later replaced by Regulation (EC) No. 177/2008, and according to a harmonized methodology adopted by Eurostat. Since 1996, ASIA covers every currently active enterprise[1] that contributes to gross domestic product, in the fields of manufacturing, trade and services, providing name, address, field of activity (at a five-digit detail of the ATECO 2007 classification, which is directly derived from the European NACE Rev. 2 nomenclature), number of employees, legal form, turnover class, and dates of creation and cessation. Economic activities not included in ASIA are: agriculture, forestry, and fishing; public administration and defense; compulsory social security; activities of membership organizations; activities of households as employers; undifferentiated goods- and services-producing activities of households for own use; activities of extraterritorial organizations and bodies; units classified as public institutions and private nonprofit institutions. ASIA is updated every year through a process that integrates several administrative and statistical sources[2], guaranteeing a proper statistical representation of active enterprises and of their identification, demographic and economic information. The register has a central role within economic statistics, and it is used for national accounting estimates. ASIA-UL (Registro statistico delle Unità Locali) is a supplement to ASIA, providing additional data (location, sector, number of employees) in respect to individual plants.

M_A, M_{MP}, M_{MT}, and M_{MR} are all estimated using

$$\widehat{M}(r) = \frac{\sum_i \frac{\sum_{j \neq i} 1\left(\|x_i - x_j^c\| \leq r\right) w\left(x_j^c\right)}{\sum_{j \neq i} 1\left(\|x_i - x_j\| \leq r\right) w\left(x_j\right)}}{\sum_i \frac{W_c - w(x_i)}{W - w(x_i)}}$$

where x_j^c are neighbour establishments in sector, c and x_j are neighbour establishments in all sectors, r is the selected distance, w a weighting for the number of employees in the establishment, W_c is the total number of employees in sector c in Sardinia, and W is the total number of employees in all establishments.

For M_A and M_{MP}, an establishment's location is its geo-located street address. For M_{MT} and M_{MR} the location is the centroid of the municipality in which the establishment is located, as determined by ISTAT. Distances for M_A and M_{MP} are calculated as great circle distances (or as the *crow-flies*); for M_{MT}, ISTAT's inter-municipality travel time in minutes; for M_{MR}, ISTAT's inter-municipality road distance, in kilometers. For both M_{MT} and M_{MR}, the distance between establishments in the same municipality is assumed to be zero. In all calculations, establishments are weighted by number of employees.

We compute M_A, M_{MP}, M_{MT}, and M_{MR} at four different distance ranges – 5, 10, 15, and 20 km, in two different years, 2007 and 2012[3]. These are, respectively, the last year before the global financial crisis, and the first postcrisis year to show an increase both in the number of firms and in the number of employees[4].

Results

Table 3 summarizes calculations of M_A for the four distance ranges for 2007, 2012, and the change from 2007 to 2012. Means and standard deviations are weighted by the number of plants in each industry. By construction, M can be computed only for industries with at least two plants, therefore industries that feature only one plant[5] are not included in the results (although those plants were still taken into account as neighbors for the computation of other industries' M). Every industry featuring over five plants shows measurable agglomeration (i.e., it has at least two plants within less than 20 km from each other) and results are consistent between 2007 and 2012, showing strikingly similar means[6]. On the other hand, the apparently large difference in maximum values between 2007 and 2012 is entirely attributable to very small industries: if we only include industries with at least 10 plants and 100 employees, the largest value that M assumes for the 5-km distance range in 2007 amounts to 32.19 for *Manufacture of cement, lime and plaster (235)*, which is also the most agglomerated industry in 2012 with a remarkably similar value of 31.89.

Table 4 shows that M does not lose much accuracy when it is computed after approximating plants' positions to the centroid of the municipality where each one is located. The weighted correlation[7] between M_A and M_{MP} is close enough to one, especially at the higher (15- and 20-km) distance ranges. This should be expected because with the shorter ranges, the impact of the approximations, and particularly the treatment of same-municipality distance as zero, will have a proportionally larger impact. The high correlation between results obtained with approximate and with actual plants' positions also holds for changes in values of the indices between 2007 and 2012.

Table 3. Statistics for M_A – Sardinia (2007, 2012, and Variation Rate 2007–2012).

	2007		2012		%Δ 2007–2012	
	Mean	Min/Max	Mean	Min/Max	Mean	Min/Max
MA (5 km)	1.43	0/262	1.43	0/112	5.32	−100/34195
MA (10 km)	1.25	0/237	1.28	0/67	4.45	−100/7033
MA (15 km)	1.17	0/102	1.20	0/52	3.85	−100/5392
MA (20 km)	1.13	0/73	1.16	0/39	3.96	−100/3685
Observations	224		219		216	

Source: Compiled by the authors.

