European Regional Science Association

SEKCJA POLSKA

WORKING PAPERS

VOLUME 2015, Issue 1

Please note that **ERSA Polish Section Working Papers** have been presented as part of the working process of ERSA Polish Section Members. They are intended to make results of ERSA Polish Section research available to other economists, geographers, planners as well as other researchers so as to encourage discussion and suggestions for revision before its final publication. The works have different level of completion. Some of the studies are in preliminary form, however some of the them might have already been submitted in either this or earlier versions for peer reviews. We highly encourage you to comment upon the presented studies. Your feedback is more than appreciated.
Should you wish to cite any of the works please use the format: *Authors, Title, ERSA Polish Section Working Papers, Volume No., Issue No.* and, as the copyright remains with the authors of the paper, please provide the corresponding author with information of the used citation.

Katarzyna Kopczewska*, Paweł Churski**, Artur Ochojski***, Adam Polko***

** University of Warsaw, Faculty of Economic Sciences*
*** Adam Mickiewicz University in Poznan, Institute of Socio-Economic Geography and Spatial Management*
**** University of Economics in Katowice, Faculty of Economics*

Corresponding Author: Katarzyna Kopczewska
University of Warsaw, Faculty of Economic Sciences
kkopczewska@wne.uw.edu.pl

SPAG: Index of spatial agglomeration[♠]

Abstract: Paper develops a measure of spatial agglomeration of economic activity based on geo-localizations of firms. Proposed here Spatial Agglomeration Index (SPAG) includes the effects of location, distance between firms and overlapping impact resulting from the size and number of companies in given sector. SPAG builds a new class of measures of spatial density of economic activity inside the region, basing on geometrical representation of firms with circles, without referring to often used Ripley's K function. SPAG detects different spatial distributions of economic activity, including clusters. We provide also the Monte Carlo significance test of SPAG, based on theoretical distribution for spatially uniform locations of business.

JEL Code: R12, R32, C43, D30

Key words: spatial agglomeration, density of economic activity, circle packing, distance-based index

1. Introduction

Both theory and practice of economics notices that spatial distributions of economic activity are far from uniform. Agglomeration of firms inside regions is followed by growth in productivity, what Krugman (1991) incorporated in New Economic Geography approach.

Measurement of density of economic activity is still lagging behind the implications of this phenomena. Cluster-based indices (Marcon & Puech, 2009) of spatial concentration and specialization (i.e. Gini, Ellison-Glaeser (1997), Location Quotient and its variations) divide the territories (i.e. countries) into arbitrary (administrative) regions, where data on economic activity are aggregated. Measures, which are based on two-dimensional matrix of employment or any economic activity in discrete space, does not comply with the criteria of 'good' specialization index (Duranton & Overman, 2005). They mainly fail with Modified Areal Unit Problem (Arbia 2001a; Marcon & Puech, 2009; Morphet, 1997) as the value depends on shape and size of territorial units, but also they do not see the heterogeneity inside the region, as they

[♠] The paper is financed by the Polish National Science Centre www.ncn.gov.pl as the research project in OPUS 6 call [grant number UMO-2013/11/B/HS4/01098].
Acknowledgements: We would like to thank Roberta Capello, Ugo Fratesi, Daniel Griffith and Grzegorz Gorzelak for their valuable comments.

compare regions and sectors between themselves as homogenous units. Improved cluster-based indices, including spatial structure expressed usually by spatial weighs matrix W (Arbia, 2001b; Bickenbach & Bode, 2008; Arbia & Piras, 2009; Guillain & Le Gallo 2010; Guimaraes, Figueiredo & Woodward, 2011; Carlei & Nuccio, 2014; Sohn, 2014), are in the middle-way as they add information on spatial autocorrelation, but keep the minorities of a-spatial indices. The group of emerging distance-based measures (Marcon & Puech, 2003, 2009; Duranton & Overman, 2005, 2008), using point geo-localizations of firms in continuous space, and exploiting concept of K Ripley's function and its modifications, solves the MAUP, but presents result as (chart of) function, not as a single index, what is highly unattractive for policy makers and practitioners. All those measures still do not give a synthetic answer, what is the density of economic activity inside given region, including agglomeration effects, area of territory, size and sectors of firms. This calls for new measure which can support researchers with good and reliable information.

Consequently, in this paper we develop distance-based geometric model of spatial agglomeration for synthetic evaluation of density of economic activity inside the region. This methods are underdeveloped in the literature, and to our knowledge we can refer to few existing papers (Marcon & Puech, 2003, 2009; Duranton & Overman, 2005, 2008; Arbia et al., 2010; Mori & Smith, 2014; Lang, Macon & Puech, 2014).

Paper is organized as follows. In section 2 we provide the an overview of distance-based measures. In section 3 we present developed SPAG model intuitively with possible interpretations. In section 4 we derive significance test. Finally we present simulations of the SPAG for different point-patterns.

2. Overview of existing indicators – what is missing?

Measurement issues of specialization, concentration and agglomeration with regard to **economic activity** – i.e. number of economic units, employment, output or value added, appear in many studies and are well developed and described (i.e. Franceschi, Mussoni & Pelloni, 2009). Measures of these concepts form the main two groups: **cluster-based methods** and **distance-based methods**. For the group of cluster-based methods, where the most often used are Gini, Ellison-Gleaser, Location Quotient etc. (see Table 2), the starting point of analysis is the two-dimensional matrix by regions and sectors with data on employment or any economic activity in discrete space delimited with arbitrary borders[1] (see Table 1). Distance-based methods, still emerging in literature, use the individual point data for single business units and mainly apply the Ripley's K function to detect density of spatial distribution of economic activity.

Following Aiginger (1999) the **specialization** measures the share of given industry in regional economy, and as the reference only the other industries in given region are taken. **Concentration** is to capture share of given industry in given region in total activity of this industry in other locations. Thus as reference it assumes the employment / value added of examined industry in all analyzed regions (see Table 1). Finally, **agglomeration** is the coverage of region with all industries, geographic concentration of firms, the measure of density of economic activity and its even or agglomerated distribution over space.

[1] Duranton and Overman (2005, p.1078) call this phenomena "*transforming dots on a map into units in boxes*". They note its main advantage – simplification of calculations, but also see the problem of losing the significant part of information, emerging aggregation problems and recalculation of index when changing the spatial scale.

Table 1: Classical approach to measuring concentration and specialization

	Territory A	Territory B	Territory C	Territory D	...	Total
Industry A	emp(*ij*)	emp(*ij*)	emp(*ij*)	emp(*ij*)		Σemp(indA)
Industry B	emp(*ij*)	emp(*ij*)	emp(*ij*)	emp(*ij*)		Σemp(indB)
Industry C	emp(*ij*)	emp(*ij*)	emp(*ij*)	emp(*ij*)		Σemp(indC)
Industry D	emp(*ij*)	emp(*ij*)	emp(*ij*)	emp(*ij*)		Σemp(indD)
...						...
Total	Σemp(terrA)	Σemp(terrB)	Σemp(terrC)	Σemp(terrD)	...	ΣΣemp

Concentration

Specialization

Source: own synthesis

Measures based on regionally aggregated values of economic activity are attractive for policy makers, as the single value index is easy in interpretation. Also data availability support those methods. Existing coefficients trying to measure agglomeration effects across industries, time and space are based on the concept of seeing local area (region) against national area (country). Thus they measure the density and spatial distribution between regions, understood as separated parts of bigger (national) territory. Because of data availability after aggregation, cluster-based methods cannot look inside the region. Thus they are measuring uniformity of distribution of economic activity among sectors and regions, and are indicating for how much given sector or region is over- or under-represented by given activity in comparison to others. In those methods space is understood as the geographical relations between regions, for which internal spatial distribution of economic activity is not being considered. That is why cluster-based methods are the relative ones. Those indicators are *de facto* the a-spatial models on local level as the internal spatial distribution is neglected. Most of them (i.e. Gini index, Theil's entropy, coefficient of variation) is not sensitive to changes in spatial patterns (so called permutionally invariant). Development works to the classical indicators are in progress, to improve statistical inference and testing (Tian, 2013). Other stream is to embed spatial structure to enable measurement of spatial patterns of agglomeration. Also the level of sectorial aggregation is being examined (Fratesi, 2008).