Table 4. Weighted Correlation Between M_A and M_{MP}, M_{MT} and M_{MR}

	2007			2012			Δ 2007–2012		
	M_{MP}	M_{MT}	M_{MR}	M_{MP}	M_{MT}	M_{MR}	M_{MP}	M_{MT}	M_{MR}
MA (5 km)	0.89	0.90	0.90	0.94	0.60	0.90	1.00	0.99	1.00
MA (10 km)	0.96	0.87	0.86	0.98	0.80	0.80	0.96	0.83	0.83
MA (15 km)	0.99	0.94	0.96	0.99	0.98	0.97	0.99	0.81	0.98
MA (20 km)	0.98	0.97	0.97	0.98	0.96	0.97	0.97	0.96	0.96

Source: Compiled by the authors.

Table 4 shows high correlation also between M_A and both M_{MT} and M_{MR}, that is when distance between centroids is computed not as the crow-flies but, respectively, through travel time in minutes and actual road distance in km. Indeed, one might speculate that the slightly lower correlation could depend just as much on the inaccuracy produced by crowflies distance ignoring physical obstacles between establishments, as on the approximation of the establishments' positions. Verifying such a hypothesis is certainly feasible – albeit probably somewhat pricey – since it would require querying Google API's for distances between every pair of establishments (instead of relying on a simple formula that provides geodesic distance between geographical coordinates), and then replicate our same work with a new – larger but still manageable – distance matrix.

Although 5 years is very short for any study of changes in industrial geography, it is in this case long enough for us to use our estimates to describe aspects of such change in Sardinia. In this particular time frame, such estimates might be viewed as showing changes brought about by the global financial crisis. The same method, applied over longer time frames, would be useful in studying changes in the agglomeration or dispersion of different industries within a larger geographical unit – region, nation state, or an entity such as the European Union.

In Fig. 1a we plot – for each sector with at least 15 plants in both 2007 and 2012 (166 industries out of 233 that had at least one operating plant in either year) – the changes in employment (2007–2012) against the sector's agglomeration index (M_A) in 2012 (using M_A for 2007 yields a very similar picture) within a 20-km radius. The aim here is to visualize both which sorts of industries are more localized, and whether there is any discernable pattern in the relationship between localization and the growth (or shrinkage) of industries. This could have implications for local housing and transport policies, as well as for industrial policies. We stress, again, that this is simply a conceptual illustration of a tool – the time period is short, and the population covered is not large; however, using the municipality-based approximation, it would be straightforward to do the same for larger places over longer periods.

There are a few outliers in terms of growth: *Other social work activities without accommodation (889)*, *Other residential care activities (879)* and *Remediation activities and other waste management services (390)* had over 500% growth in terms of employees, but also *Sale of motor vehicle parts and accessories (453)*, *Landscape care and maintenance service activities (813)*, *Other information service activities (639)*, *Gambling and betting activities (920)* and *Investigation activities (803)* increased their employees by well over 100%. Trimming these growth outliers, we can zoom in to the sub-set plotted in Fig. 1b.

Most industries do not show a particularly strong level of either agglomeration or dispersion, with their M hovering between 0.5 and 1.5. Among these are those with the highest levels of employment change: *Waste treatment and disposal (382)*, *Other financial service activities except*