Features of good measures were already defined. Following Duranton & Overman (2005) good measures of spatial activity from informational side should be comparable across industries, control for spatial agglomeration effects as well as industrial concentration. From technical side, should be resistant to geographical scale, administrative units aggregation and MAUP, as well as provide significance test. Fratesi (2008) in measuring localization requires two corrections: for overall agglomeration of activity and for sectoral structure. Guillain & Le Gallo (2010) add to that it should be feasible due to data availability and confidentiality restrictions. Kominers (2007) adds the requirement of being justified by suitable model. Palan (2010) adds few more: axiom of anonymity, axiom of progressive transfer, bounds and decomposability. However, these criteria are mainly to improve the cluster-based methods. The answer to this was the significance test applied to traditional measures as bootstrap test for LQ (Tian, 2013) or adding spatial components to them (Guillain & Le Gallo, 2010; Arbia & Piras, 2009; Carlei & Nuccio, 2014; Guimaraes, Figueiredo & Woodward, 2011; Bickenbach & Bode, 2008; Sohn, 2014).

Table 2 reports most of the existing indicators of concentration, specialization and agglomeration with regard to their features. Basic models anchored on two-dimensional matrix of economic activity aggregated inside given regions for given sector are being supplemented with information on size of companies, as spatial information as area of territory, distance between regions and spatial neighborhood structure.

Table 2: Properties of cluster-based methods

Index	n x k matrix of economic activity	Neighbour -hood relations - W spatial weights matrix	Distance between regions	Area of territo-ry	Size of compa-nies
Gini index, Location Quotient (Hoover-Balassa coefficient), Theil's Entropy, Shannon Entropy, Ogive index, Diversification index, Krugman Specialization index, Inequality index, Index of agglomeration, Index of specialization Hallet (2000), Entropy index of overall localization (Cutrini, 2009), Entropy measure (Bruelhart & Traeger, 2005)	V	---	---	--	---
Herfindahl index, absolute and relative Theil index (Bickenbach, Bode & Krieger-Bode, 2012), Relative Diversity Index (Duranton & Puga, 2001)	---	---	---	----	V
Ellison & Glaeser index, excessive concentration (Ellison & Glaeser, 1997), Herfindahl index, Isard index, Maurel & Sedillot, 1999	V	---	---	---	V
Clustering index (Bergstrand, 1985)	V	---	V	---	---
Concentration index (Spiezia, 2002), Regional Industrial Mass and Regional Industrial Concentration (Franceschi, Mussoni & Pelloni, 2009)	V	---	---	V	---
Gini with ESDA (Guillain & Le Gallo 2010), Using Gini together with Moran's I and Getis-Ord (Arbia, 2001b), Spatial Concentration Measure (Arbia & Piras, 2009), Relative Industrial Relevance (Carlei & Nuccio, 2014), inflation factor as correction of other measures (Guimaraes, Figueiredo & Woodward, 2011), Disproportionality Measures (Bickenbach & Bode, 2008), Spatial distribution (Sohn, 2014)	V	V	V	---	---

Source: own synthesis

As Guillain & Le Gallo (2010) state, regional science is now targeted on measurement and comparison of the spatial distributions of economic activity. Most of technical drawbacks of cluster-based measures, regarding the problems of border delimitation, using all geographical scales simultaniously and the reference area when measuring agglomeration, indicated by Guillain & Le Gallo (2010), are being automatically solved by applying individual data based measures, which are not aggregated over arbitrarily selected space. Necessity of using continuous-space models based on micro data was raised by Arbia (2001a). This emerging class of measures, the distance-based methods (see Table 3), allow for the look inside the region, also without referring to other regions. Density and spatial structure can be measured inside single region and if needed compared with other regions, what makes those measures absolute.

Typology of existing distance-based methods (Marcon & Puech, 2014) proves that many of them (Marcon & Puech, 2003, 2009; Duranton & Overman, 2005, 2008; Arbia et al., 2010) are anchored on the same basic methodology of point pattern and coming from K-density Ripley's function counting the number of neighbors for given distance, standardizing this value with space or other number of neighbors, averaging and normalizing the result with regard to edge effect. Those methods try to gauge the spatial concentration and determine the spatial structure of economic activity, mainly to compare the deviations of sectorial spatial distribution from the aggregated general economic activity distribution. The result is two-dimensional plot, where for distance (x), K-function (y) is being plotted. Duranton & Overman (2005, 2008) test for co-location of linked industries and subgroups (exit and entry firms, FDI and home firms, big and small firms) location. For analyzed data they construct bootstrapped confidence intervals. Arbia et al. (2010) and Kang (2010) develop the space-time K function to model the temporal dynamics of spatial pattern and the result is three-dimensional: for time (x) and distance (y), the K surface (z) is being plotted. Analytical expression for variance of two-dimensional indices by Marcon & Puech (2009) and Duranton & Overman (2005) was provided by Jensen & Michel (2011).

There are few other trials. Do & Campante (2009) propose Gravity-based Centered Index of Spatial Concentration (G-CISC) to measure the concentration around given point, instead of concentration over give area. Designed as index universal over any space, in 2D is represented by "*decreasing log-linear functions of the distance between individual observations and the center*". For gridded data, concentration of variable is calculated and standardized with maximum distance across / inside country. Lang, Marcon & Puech (2014) propose the relative density *m* function, which completes the typology of distance-based function (Marcon & Puech, 2014). Based on kernel estimation, it goes beyond the class measures based on Ripley's K. It can detect local spatial structures of point data in relative way and confidence intervals can be simulated.

Both distance-based and cluster-based mixed approach is presented by Mori & Smith (2014). They use very local areal data to determine economic area (usable area for firms), distances between those spatial units, as well as individual data on firms' location. Using probabilistic methods they model two component indicator of industrial spatial agglomeration: global extent (GE) and local density (LD). This approach can define scale and degree of industrial agglomeration and detect the spatial clusters. GE is thus a measure of spatial spread of clusters in areal approach between regions, and LD is internal measure of area covered by selected industry in region, where as the reference is assumed the random distribution of industry firms' location. Clusters are defined as overrepresentation of firms in given industry in given local unit. Indicators by Mori & Smith (2014) even though are computationally advanced, fail MAUP, as the basis of calculation are local administrative units. This is also a relative measure as it compares local spatial distributions with the aggregated and random ones.

Distance-based measures fulfill most of the criteria specified by Duranton & Overman (2005). They are comparable across industries, by construction try to control for spatial agglomeration effects as well as industrial concentration. Using micro point data rejects MAUP and developments provide the significance test and confidence intervals. Data availability becomes less important issue. Even though, distance-based measures are not perfect tools applicable in precise measurement of economic activity inside the region. That is why new set of supplementary criteria of good index should be defined, what we propose below.

In reference to point data there exist technical conditions of index, which are based on mathematical properties of measures. Do & Campante (2009) for grid-based data give basic and refinement anxioms of decomposability and monotonicity to be satisfied by function being the index of spatial concentration. In those indicators, main interest is on the density of economic activity and its spatial distribution over the territory. However set of substantive criteria for point-data indices must be defined.

First condition is that those measures should be sensitive to different spatial distributions. Marcon & Puech (2014) show that most of distance-based measures applies the same mechanism of Ripley's K functions. As proved on Plot 2, Ripley's K function is sensitive to size sample and poorly distinguishes between spatial uniform and border-dispersed distributions of points over given territory.

Second condition is that indicator should include area of territory, as well as size and importance of companies. Firms are not homogenous in terms of employment or turnover, and few big companies on small area means not the same as few small firms in big region. Thus indicator should combine and relate the both values. Most of distance-based models neglect this issue, only Duranton & Overman (2005) include size of firms as characteristics of subsample. Fratesi (2008) notes that small and big firms should not be carelessly permutated over space because of different requirements and conditions of its existence.

Third condition is the easy and unequivocal interpretation. Continuous-space indicators give the result as a line graph of function, some with confidence interval, what is difficult in operationalization and decreases its attractiveness and usefulness in applied policy. Distance-based measures will gain understanding, when their result will be given as single measure, scaled and comparable between sectors, areas, time, firms density etc.

Fourth condition is on applying absolute measures. As Marcon & Puech (2009) state: "*Relative measures detect whether each industry is overrepresented or underrepresented with respect to a baseline distribution, for example, the overall location pattern of industries*". Most of the cluster-based measures are relative and only comparable between each other. Relative measures requires other regions to conclude on the studies area. Although is allows for almost automatic ranking and easy comparisons, it bears the problem of extreme values or activity of other sectors in other regions, which strongly impacts the results. Idea of constructing absolute measure meets the postulate of independent analysis of region, without reference to neighbors and others. This kind of measures are universal, as the economic interpretation results from comparison with well-defined reference point. Distance-based measures, because of construction, mainly are absolute.

Fifth condition is on the information obtained from the results. Cluster-based measures conclude on specialization, concentration and agglomeration, even if the names of concepts are used in chaotic manner (Churski et. al, 2015). Following the Marcon & Puech (2009) this is mainly measurement of over- and under-representation sector in region in comparison to overall location pattern. Distance-based measure ask other questions. They are designed to detect concentration and dispersion over space. With Ripley's K they detect the distance, at which this concentration or dispersion exists.