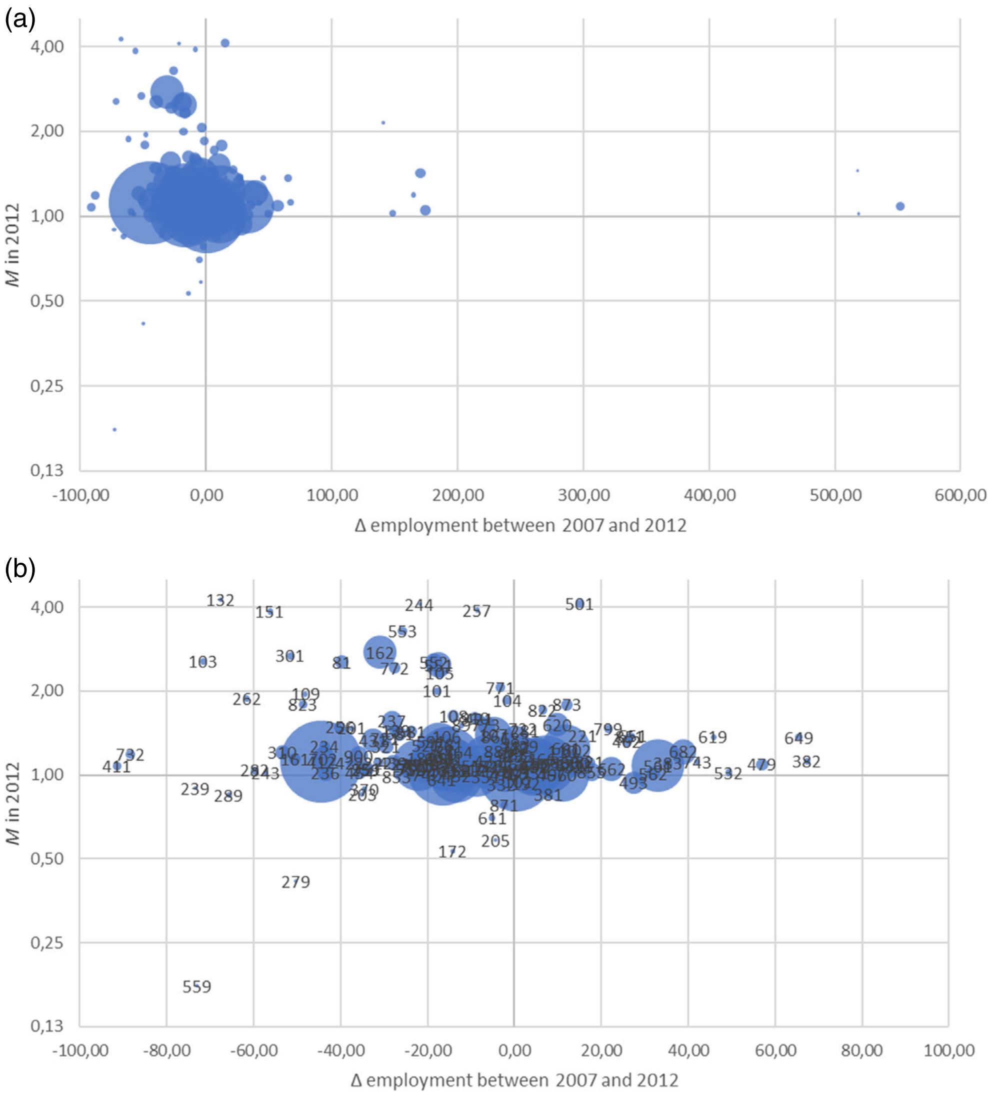

Figure 1. (a) *M* (2012) versus change in employment (2007–2012) by industry. Source: Compiled by the authors. (b) *M* (2012) versus change in employment (2007–2012) by industry (outliers removed). Source: Compiled by the authors.

insurance and pension funding activities (649), *Retail trade not in stores, stalls or markets (479)* which have seen an increase in employment between 50% and 100%, and on the other side of the spectrum *Project management activities related to construction (411)*, *Market research and public opinion polling (732)* and *Manufacture of other nonmetallic mineral products N.E.C. (239)* which have seen a comparable reduction.

Looking upward on the plot, we start with the north-east quadrant. This is where we should see sectors with high localization and high employment growth. It is, essentially, empty since the only industry with a somehow high *M* is *Sea and coastal water transport (501)* with a 15% growth in both employment and number of plants between 2007 and 2012. Except the aforementioned outliers and the really small industries excluded from the plot, every other growing industry has an M lower than 1.8 – and even lower than 1.4 if one wants to find employment growth higher than the 20% threshold.

In the north-west quadrant, we see industries with above-average localization and above-average employment shrinkage. Not only this quadrant is far more populated, but many strongly agglomerated industries show a very high reduction in employment: among them, *Weaving of textiles (132)*, which has the highest M in the plot and saw a decrease in employment of almost 70% between 2007 and 2012; other agglomerated industries that showed strong reduction in employment are *Tanning and dressing of leather; manufacture of luggage; handbags; saddles and harness; dressing and dyeing of fur (151), Processing and preserving of fruit and vegetables (103), Building of ships and boats (301)* and also a very large industry such as *Manufacture of products of wood, cork, straw and plaiting materials (162)* that started out with almost 1500 plants and 5000 employees and was almost three times more agglomerated than business activities in general.

Another use of this tool is simply to rank industries by degree of localization. Again, ignoring the industries with the smallest employment numbers – and, by construction, those with only a single plant – we see the following:

> *Manufacture of basic precious and other nonferrous metals; reprocessing of nuclear fuels (244), Manufacture of knitted and crocheted apparel (143)*, *Manufacture of agricultural and forestry machinery (283)*, and *Manufacture of refined petroleum products (192)* are consistently among the most agglomerated in both years.
> Localization is certainly not limited to manufacturing industries: *Camping grounds, recreational vehicle parks and trailer parks (553)*, and *Sea and coastal water transport (501)* are consistently on top of the ranking in both years, as it is quite natural for activities respectively related to tourism and to water transport, as demonstrated – albeit less strongly – by *Hotels (551)* and *Other short term accommodation activities (559)*. Even more predictably, those industries "for which location is constrained by natural advantages" (Guillain and Le Gallo, 2010, p. 969) – such as *Building of ships and boats (301)* and, especially, *Quarrying of stone, sand and clay (081), Mining and quarrying N.E.C. (089)*, and *Mining of hard coal (051)* – are all among the most agglomerated.

Whereas the persistency on the highest positions of the ranking might indicate actual agglomeration for small industries instead of mere chance, we might have a harder time when trying to infer dispersion for industries on the bottom: indeed, while we could argue that an industry consisting of five plants located close to each other for the best part of a decade might hint to an actual reason behind such proximity, the same number of plants located far from each other might very well not be driven by any particular dispersion force but by random chance. Therefore, it seems far more sensible to focus on those industries that manage to keep their plants decently far away from each other despite featuring many of them. Some of these industries are certainly predictable and this might be interpreted as a sign that our index is indeed representing dispersion as we would expect it to (Table 5):

- *Postal activities (531), Monetary intermediation (641), Waste collection (381), Medical and dental practice activities (862), Electric power generation, transmission and distribution (351), Other passenger land transport (493), and Wired telecommunications activities (611)* satisfy a public interest that requires them to be geographically dispersed and to follow the general population pattern rather than economic activity (with the latter supposedly more geographically concentrated than the former);

Table 5. 10 Most Agglomerated and 10 Most Dispersed Industries within a 15 km Radius in Sardinia in 2007 and 2012 (Only Industries with >15 Plants and >50 Employees in Both Years)

Industry code	Industry description	Rank 2007	Rank 2012	Employees 2012	Plants 2012
257	Manufacture of cutlery, hand tools, and general hardware	1	1	54	34
501	Sea and coastal water transport	3	2	594	114
132	Weaving of textiles	38	3	248	30
553	Camping grounds, recreational vehicle parks, and trailer parks	2	4	502	85
192	Manufacture of refined petroleum products	7	5	1449	31
201	Manufacture of basic chemicals, fertilizers and nitrogen compounds, plastics and synthetic rubber in primary forms	4	6	1289	34
105	Manufacture of dairy products	11	7	1674	161
551	Hotels	10	8	7000	744
301	Building of ships and boats	16	9	139	58
162	Manufacture of products of wood, cork, straw, and plaiting materials	12	10	3432	1088
332	Installation of industrial machinery and equipment	141	153	575	157
493	Other passenger land transport	151	154	4957	727
243	Casting of semi-finished steel products	105	155	70	20
381	Waste collection	153	156	3155	143
611	Wired telecommunications activities	157	157	1724	41
221	Manufacture of rubber products	86	158	263	24
205	Manufacture of other chemical products N.E.C.	154	159	89	17
239	Manufacture of other non-metallic mineral products N.E.C.	24	160	68	23
203	Manufacture of paints, varnishes and similar coatings, printing ink, and mastics	160	161	105	23
279	Manufacture of other electrical equipment	162	162	100	18

Source: Compiled by the authors.

Table 6. 10 Industries with the Largest % Increase and 10 Industries with the Largest % Decrease in Agglomeration in Sardinia between 2007 and 2012 (only Industries with ≥15 Plants and >50 Employees in Both Years)

Industry code	Industry description	Δ% Employees	Δ% Plants	Δ% *M*
132	Weaving of textiles	−67.55	−18.92	206.31
772	Renting and leasing of personal and household goods	−27.60	−17.17	109.21
429	Construction of other civil engineering projects	2.11	−3.55	94.10
619	Other telecommunications activities	45.74	73.33	71.53
329	Other manufacturing N.E.C.	−0.09	20.41	70.45
782	Temporary employment agency activities	0.38	−21.54	66.70
421	Construction of roads and railways	−8.11	−1.74	63.82
89	Mining and quarrying N.E.C.	−12.18	−14.29	56.34
803	Investigation activities	140.78	65.00	51.86
301	Building of ships and boats	−51.54	−25.64	48.86
139	Manufacture of other textiles	−27.39	−28.08	−29.46
257	Manufacture of cutlery, hand tools, and general hardware	−8.59	−10.53	−31.48
233	Manufacture of clay building materials	1.04	27.78	−33.23
532	Courier activities	49.36	51.43	−33.62
205	Manufacture of other chemical products N.E.C.	−4.41	−10.53	−34.76
203	Manufacture of paints, varnishes and similar coatings, printing ink, and mastics	−35.07	−34.29	−36.17
172	Manufacture of corrugated paper and paperboard and of containers of paper and paperboard	−14.27	−18.18	−39.72
871	Residential nursing care facilities	−2.32	−51.47	−42.30
221	Manufacture of rubber products	15.72	20.00	−43.90
239	Manufacture of other non-metallic mineral products N.E.C.	−73.44	−20.69	−72.11