This review proves that there is still need to build a new measure of spatial agglomeration, which would follow all mentioned conditions as there is no measure giving a synthetic answer, what is the density of economic activity inside given region, including agglomeration effects, area of territory, size and sectors of firms, and locations around core cities. In the next section we develop distance-based geometric model of spatial agglomeration for synthetic evaluation of density of economic activity inside the region, which fulfills the above mentioned criteria.

Table 3: Properties of distance-based methods

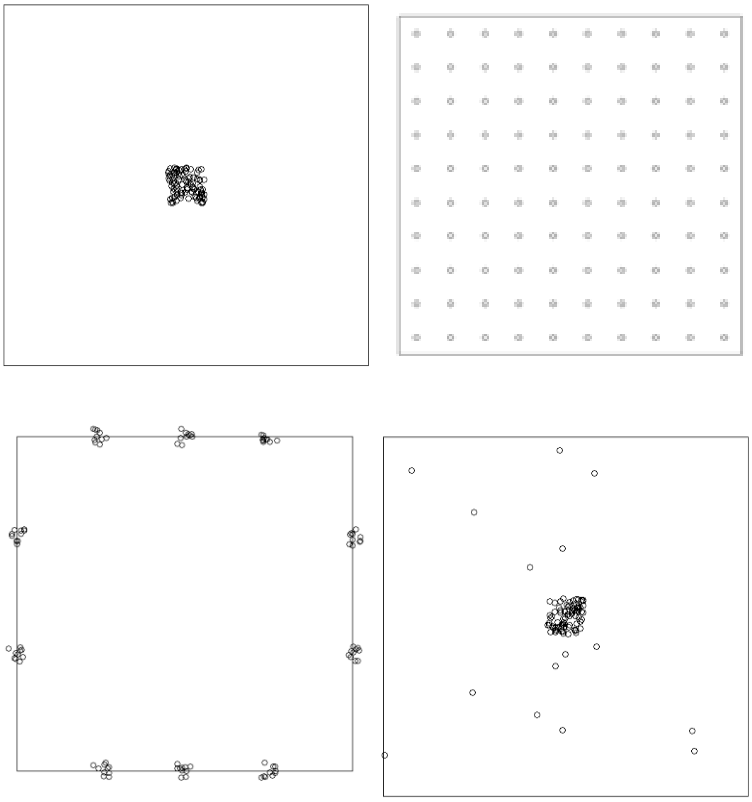
	Based on Ripley's K function	Indivi-dual firms given as point	Dense grid or very small spatial units	Area of terri-tory	Size of compa-nies	Sectorial approach	Easy in interpret-tation point result	Signifi-cance test
Marcon & Puech, 2003	V	V	---	---	---	V	---	V
Duranton & Overman, 2005	V	V	---	---	V	V	---	V
Do & Campante, 2009	---	---	V	V	---	---	---	---
Duranton & Overman, 2008	V	V	---	---	V	V	---	V
Marcon & Puech, 2009	V	V	---	---	V	V	---	V
Space-time K function (Arbia, Espa, Giuliani & Mazzitelli, 2010)	V	V	---	---	---	---	---	V
global extent (GE) and local density (LD) (Mori & Smith, 2014)	---	V	V	V	---	V	---	V
relative density *m* function (Lang, Macon, Puech, 2014)	---	V	---	---	---	---	---	V
SPAG (proposed here)	---	V	---	V	V	V	V	V

Source: own synthesis

3. Why we need new index?

Spatial literature notices the importance of space and spatial distribution of activity for socio-economic patterns of development. This interest in **density of economic activity** is visible both in theoretical literature explaining the mechanisms as well as in methodological papers when this issue is to be measured. However, it still hard to measure this density, mainly because of the lack of good measure. Even if the patterns of spatial distribution of employment are as on Plot 1, popular measures of specialization and agglomeration give the same results, as the number of firms on the plot is the same.

Plot 1: Different spatial patterns of economic activity with different volume of business: a) agglomeration, b) uniform spatial distribution, c) border-dispersed distribution, d) agglomeration with few units on peripheries

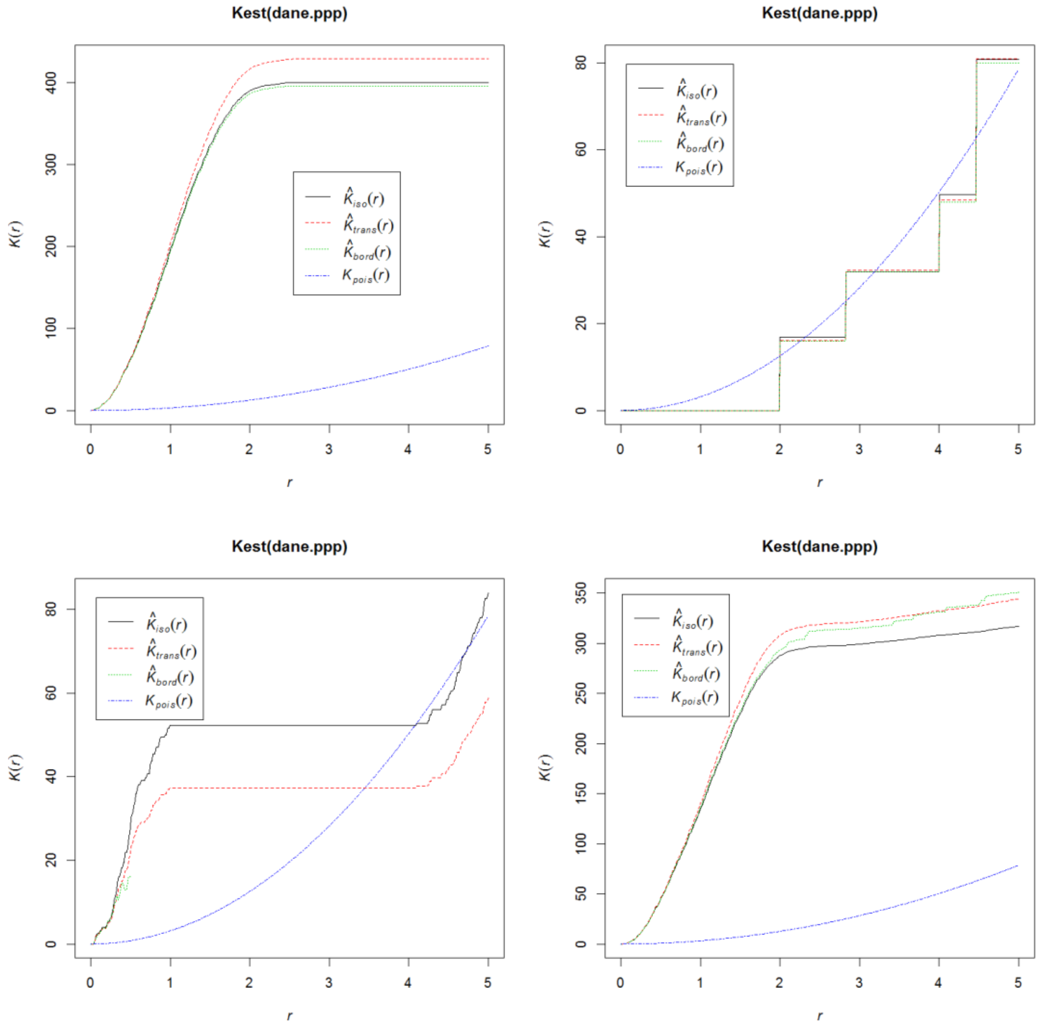


Source: own simulation

If the localization of firms of any sectorial composition (the same in call cases) is as on the Plot 1 and all panels – “regions” compound one “country”, then because the number of firms on each panel is the same, all cluster-based measures, with and without spatial component will be the same. Even if the spatial distribution is different inside regions, the aggregation of the data (i.e. on employment) for administratively delimited regions smooths this difference.

Literature on distance-based indices, exploiting Ripley’s K function, is trying to look inside the region to assess the spatial distribution. Expectation is that when firms are clustered, then they have more neighbors in given radius than if they are randomly or uniformly distributed. Results on the Plot 2 prove that this approach gives very similar results in case of diametrically different spatial distributions. Presented Ripley’s K functions (Plot 2) for spatial distributions as on Plot 1 poorly distinguish the underlying spatial patterns. Secondly, this approach does not fulfill the condition of easy in interpretation result, as the output is mainly graphical and functional, not numeric and index.