Source: Compiled by the authors.

- *Electrical, plumbing, and other construction installation activities (432), Cleaning activities (812), Photographic activities (742)*, and most types of retailing activities, professional services and restoration do not directly satisfy a public interest strictu sensu, but still rely mostly on individual customers, with less room for economies of scale at plant level and, consequently, geographic concentration (Table 6).

As concerns manufacturing activities, *Manufacture of structural metal products (251)* and *Manufacture of medical and dental instruments and supplies (325)* are the only adequately sized industries that appear dispersed for every distance range both in 2007 and 2012.

In apparent contrast with Cainelli, Ganau, and Jiang (2020, p. 443), who found that "Italian manufacturing sectors experienced a process of space-time dispersion during the period of the Great Recession, although with slightly different intensity and patterns," descriptive statistics provided in Table 3 shows a tiny increase in the weighted mean of agglomeration results for every distance range. Indeed, even when focusing on the change for individual industries (accounting for the number of plants each industry consists of), there seems to be a slight percentage increase in M between 2007 and 2012. Moreover, our findings also contrast with De Dominicis, Arbia, and De Groot (2013, p. 5), who observed that "whereas manufacturing has been spreading out, service activities have become increasingly clustered," and instead they seem to suggest that manufacturing activities have generally clustered more than service industries, not less.

Conclusion

Distance-based measures of industrial localization have clear advantages over earlier ones based on summary statistics for administratively- or politically defined areas. In economic geography, the distance-based tool of choice has been Duranton and Overman's (2005) Kd, a marginal estimator. Marcon and Puech (2010) have made a strong argument for the use of a cumulative estimate, in addition to marginal ones. The tool they propose, M, is computationally intensive, and probably for that reason is far more cited than used. We propose that, by using for a plant's location a central point in a small administratively-defined area, most of the advantages of distance-based measures can be retained, even as the computational burden is sharply reduced. We demonstrate this with data from a census of nonagricultural business establishments in the Italian island-region of Sardinia: correlations between M_A (based on crow-flies distances between geo-located street addresses) and M_{MP} (based on crow-flies distances between municipal centroids), M_{MR} or M_{MT} (both based on road distances and travel times between municipal centroids, respectively), are very high, especially when M_A is measured at a greater distance (15–20 km) where the correlation is on the order of 0.97–0.98. This is in keeping with Marcon and Puech's proposition that "cumulative functions are insensitive to errors at smaller scales than the distance they consider: if the uncertainty is a few hectometers, the number of neighbors up to a few kilometers is known with no error except for the more distant ones, which are a small proportion" (2017, p. 30). On the other hand, one should also be aware that clustering often occurs at a scale that is smaller (Feser and Sweeney, 2002) than the mean distance between Sardinian administrative boundaries, and in that case our approximation would be unable to pick it up correctly.

With the small adjustment we have employed, and with the increasing availability of micro-data, a cumulative tool such as M could be used to study industry localization and dispersion across, for instance, a large nation state. It is also of course useful in cases where individual firm- or establishment data is available, but is not precisely geolocated.

We have relied on comprehensive data provided by ISTAT – the Italian Institute of Statistics – to measure agglomeration for Sardinian industries in 2007 and in 2012. We believe our contribution is relevant with respect to both the methodological approach and the results obtained. Indeed, our operationalization validates an innovative way to use an accurate measure such as Marcon and Puech's M, whose experimentation had so far been restricted by its unmanageable computational intensity to the limited scope of individual city neighborhoods. Ours should not be taken as a general solution, though: mean distance between municipalities is about 5 km, so our proposed approximation would be unreliable for exploring agglomeration at a lower scale than such a threshold: its purpose is not to localize agglomerations exactly, but to assess the propensity of industries to either agglomerate or disperse in an area large enough to contain most sectors. With such a caveat in mind, our method extends M's implementation possibilities to the study of larger geographic regions and even entire countries, as already pioneered by Tidu (2021). This is of the utmost importance because it offers an alternative to the passive acceptance of the distortions caused either by the Modifiable Areal Unit Problem or, alternatively, by the absence of a benchmark when relying on more commonly used distance-based methods, such as Duranton and Overman's Kd. With micro-geographic data becoming increasingly available (Arbia, 2001), it is crucial to learn how to exploit their whole potential when researching economics. Sardinia was chosen as the target of our study because of a demographic and economic size that make the island's data at the same time computationally manageable but economically relevant.