Plot 2: Ripley's function for different spatial patterns: a) agglomeration, b) uniform spatial distribution, c) border-dispersed distribution, d) agglomeration with few units on peripheries



Source: own simulation

This paper is to construct new measure of density of economic activity based on geo-location of firms to test hypothesis of independence of localization, which would be sensitive for spatial patterns, size and volume of economic activity and easy in interpretation.

4. Construction of the index

Below we propose the new approach to measure the density of economic activity over territory. This measure provides with information about inside of region, what is unavailable in case of cluster-based measures. As the other distance-based measures, it starts with individual geo-located firms so might be applicable over territorial division and the problem of zoning (MAUP) is not present here. The first novelty is that we do not use the concept of Ripley's K function neither kernel estimation. Instead, we propose the index of spatial agglomeration (SPAG) based on geometrical representation of firms with circles. It evaluates the regions' coverage with the economic activity. By construction, this is an absolute measure, with reference value of uniform spatial distribution equal 1. SPAG index incorporates the information on area of territory as well as the size and sectors of companies. The second novelty is that the result is a decomposable point-value index, and also geometric (graphical) representation is available. All three multiplier components are scaled around 1 and have their economic interpretation. Thus the application of SPAG by policy makers, as well as the comparability between regions, sectors, over time and with uniform spatial distribution is easy and powerful. Finally we provide the confidence interval to test the empirical values of SPAG.

SPAG index is designed to:

a) measure the degree of spatial agglomeration (spatial density of activity, geographical concentration, clustering) and distance from uniformity of economic activity. Reference value SPAG=1 is for the same size companies distributed evenly over the territory. Values of SPAG<1 reveal patterns of clustering, with extreme value SPAG~0 at one-point cluster. Values of SPAG>1 prove the existence of border-dispersed pattern and the mechanisms of repulsion.
b) supplement the traditional cluster-based measures by providing the coherent to *kxn* matrix of density of economic activity, the *kxn* matix of SPAG by regions and sectors. These results might be applied to correct for spatial dimension the traditional concentration and specialization measures.
c) compare regions, compare the same region over time, compare sectors inside region and between regions and track dynamics of agglomeration or repulsion

The starting point is the geo-location of *n* business units. By construction, index compares the empirical and theoretical distributions of circles representing firms. In empirical distribution, each point (x,y) of *n* firms' location is appointed by the circle, which area is proportional to employment $empl_i$ in the company.

$$\sum_{i=1}^{n} a_i(r_i) = \sum_{i=1}^{n} \pi \cdot r_i^2 = A \quad \text{and} \quad a_i \sim empl_i \quad \text{and } (x_i, y_i) \sim \text{empirical} \tag{1}$$

where a_i is the area and r_i is the radius of *i*-th circle representing *i*-th firm, $empl_i$ is an employment in *i*-th firm, A is the area of region and (x_i, y_i) are empirical geographical coordinates of *i*-th firm. Radii r_i of circles result from optimization, that sum of the a_i areas of *n* circles is equal to the area *A* of the region. Radii r_i of *n* circles might be continuous variable for precise data on employment or discrete for interval data. In case of interval data with *k* classes of employment in firms, the optimization is as follows:

$$\sum_{k=1}^{K} n_k \cdot \pi \cdot r_k^2 - A = 0 \tag{2}$$

When employment in larger firms is *d* times bigger than in the smallest ones (treated as reference size firm), than the optimization takes the form:

$$\sum_{k=1}^{K} n_k \cdot \pi \cdot d_k \cdot r_{base}^2 - A = 0 \tag{3}$$

where r_{base} is the radius of smallest circles representing the group of smallest firms. Radii of the circles create the *business impact zones*, which are automatically bigger in case of bigger firms. Fulfilling the optimization condition guarantees that radii of circles are not random, but are well-linked with both size of region and volume of economic activity. The interpretation of proportional to employment radii is as follow: the whole economic activity of region (employment) is projected on the area of region and each firm with its employment is participating in this total employment. The share of *i-th* circle area in total area is the same as share of *i-th* firm employment in total employment, thus circles represent firms in coherent way.

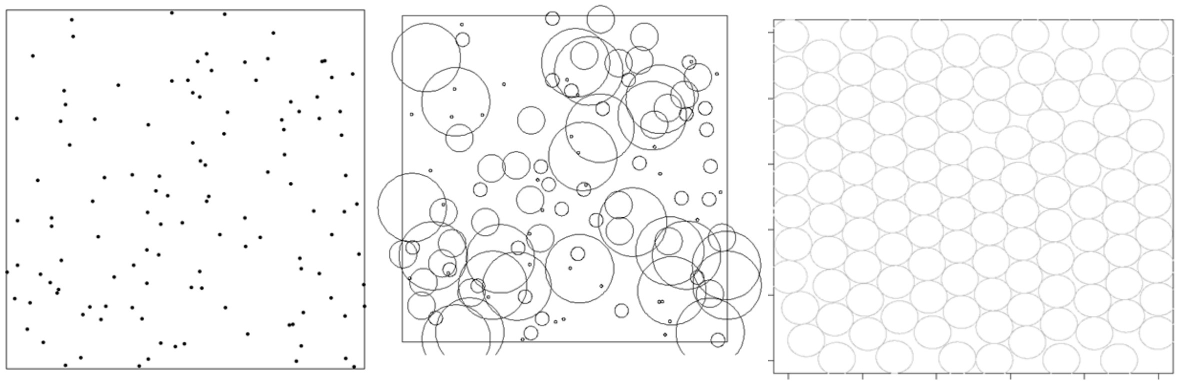
Setting circles in real business locations is to reflect the phenomena of spatial agglomeration or other spatial patterns.

In theoretical distribution one assumes the same number of n firms as in empirical distribution, but of equal size. Again, radius r_t of circles fulfills the condition, that sum of the areas of n circles is equal to the area of the region. Theoretical locations follow the spatially uniform distribution, with no agglomeration or dispersion, gridded in case of cuboid regions and resulting from circle packing in case of non-regular shape of region.

$$\sum_{i=1}^{n} a_i(r_t) = n \cdot \pi \cdot r_t^2 = A \quad \text{and} \quad a_i = \frac{A}{n} \quad \text{and} \quad (x_i, y_i) \sim \text{spatially uniform} \tag{4}$$

The above framework can be visualized with an example as below (Plot 3). Empirical locations of n firms (3a) are represented with circles of proportional to employment area (3b) and compared with theoretical spatially uniform distribution of equal size circles (3c).

Plot 3: Framework of SPAG *(n=118)*

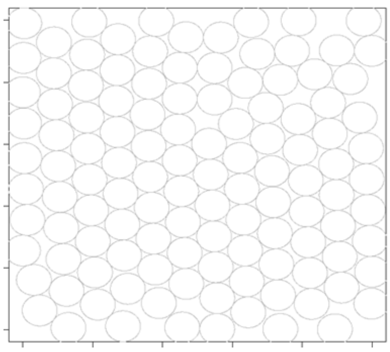


Source: own simulation in R with use of Bedward (2010) code

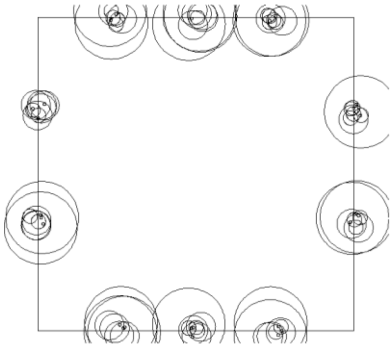
The above example is for random pattern of location of firms. There are also many other possible spatial distributions of economic activity (see Plot 4). As border examples one can assume two extreme cases: i) in the case of uniform spatial distribution circles will not overlap, what gives the maximum coverage with an economic activity and minimum agglomeration effect (Plot 4a), ii) in case of an extremely uneven coverage, all companies will be located at a single point and all zones cover a range, which will mean a minimum coverage with business and extreme spatial agglomeration (Plot 4f). Between these extremes there are several intermediate states, including iii) impact zones spatially separated (Plot 4b), iv) few not overlapping clusters of impact zones (Plot 4c), v) impact zones partly overlapping (Plot 4d) or clusters with weak dispersion (Plot 4e). In all cases there is different degree of coverage of territory with the business.