Our results show an extremely high degree of correlation between M computed with actual plants' locations and M computed by approximating such locations to the centroids of the municipalities where the plants are located. The picture that emerges from our estimates of M for Sardinia broadly accords with the stylized facts which emerge from the literature on localization. In particular, we see localization slightly decreasing (Behrens and Bougna, 2015; Almeida, Neto, and Rocha, 2020) during the Great Recession, albeit with the most agglomerated industries – especially manufacturing ones – maintaining a high degree of agglomeration, and sometimes even showing an increase (Behrens and Bougna, 2015). Literature is conspicuous on the matter and it would certainly be interesting to compare it with results obtained through M, even distinguishing on the basis of firms' characteristic, such as plant's size (Sweeney and Feser, 1998). Such results, however, are clearly preliminary and – at best – only accessory to our actual goal: showing that our proposed approximation is reliably correlated with Marcon and Puech's method. Therefore, we do not aim to disprove previous literature, especially since we have not run a significance test to verify which results should be discarded for any further analysis. On the other hand, we believe that – as preliminary as they might be – our results hint to paths that could be explored through M, now that our methodology has shown a viable way to exploit it even for large areas and without the need to develop complex algorithms or to deal with overbearing numbers of computations.

Acknowledgements

We feel thankful toward Ernesto Batteta for his technical expertise, and Alfredo Tidu who had the initial intuition of the geographical approximation whose validity we explore in this contribution. Furthermore, we also thank Simona Iammarino and Emanuela Marroccu who helped develop the conceptual framework in the initial phase of our work, and everyone who attended our seminars and the conferences where we presented this work, contributing with precious advice and suggestion. Similarly, we strongly believe that the referees that reviewed this contribution were fundamental in defining some aspects that improved substantially our work.

We also gratefully acknowledge the Sardinian Regional Government for the financial support during the initial development of the work that led to this article, through a PhD scholarship (P.O.R. Sardegna F.S.E.).

Notes

[1] Defined by ISTAT's quality report (https://www.istat.it/it/archivio/216767), in accordance with European Council Regulation No. 696/93, as "the smallest combination of legal units that is an organizational

unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit."

2 Agenzia delle Entrate; INAIL (Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro); CCIAA (Camere di Commercio, Industria, Agricoltura e Artigianato); Banca d'Italia, INPS (Istituto Nazionale della Previdenza Sociale); Seat – pagine gialle Spa; ISVAP (Istituto per la Vigilanza sulle Assicurazioni Private e di Interesse Collettivo).

3 It is not a coincidence that those same years were also chosen by Cainelli, Ganau, and Jiang (2020), who acknowledged that 2007 "is generally regarded as a pre-crisis year" and that 2012 "corresponds to the first year the Italian economy entered a second wave of downturn after the recovery peak reached in 2011."

4 https://www.istat.it/it/files/2018/12/C14.pdf

5 Eight in both 2007 and 2012.

6 As expected, means are slightly higher than 1 and become closer to such a value as the distance considered increases: a value of precisely 1 would be obtained by an industry whose plants were distributed in a pattern exactly mimicking the general distribution of every economic activity within the entire territory analyzed; since most industries tend to show some degree of agglomeration and since the distance ranges we selected are far shorter than a radius that could include the whole island, our values are easily explained.

7 Since a significance test would require a high number of simulations to verify the significance of results for each industry, and this would be very difficult especially for M_A, we applied frequency weights to correlations so that each industry counts as much as his size in terms of number of plants. We believe that such approach follows the same logic of simulations: these would discard smaller industries as providing results based on chance ("re-shuffling" their plants over the whole set of locations would almost certainly provide very different results than their actual locations do) and maintain larger industries where such reshuffling would involve so many plants that most scenarios would provide very similar results to the real world. Perhaps less binarily, weighted correlations are more influenced by large industries and far less so by very small industries, without discarding them entirely.