Plot 4: Extreme spatial distributions of impact zones

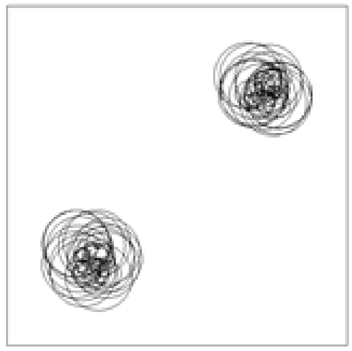
a)spatially uniform distribution



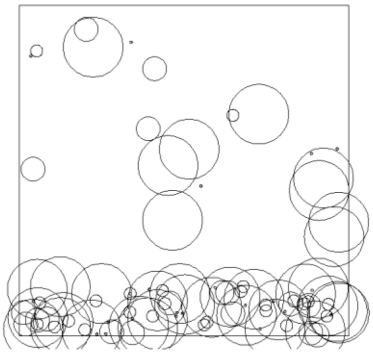
b) border-dispersed location



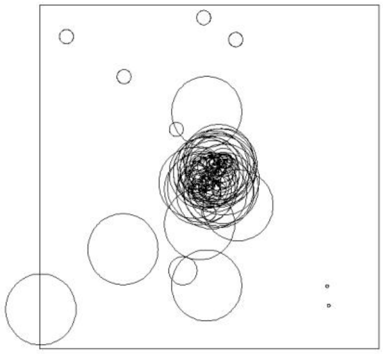
c) not overlapping clusters of impact zones



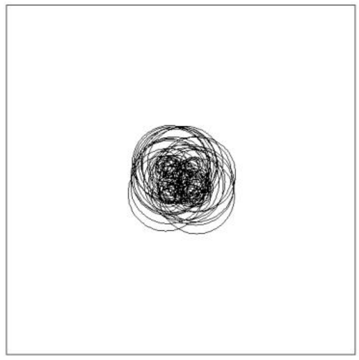
d) impact zones partly overlapping



e) Weak dispersion around cluster



f) Spatial concentration in one point



Source: own simulation

To reflect all possible localization scenarios, the construction of SPAG includes three elements:

a) coverage of territory by circles, to enable calculations of relative coverage, with selected sector in relation to all business units

b) average distance between locations, to cover the extreme effects of full concentration and border-dispersed points, as well to distinguish between non overlapping circles strongly dispersed and tightly located

c) the ratio of overlapping circle areas, to measure the degree of departure from spatially uniform (non-overlapping) distribution towards full concentration in single point.

Thus, the index of Spatial Agglomeration (SPAG) is the product of three components: index of coverage $I_{coverage}$, index of distance $I_{distance}$ and index of overlap $I_{overlap}$.

$$SPAG=I_{coverage}*I_{distance}*I_{overlap} \tag{5}$$

Multiplicative form of SPAG with equal weights of components allows for obtaining minimum level of SPAG (around 0) for extreme spatial concentration as well keeping the value around 1 in case of equal spatial distribution. It also strengthens the effects of spatial agglomeration as the $I_{distance}$ and $I_{overlap}$ react the same way, although they are designed to capture different spatial patterns. SPAG takes values from 0 to *r,* where *r* is the radius of the circle described on the geometrical figure representing the studied area. For full concentration of all units in one single point, value of SPAG=0, as the average distance = 0 between units. For fully uniform distribution over territory, with no overlapping of business impact zones, SPAG=1. SPAG>1 is typical for border-dispersed locations, with poor overlapping, as then the average distance is bigger than in spatially uniform distribution.

In the construction of SPAG and its components, the general rule is that empirical estimations are being compared with theoretical one from spatially uniform distribution. Thus the nominator of index is based on empirical spatial distribution of circles, and the denominator is the theoretical benchmark from uniform distribution of location and size of companies (double uniform over space and size). In empirical distribution one assumes circles with the area proportional to employment in the companies. Radii of circles are optimized to fulfill the condition that sum of the areas of circles is equal to the area of the region. This areas of circles create the business impact zones. In theoretical benchmark distribution, which is used in denominator, the uniformly agglomerated / dispersed location of circles of equal size is being assumed. Below all three components of SPAG are presented.

Index of coverage, $I_{coverage}$, is the coefficient of spatial coverage of studied area with circles representing firms. Value of this index is:

$$I_{coverage} = \frac{\sum P_i}{P_r} \tag{6}$$

where P_i is the area of selected circles representing firms, and P_r is the area of region. In case when all *n* out of *n* firms located on given area are being analyzed, $I_{coverage}$ should be 1. For sectorial analysis, when all *n* firms are treated as the reference point and only sectorial sample $k<n$ is being selected, $I_{coverage}$ is less than 1. $I_{coverage}$ by assumption cannot be more than 1, as the analysis cannot include more firms than located on the study area ($k>n$). This index is to distinguish between different number of firms under analysis. It defines the percentage magnitude of given sector for whole market.

Assumption for this index is about the size of circles. Benchmark distribution applied in denominator, is based on equal-sized circles uniformly distributed over bounding box (region). This can be achieved with the procedure of circle packing or grid. Thus, in the denominator benchmark distribution only a number of firms and total employment in region do matter and circles represent the average firm. On the contrary, empirical distribution used in the nominator, takes into account the size of companies (precise or in interval) and their real locations. Size of circles would be different, depending on size of company and total employment in region. In both distributions the condition is that radii of circles are optimized to fulfill the condition that sum of the areas of circles is equal to the area of the region.

One should also note one issue in visualization. Packing circles procedure assumes that nothing extends beyond bounding box, while in mapping the real location, this condition does not have to be fulfilled. In computational sense, areas of circles and regions are balanced, but in graphical sense this matching is not full.

Index of distance, $I_{distance}$, is the coefficient representing the average distance between locations. We compare average real distance of selected points to average theoretical distance of selected points. As the reference point we assume uniform distribution of circles over territory. Value of this index is:

$$I_{distance} = \frac{\sum_i \sum_j d_{ij}/k}{\sum_i \sum_j \widehat{d_{ij}}/k} \quad (7)$$

where $\sum_i \sum_j d_{ij}$ is the sum of distances between pairs of analyzed centroids, $\sum_i \sum_j \widehat{d_{ij}}$ is the sum of distances between pairs of centroids under uniform benchmark distribution, k is the number of firms selected to the analysis out of total number of n firms in region. This coefficient is designed to measure for how far on average are the firms from each other. In case they are concentrated in one point, the average distance will be 0, thus $I_{distance}$ coefficient will 0. If firms are uniformly distributed over territory, their empirical average distance should be the same as benchmark average distance, making $I_{distance} = 1$. This index might be >1 in situation of border-locations of k<n firms, when distant localization increases the average distance between points. Index of distance matters for SPAG especially when k<n. Then for non-overlapping case average distance may vary if points located in non-overlapping cluster or separately.

In the denominator of this index, the analyzed sample of k firms should be included. Benchmark uniform distribution might have thus twofold form, one in case of full coverage analysis when $k=n$, the second in case of not full coverage analysis when $k<n$. For full coverage analysis, the benchmark uniform distribution assumes that all circles of equal size are packed in bounding box. This distribution is the same as in case of denominator of index of coverage. For not fully coverage some other uniform distribution must be assumed. The best option seems to use grid division of region with k cells and locate there k equal-sized circles, what will distribute circles without full coverage uniformly over territory of bounding box[2]. It is also worth to note, that again for denominator benchmark distribution circles of equal size are assumed. In index of distance size of circle does not matter as the distances are measured between centroids of circles. Different size of circles could affect the grid division only and complicate the algorithm without value added. The equal-sized firms in denominator are thus comparable between index of coverage and index of distance.

In the nominator of the index of distance the same empirical distribution of firms as in index of coverage is being used. In this index, the average distance between centroids, instead of total area of circles as in index of coverage, is calculated. Even if the empirical distribution includes different-sized circles, this information is not reflected in the result.

Index of overlap, $I_{overlap}$, is the coefficient indicating for how much the impact zones of firms overlaps or to what degree firms are uniformly distributed over territory. As the reference point we assume uniform distribution of circles over territory. We calculate the union as the total area covered by circles, including the overlapping effects. Value of this index is:

$$I_{overlap} = \frac{P(\bigcup_i P_i)}{\sum P_i} \quad (8)$$

where $\sum P_i$ is the total area of circles representing firms and $P(\bigcup_i P_i)$ is the union of areas of circles and (1- $P(\bigcup_i P_i)$) is the area uncovered by circles. This coefficient measures for how much the territorial impact of firms is concentrated in narrow area with covering mutually business impact zones, or on the contrary to what degree this impact is being expanded over the whole territory.

[2] In spatial sampling, reaching the equal spatial coverage is possible when applying the minimization of the mean for the shortest distances (MMSD) in Euclidean sense. This is feasible when space is being discretized into grid (van Groenigen et al. 1999, Wang et al, 2012).

Denominator of this index is defined as in the index of distance, but the total area covered by circles instead of average distance is being calculated. For case of full cover the circle packing procedure is used and for not full cover the grid with equal-sized circles is applied. Distribution in the nominator is the same as in previous both indices and union, understood as total area covered with circles, is calculated.

Distributions of nominator and denominator in all three component indices is presented in the table below (see Table 4).

Table 4: Summary of distributions used in model

	Index of coverage	**Index of distance**	**Index of overlap**
Measure to be calculated in counter and nominator	Total area of selected *k* circles	Average distance between *k* circles	Area covered by selected *k* circles
Counter Empirical distribution of *k* circles	Circles of different size Radii of circles from optimization	Circles of different size Radii of circles from optimization	Circles of different size Radii of circles from optimization
Nominator Theoretical distribution under full coverage *k=n*	Real location Circles of equal size Radii of *n* circles from optimization Circle packing algorithm to locate circles	Real location Circles of equal size Radii of *n* circles from optimization Circle packing algorithm to locate circles Circles of equal size	Real location Circles of equal size Radii of *n* circles from optimization Circle packing algorithm to locate circles Circles of equal size
Nominator Theoretical distribution under not full coverage *k<n*	---	Selection of *k* out *n* circles from full coverage scenario Grid division of bounding box for *k* cells to locate *k* circles inside	Selection of *k* out *n* circles from full coverage scenario Grid division of bounding box for *k* cells to locate *k* circles inside

Source: Own summary

Interpretation and possible results from component indices are given in a table below (see Table 5).

Table 5: Possible results – potential values of indices

	Index of coverage	**Index of distance**	**Index of overlap**
I=0	For *k<n* no (*k=0*) circles selected – no firms of given sector in region For *k=n* Impossible result*	For *k<n* All circles in the same point The same location of all firms Extreme spatial agglomeration For *k=n* All circles in the same point The same location of all firms Extreme spatial agglomeration	For *k<n* Impossible result*** For *k=n* Impossible result***
0<I<1	For *k<n* Natural result when not all	For *k<n* Circles located closer than on	For *k<n* Overlapping of circles

	circles selected Only firms of given sector are analyzed in reference to full business activity Firms can be both smaller or bigger than on average For *k*=*n* Impossible result**	average Firms follow agglomeration pattern For *k*=*n* Circles located closer than on average Firms follow agglomeration pattern	Agglomeration of firms on small part of region For *k*=*n* Overlapping of circles Agglomeration of firms on small part of region
I=1	For *k*<*n* Area of circles bigger than average *k* firms of selected industry are relatively big and employ more than average For *k*=*n* Natural result when all firms selected	For *k*<*n* Uniform distribution of empirical sample Firms cover with their impact zone full territory For *k*=*n* Uniform distribution of empirical sample Firms cover with their impact zone full territory	For *k*<*n* Poor overlapping of circles Firms uniformly distributed over territory Poor agglomeration forces Overlapping possible when firms bigger than on average For *k*=*n* Poor overlapping of circles Firms uniformly distributed over territory Poor agglomeration forces Overlapping possible when firms bigger than on average
I>1	For *k*<*n* Area of circles bigger than average *k* firms of selected industry are relatively big and employ more than average For *k*=*n* Impossible result**	For *k*<*n* Circles located on the border of bounding box, with possible overlapping Empty center area of bounding box Firms escape from central location For *k*=*n* Circles located on the border of bounding box, with possible overlapping Empty center area of bounding box Firms escape from central location	For *k*<*n* Poor overlapping of circles Firms uniformly distributed over territory Poor agglomeration forces Overlapping possible when firms bigger than on average For *k*=*n* Poor overlapping of circles Firms uniformly distributed over territory Poor agglomeration forces Overlapping possible when firms bigger than on average

** by assumption we select all firms on given territory*
*** because of condition of radii length optimization*
**** even with one circle some area must be covered*
Source: Own summary

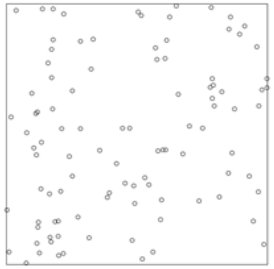
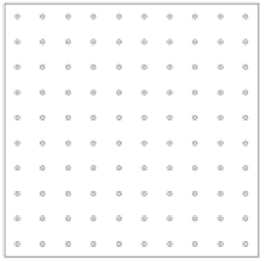
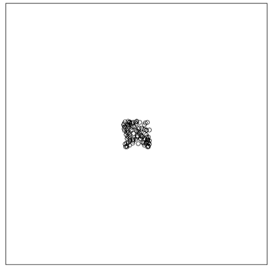
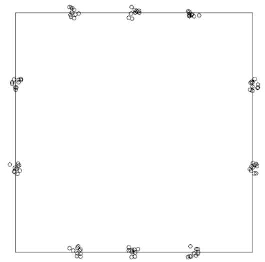
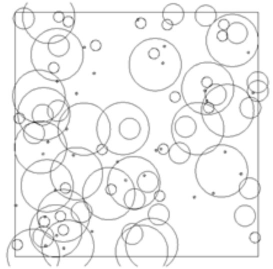
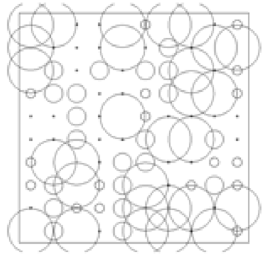
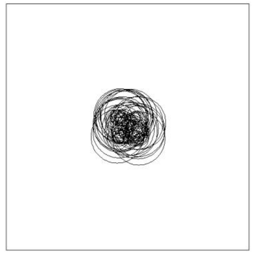
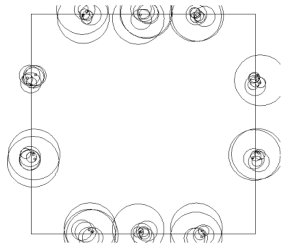
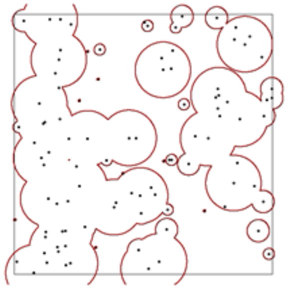
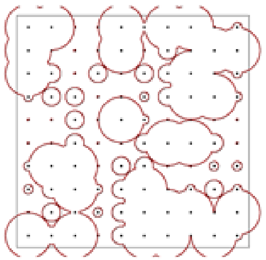
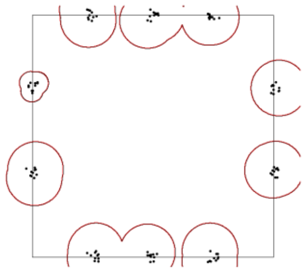
Table 6 presents simulation of SPAG for different spatial distributions of points. In all simulations theoretical data for n=100 firms located on the 10x10 square and four classes of companies' size with equal frequency distribution of size were taken. This simulation proves that SPAG distinguishes different spatial distributions of firms' locations. In extreme case of firms concentrated in one single point (with one agglomeration center and no units on peripheries) value of SPAG is low. The highest values of SPAG is for spatially uniform distribution of firms.

Interpretation of SPAG is multidirectional. All three components and its final value should be considered jointly. **Coverage** component sets the dimension of analysis. When all business units selected it takes naturally value of 1. But for selected industries, depending on research perspective, it can take value equal 1 (full coverage of sector) as well as less than one (share of industry). With $I_{coverage}=1$ business concentration analysis is limited to this industry only, and co-locations does not matter. Optimization of radius is performed for every industry, so circle sizes are not comparable between sectors. However this is autonomous analysis and SPAG can easily supplement the information from traditional cluster-based analysis, and it does not replicate information on sectorial concentration. With $I_{coverage}<1$, which reflects the share of industry in regional employment, radii of circles are the same for all sectors, with information on sectorial concentration already included. Thus in this case the reference point for comparative analytics is the whole regional economy, not the behavior of sector itself. **Distance** component is the measure of density of economic activity in space. This measure is always comparable, independent on sector selected and value of coverage. Low values of distance reflect spatial proximity of firms what appears for spatially concentrated locations of business units. Values of $I_{distance}$ close to 1 indicate that empirical points are distributed on average similarly to the theoretical gridded points, what suggests the equal saturation of territory with business. In case of regions with space impossible to develop (lakes, mountains etc.). $I_{distance}$ will be <1, what truly reflects the fact that part of territory is economically inactive. **Overlap** component supplements the information on distance. It includes information on companies' size and can indicate for how much big firms co-locate with small firms. In economies with small firms only, the overlap is automatically lower, existence of big firms interacts with other firms. It measures the agglomeration, for which overlapping of business zones is evident. Low values of $I_{overlap}$ prove the pattern of agglomeration, and on contrary values close to 1 reflect spatially uniform locations.