References

Aleksandrova, E., K. Behrens, and M. Kuznetsova. (2020). "Manufacturing (co) Agglomeration in a Transition Country: Evidence from Russia." *Journal of Regional Science* 60(1), 88–128.

Almeida, E. T., R. M. S. Neto, and R. M. Rocha. 2020. Manufacturing Location Patterns in Brazil. PhD Dissertation in Economics. PIMES, Federal University of Pernambuco – UFPE.

Arbia, G. (2001). "Modelling the Geography of Economic Activities on a Continuous Space." *Papers in Regional Science* 80(4), 411–24.

Arita, T., and McCann, P. (2000). "Industrialalliances and firm location behaviour: some evidence from the US semiconductorindustry." *Applied Economics* 32(11), 1391–403.

Baddeley, A. J., J. Møller, and R. Waagepetersen. (2000). "Non-and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns." *Statistica Neerlandica* 54(3), 329–50.

Barlet, M., A. Briant, and L. Crusson. (2013). "Location Patterns of Service Industries in France: A Distance-Based Approach." *Regional Science and Urban Economics* 43(2), 338–51.

Becattini, G., F. Pyke, and W. Sengenberger, eds. (1990). *Industrial Districts and Inter-Firm Co-Operation in Italy*. Geneva: International Institute for Labour Studies.

Behrens, K., and T. Bougna. (2015). "An Anatomy of the Geographical Concentration of Canadian Manufacturing Industries." *Regional Science and Urban Economics* 51, 47–69.

Bonneu, F., and C. Thomas-Agnan. (2015). "Measuring and Testing Spatial Mass Concentration with Micro-Geographic Data." *Spatial Economic Analysis* 10(3), 289–316.

Brakman, S., H. Garretsen, and Z. Zhao. (2017). "Spatial Concentration of Manufacturing Firms in China." *Papers in Regional Science* 96, S179–S205.

Cainelli, G., R. Ganau, and Y. Jiang. (2020). "Detecting Space–Time Agglomeration Processes over the Great Recession Using Firm-Level Micro-Geographic Data." *Journal of Geographical Systems* 22(4), 419–45.

Catalini, C., C. Fons-Rosen, and P. Gaulé. (2020). “How Do Travel Costs Shape Collaboration?” *Management Science* 66(8), 3340–60.

De Dominicis, L., G. Arbia, and H. L. De Groot. (2013). “Concentration of Manufacturing and Service Sector Activities in Italy: Accounting for Spatial Dependence and Firm Size Distribution.” *Regional Studies* 47(3), 405–18.

Diggle, P. J., and A. G. Chetwynd. (1991). “Second-Order Analysis of Spatial Clustering for Inhomogeneous Populations.” *Biometrics* 47(3), 1155–63.

Dunford, M. (2006). “Industrial Districts, Magic Circles, and the Restructuring of the Italian Textiles and Clothing Chain.” *Economic Geography* 82(1), 27–59.

Duranton, G., and H. G. Overman. (2005). “Testing for Localization Using Micro-Geographic Data.” *The Review of Economic Studies* 72(4), 1077–106.

Ellison, G., and E. L. Glaeser. (1997). “Geographic Concentration in US Manufacturing Industries: A Dartboard Approach.” *Journal of Political Economy* 105(5), 889–927.

Feser, E. J., and S. H. Sweeney. (2002). “Theory, Methods and a Cross-Metropolitan Comparison of Business Clustering.” *Industrial Location Economics* 1, 222.

Floch, J. M., E. Marcon, and F. Puech. (2018). “Spatial Distribution of Points.” In *Handbook of Spatial Analysis: Theory and Application with R*, 77–111, edited by V. Loonis and M. P. Bellefon. Montrouge, France: Institut National de la Statistique et des Etudes Economiques.

Fratesi, U. (2008). “Issues in the Measurement of Localisation.” *Environment and Planning A* 40, 733–58.

Guillain, R., and J. Le Gallo. (2010). “Agglomeration and Dispersion of Economic Activities in and around Paris: An Exploratory Spatial Data Analysis.” *Environment and Planning B: Planning and Design* 37(6), 961–81.

Jensen, P., and J. Michel. (2011). “Measuring Spatial Dispersion: Exact Results on the Variance of Random Spatial Distributions.” *The Annals of Regional Science* 47(1), 81–110.

Koh, H. J., and N. Riedel. (2014). “Assessing the Localization Pattern of German Manufacturing and Service Industries: A Distance-Based Approach.” *Regional Studies* 48(5), 823–43.