SPAG can also be used in analysis theoretical analysis of spatial concentration. Following Hooverian approach (Hoover, 1937), economies of scale appear when firms prefer to operate as big business because of increasing returns to scale. SPAG by inclusion of size of company in its construction reflects the issue of economies of scale, mainly in overlap component. Significant overlap is typical for economies saturated with big firms. Also urbanization economies are reflected in SPAG because of point location of data used in this index. Joint benefits of operating in cluster and use the same advantages of given location are proved by inter-sectoral low distance component.

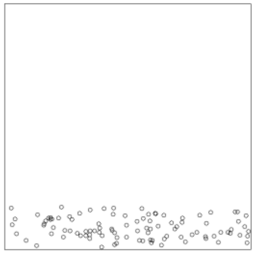
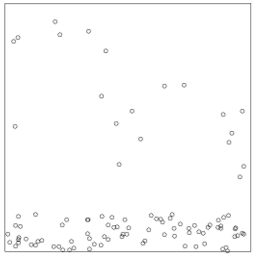
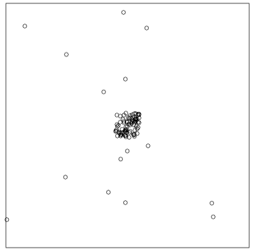
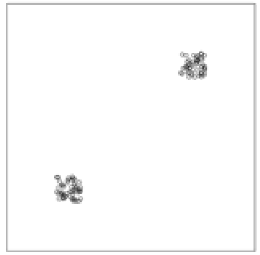
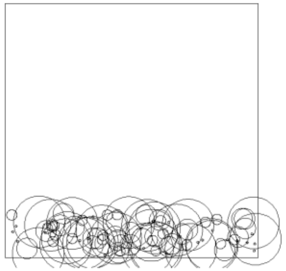
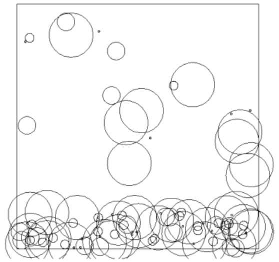
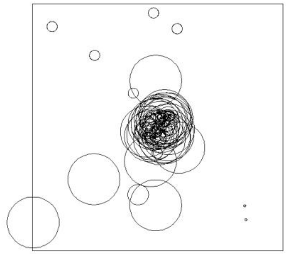
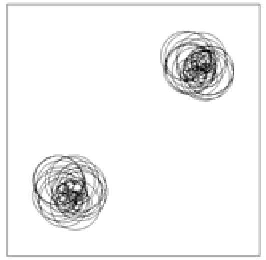
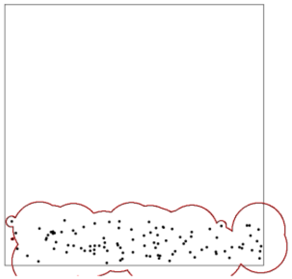
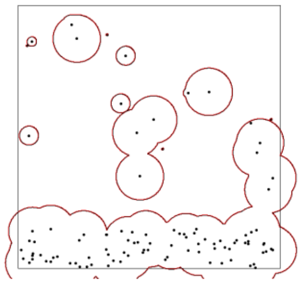
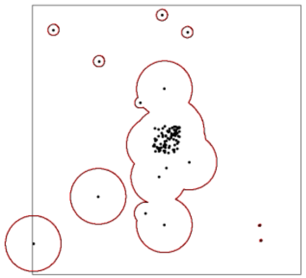
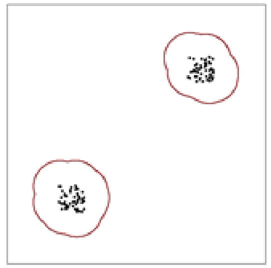
SPAG also meets the criteria of point-data index mentioned in point 2. As proved in Table 6 it is sensitive to different spatial distribution, what fulfills the 1st criterion. By construction it meets the 2nd and 4th criteria of referring to area of territory, size of firms and being an absolute measure. Its point results together with the confidence interval support 3rd condition of unequivocal interpretation. Finally one can claim that SPAG measures well the agglomeration over space what meets the 5th condition.

Table 6: Simulation of SPAG for all companies from given territory

Spatial distributions	**Random**	**Spatially uniform**	**Concentrated – one agglomeration and no units on peripheries**	**Border-dispersed**
Random localization of firms given with geo locations				
Firms represented by circles with radius r, that sum of areas of circles equals area of region located in their real locations, size of firms matters				
Union of overlapping circles A bird's eye view				
SPAG	$I_{coverage} = 1$ $I_{distance} = 1.03$ $I_{overlap} = 0.69$ SPAG = 0.72	$I_{coverage} = 1$ $I_{distance} = 1$ $I_{overlap} = 0.74$ SPAG = 0.74	$I_{coverage} = 1$ $I_{distance} = 0.1$ $I_{overlap} = 0.08$ SPAG = 0.008	$I_{coverage} = 1$ $I_{distance} = 1.31$ $I_{overlap} = 0.44$ SPAG = 0.57

Source: own simulation

Table 6: (continued) Simulation of SPAG for all companies from given territory

Spatial distributions	Costal distribution without interior locations	Costal distribution with interior locations	One agglomeration with few units on peripheries	Concentrated – two agglomerations and no units on peripheries
Random localization of firms given with geo locations				
Firms represented by circles with radius r, that sum of areas of circles equals area of region located in their real locations, size of firms matters				
Union of overlapping circles A bird's eye view				
SPAG	$I_{coverage} = 1$ $I_{distance} = 0.61$ $I_{overlap} = 0.28$ SPAG = 0.17	$I_{coverage} = 1$ $I_{distance} = 0.89$ $I_{overlap} = 0.46$ SPAG = 0.41	$I_{coverage} = 1$ $I_{distance} = 0.26$ $I_{overlap} = 0.23$ SPAG = 0.06	$I_{coverage} = 1$ $I_{distance} = 0.74$ $I_{overlap} = 0.13$ SPAG = 0.1

5. Confidence interval for SPAG

In its construction, SPAG is the function of two variables: spatial distribution of location and distribution of companies' size.

$$SPAG = f(\text{ location, size }) \tag{9}$$

Empirical SPAG is calculated for empirical locations and empirical size of companies.

$$SPAG_{empir} = f(\text{ location}_{empir}, \text{size}_{empir}) \tag{10}$$

In research on spatial location, when number of companies and its size is known, the question under examination is whether spatial distribution of impact of firms is uniform. This should be understood as the question if companies of given size in their real location cover with their range (business impact zone) the territory similarly as if they were located uniformly (with the size of companies etc. ceteris paribus). Under null hypothesis the permutation of SPAG can examine the uniformity of spatial location. On contrary under alternative hypothesis, one can test if spatial dispersion of companies' impact is towards the borders (repulsion) or towards single point (agglomeration). Thus in permutation test, the companies of size given empirically are permutated over the theoretical space – uniformly distributed points[3]. This allows for testing for how much uniform location would change the SPAG in comparison to empirical location, keeping size of business units as given. Permutations of $SPAG_{test}$ index depending on empirical size of companies in theoretical locations is being calculated as follows

$$SPAG_{test} = f(\text{ location}_{theor}, \text{size}_{empir}) \tag{11}$$

This allows for constructing (1-α) confidence interval of $SPAG_{test}$, with lower and upper border values $SPAG_{test,\,L}$, $SPAG_{test,\,U}$, and assumed significance level α:

$$Pr(\, SPAG_{test,L} < SPAG_{test} < SPAG_{test,U})=1-\alpha \tag{12}$$

Distribution of permutated SPAG shows what are the possible results of SPAG if firms of empirical size located uniformly (grid-like). Different values of SPAG result from overlapping effects which is because of the different size of firms. When $SPAG_{empir}$ belongs to the distribution of $SPAG_{test}$ one can assume the null hypothesis that impact zones of firms are as in situation of spatial uniform distribution, what suggests no agglomeration effects. Low SPAG values are typical for spatial concentration patterns. When $SPAG_{empir}$ is significantly lower than lower border of confidence interval of $SPAG_{test,L}$, one can conclude on agglomeration effects and clustering of firms over territory of region. High values of SPAG appear when no overlapping and the distance between firms is relatively high. When $SPAG_{empir}$ is significantly higher than upper border of confidence interval of $SPAG_{test,U}$, than one conclude on border-dispersed location of business, the so-called "donut" model of location.

Permutations of SPAG give the results which build the confidence intervals for testing the hypothesis on spatial uniformity of business impact zones. The critical centiles for α (e.x. 5% and 95%) give the lower and upper borders of $SPAG_{test}$. Also the whole distribution of all permutated values can be tested against normality (e.g. with normality Shapiro test). When normality hypothesis confirmed, empirical distribution is symmetric and no bias phenomena appear.