Kolko, J. (2010). “Urbanization, Agglomeration, and Coagglomeration of Service Industries.” In *The Economics of Agglomeration*, 151–80. Cambridge: National Bureau of Economic Research.

Kopczewska, K. (2018). “Cluster-Based Measures of Regional Concentration. Critical Overview.” *Spatial Statistics* 27, 31–57.

Marcon, E., and F. Puech. (2003). “Evaluating the Geographic Concentration of Industries Using Distance-Based Methods.” *Journal of economic geography* 3(4), 409–28.

Marcon, E., and F. Puech. (2010). “Measures of the Geographic Concentration of Industries: Improving Distance-Based Methods.” *Journal of Economic Geography* 10(5), 745–62.

Marcon, É., and Puech, F. (2014). “Mesures de la concentration spatiale en espace continu: théorie etapplications.” *Économie et Statistique* 474(1), 105–31.

Marcon, E., and F. Puech. (2017). “A Typology of Distance-Based Measures of Spatial Concentration.” *Regional Science and Urban Economics* 62, 56–67.

Méndez-Ortega, C., and J. M. Arauzo-Carod. (2019). “Locating Software, Video Game, and Editing Electronics Firms: Using Microgeographic Data to Study Barcelona.” *Journal of Urban Technology* 26(3), 81–109.

Moreno-Monroy, A. I., and G. A. G. Garcìa-Cruz. (2016). “Intra-Metropolitan Agglomeration of Formal and Informal Manufacturing Activity: Evidence from Cali, Colombia.” *Tijdschrift voor economische en sociale geografie* 107(4), 389–406.

Nakajima, K., Y. U. Saito, and I. Uesugi. (2012). “Measuring Economic Localization: Evidence from Japanese Firm-Level Data.” *Journal of the Japanese and International Economies* 26(2), 201–20.

Openshaw, S., and P. Taylor. (1979). “A Million or So Correlation Coefficients: Three Experiments on the Modifiable Unit Area Problem.” In *Statistical Applications in the Spatial Sciences*, edited by N. Wrigley. 127–44. London: Pion.

Ripley, B. D. (1976). “The Second-Order Analysis of Stationary Point Processes.” *Journal of Applied Probability* 13(2), 255–66.

Ripley, B. D. (1977). “Modelling spatial patterns.” *Journal of the Royal Statistical Society: Series B (Methodological)* 39(2), 172–92.

Rosenthal, S. S., and W. C. Strange. (2001). “The Determinants of Agglomeration.” *Journal of Urban Economics* 50(2), 191–229.

Saxenian, A. (1996). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128, with a New Preface by the Author*. Cambridge: Harvard University Press.

Scholl, T., and T. Brenner. (2015). “Optimizing Distance-Based Methods for Large Data Sets.” *Journal of Geographical Systems* 17, 333–51.

Sorenson, O., and T. E. Stuart. (2001). “Syndication Networks and the Spatial Distribution of Venture Capital Investments.” *American Journal of Sociology* 106(6), 1546–88.

Storper, M., and A. J. Venables. (2004). “Buzz: Face-to-Face Contact and the Urban Economy.” *Journal of Economic Geography* 4(4), 351–70.

Sweeney, S., and M. Gómez-Antonio. (2016). “Localization and Industry Clustering Econometrics: An Assessment of Gibbs Models for Spatial Point Processes.” *Journal of Regional Science* 56(2), 257–87.

Sweeney, S. H., and E. J. Feser. (1998). “Plant Size and Clustering of Manufacturing Activity.” *Geographical Analysis* 30(1), 45–64.

Sweeney, S. H., and E. J. Feser. (2004). “Business Location and Spatial Externalities: Tying Concepts to Measures.” In *Spatially Integrated Social Science*, 239–62, edited by F. Michael Goodchild and D. G. Janelle. New york: Oxford University Press.

Sweeney, S. H., and K. J. Konty. (2005). “Robust Point-Pattern Inference from Spatially Censored Data.” *Environment and Planning A* 37(1), 141–59.

Tidu, A. (2021). Economic agglomeration in Italy before and after the Great Recession. PhD *Dissertation*. University of Cagliari.

Van Ham, M., P. Hooimeijer, and C. H. Mulder. (2001). “Urban Form and Job Access: Disparate Realities in the Randstad.” *Tijdschrift voor economische en sociale geografie* 92(2), 231–46.

Zhang, X., J. Yao, K. Sila-Nowicka, and C. Song. (2021). “Geographic Concentration of Industries in Jiangsu, China: A Spatial Point Pattern Analysis Using Micro-Geographic Data.” *The Annals of Regional Science* 66(2), 439–61.