[3] Fratesi (2007) notes that permutation of companies requires the assumption on the same size of companies, when size of company not included in the measure. This is because units of different size could have different impact in different locations. In case of SPAG, size of unit is reflected in the area of circle. Thus permutation's goal is to see different impacts of size in uniform locations.

6. Summary

Traditionally, the analysis of concentration, specialization and agglomeration is conducted with cluster-based indices based on the two-dimensional employment matrix (Gini, Ellison-Glaeser etc.), its extensions by spatial weights matrix W or finally with distance-based indices using Ripley's K function and its modifications. These measures hardly include the real spatial locations and are not sensitive to different point patterns, or do not measure with single index the density of economic activity, or do not allow for analysis of chain values.

We tackle with these problems by constructing new measure of spatial agglomeration, the SPAG index. SPAG index is to determine to what extent the companies on the territory (e.x. in the region) are evenly distributed over space or follow spatial agglomeration pattern. The novelty and uniqueness of this model is that the index is based on geo-location of business units, and not depending on the administrative areas. SPAG is anchored in the geometrical model of representation of firms with circles, and not location quotient or Ripley's K function. SPAG index is a product of three indices: coverage, distance and overlap. Proposed measure can assess in absolute terms the density of economic activity inside the region and not only the saturation of economic activity in interregional comparison. It takes into account the spatial interactions and functional relationships, the importance of administrative borders or core centers instead of treating the region as a an isolated atom. Model can also be used in assessment of location of the horizontal and vertical chains. SPAG produces information which was not available till now.

We propose a confidence interval to determine if empirical value of SPAG can be interpreted as uniform distribution of firms over territory or if spatial distributions significantly differ. Values of SPAG can be calculated for any territory and sector, for firms $n \geq 2$. Its value is absolute without reference to other regions. Values of SPAG can supplement the traditional measures based on two-dimensional table of economic activity (by regions and sectors).

In the paper we propose also new criteria for emerging group of continuous-space individual data indicators, where single firm is represented by the point on the surface. There are sensitivity to different spatial distributions, including area of territory, as well as size and importance of companies, unequivocal interpretation, being an absolute measure and detecting concentration and dispersion over space. We also prove that our model fulfills all criteria mentioned above.

References:

Aiginger, K. (1999). Do Industrial Structures Converge? A survey on the empirical literature on specialisation and concentration of industries. *WIFO working paper* No. 116, 1999.

Arbia, G. (2001a). Modelling the geography of economic activities on a continuous space. *Papers in Regional Science*, 80, 411-424

Arbia, G. (2001b). The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems*, (2001) 3:271-281

Arbia, G., Espa, G., Giuliani, D., &Mazzitelli, A. (2010). Detecting the Existence of Space-time Clustering of Firms. *Regional Science and Urban Economics* 40 (5): 311-323.

Arbia, G., & Piras, G. (2009). A new class of spatial concentration measures. *Computational Statistics and Data Analysis* 53 (2009) 4471-4481

Bedward, M. (2010). Circle packing with R, *R-bloggers*, http://www.r-bloggers.com/circle-packing-with-r/

Bergstrand, J. H. (1985).The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence. *The Review of Economics and Statistics*, MIT Press, vol. 67(3), 474-81.

Bickenbach, F., & Bode, E. (2008). Disproportionality measures of concentration, specialization, and localization. *International Regional Science Review* 31 (4), 359_388.

Bickenbach, F., Bode E. & Krieger-Bode, Ch. (2012). Closing the gap between absolute and relative measures of localization, concentration and specialization. *Papers in Regional Science*, vol.92 no 3 August 2013

Bruelhart, M., & Traegerb, R. (2005). An account of geographic concentration patterns in Europe. *Regional Science and Urban Economics* 35 (2005) 597– 624

Carlei, V., & Nuccio, M. (2014). Mapping industrial patterns in spatial agglomeration: A SOM approach to Italian industrial districts. *Pattern Recognition Letters* 40 (2014) 1-10

Churski, P., Polko, A., Ochojski, A., & Kopczewska, K. (2015). How to understand regional specialization? Economic and regional science perspective, *Working Paper*

Cutrini, E. (2009). Using entropy measures to disentangle regional from national localization patterns. *Regional Science and Urban Economics* 39 (2):243-250. doi:10.1016/j.regsciurbeco.2008.08.005

Do, Q.-A., & Campante, F. (2009). A centered measure of spatial concentration: A gravity-based approach with an application to population and capital cities. *Harvard University*, mimeo.

Duranton, G., & Puga, D. (2001) .Nursery Cities: Urban Diversity, Process Innovation, and the Life-Cycle of Products. *American Economic Review* 91(5), 1454.1477.

Duranton, G., & Overman,. H.G. (2005). Testing for Localization Using Micro- geographic Data. *The Review of Economic Studies* 72(4): 1077–1106.

Duranton, G., & Overman, H. G. (2008). Exploring the detailed location patterns of UK manufacturing industries using microgeographic data. *Journal of Regional Science*, 48: 213–243.

Ellison, G., & Glaeser, E. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *The Journal of Political Economy* 105 (5), 889-927.

Franceschi, F., Mussoni, M., &Pelloni,. G. (2009). A Note on New Measures of Agglomeration and Specialization. *Unpublished, University of Bologna*. amsacta.cib.unibo.it/2683/

Fratesi, U. (2008). Issues in the measurement of localization. *Environment and Planning A* 2008, vol. 40, pp. 733-758, doi:10.1068/a39223

Guillain, R, & Le Gallo, J. (2010). Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis. *Environment and Planning B* 37 961–81

Guimaraes, P., Figueiredo, I., & Woodward D. (2011)., Accounting for neighboring effects in measures of spatial concentration. *Journal of Regional Science*, vol. 51 no. 4, 2011, pp.678-693

Hallet, M. (2000) .Regional specialisation and concentration in the EU. *Economic Papers*, 141, Brussels: European Commission.

Hoover, E.M. (1937). Location Theory and the Shoe and Leather Industries. *Harvard University Press*, Cambridge, MA

Jensen, P., & Michel, J. (2011). Measuring spatial dispersion: exact results on the variance of random spatial distributions. *Annals of Regional Science* (2011) 47:81-110

Kang, H. (2010)., Detecting agglomeration processes using space–time clustering analyses. *The Annals of Regional Science* (2010), 45:291-311

Kominers, S. D. (2007). Measuring Agglomeration. 2007. Unpublished expository paper. www.scottkom.com/articles/measure_agglomeration.pdf.

Krugman, P. R. (1991). Geography and trade. *MIT press*, 1991.

Lang, G., Marcon, E., & Puech, F. (2014). Distance-Based Measures Of Spatial Concentration: Introducing A Relative Density Function. <hal-01082178> https://hal.archives-ouvertes.fr/hal-01082178

Marcon, E., & Puech, F. (2003). Evaluating the Geo-graphic Concentration of Industries Using Distance-based Methods. *Journal of Economic Geography* 3(4): 409–428.

Marcon, E., &Puech,. F. (2009). Measures of the Geographic Concentration of Industries: Improving Distance-based Methods*. Journal of Economic Geography* 10(5):745–762.

Marcon, E., & Puech, F. (2015). A typology of distance-based measures of spatial concentration, <halshs-00679993v3> version 3

Maurel, F., & Sedillot, B. (1999). A measure for geographical concentration of French manufacturing industries. *Regional Science and Urban Economics* 29 (5), 575_604.

Mori, T., & Smith, T. (2014). A probabilistic modeling approach to the detection of industrial agglomeration, *Journal of Economic Geography* 14(3):547-588

Morphet, C. S. (1997). A statistical method for the identification of spatial clusters. *Environment and Planning A*, 29: 1039–1055.

Palan, N. (2010). Measurement of Specialization – The Choice of Indices, *FIW Working Paper* No 62, December 2010

Sohn, J. (2014). Industry classification considering spatial distribution of manufacturing activities. *Area*, 2014 46.1 101-110

Spiezia, V. (2002). Geographic Concentration of Production and Unemployment in OECD countries. *Cities and Regions*, December Issue.

Tian, Z., (2013). Measuring Agglomeration Using the Standardized Location Quotient with A Bootstrap Method. *Journal of Regional Analysis & Policy* 43(2): 186-197

van Groenigen, J.W., Siderius, W., Stein, A. (1999). Constrained optimisation of soil sampling for minimization of the kriging variance. *Geoderma* 87, 239–259.

Wang, J., Stein, A., Gao, B., & Ge, Y. (2012). A review of spatial sampling. *Spatial Statistics* 2 (2012) 1-14 doi: 10-1016/j.spasta.2012.08.001