POINT PATTERN METHODS FOR ANALYZING INDUSTRIAL LOCATION

Miguel Gómez-Antonio
  Universidad Complutense de Madrid and Instituto de Estudios
  Fiscales (Spain)
Ángel Alañón-Pardo
  Universidad Complutense de Madrid and Instituto Complutense
  de Estudios Internacionales (Spain)
  Corresponding author: aalanonp@ucm.es

ABSTRACT

Literature on point pattern methods for analyzing geographical concentration of firms has increased dramatically over the last decade. Revision of the state of the art in empirical applications shows that most methods are mainly exploratory while others focus on the identification of cluster determinants. We contribute in this regard by analyzing key features that underline the differences among exploratory methods: Functional form, selection of controls, significance of results, and treatment of edge effects. We also stress the potential and complementarity of new methods such as Gibbs models.

Keywords: Industrial location, point pattern analysis, Gibbs models, distance-based measures.

JEL Classification: C40, R12, R30.

MÉTODOS DE PATRONES DE PUNTOS PARA ANALIZAR LA LOCALIZACIÓN INDUSTRIAL

RESUMEN

La literatura sobre métodos de análisis de patrones de puntos para estudiar la concentración geográfica de las empresas ha aumentado espectacularmente en la última década. La revisión de la literatura empírica muestra que la mayoría de los métodos son principalmente exploratorios, mientras que otros se centran en la identificación de los determinantes de la aglomeración. En este artículo se analizan las características clave que subrayan las diferencias entre los métodos exploratorios: forma funcional, selección de controles, significación de los resultados y tratamiento de los efectos borde. Además, se destaca el potencial y la complementariedad de nuevos métodos como los modelos de Gibbs.

Palabras clave: localización industrial, análisis de patrones de puntos, modelos de Gibbs, medidas basadas en la distancia.

Clasificación JEL: C40, R12, R30.

1. INTRODUCTION

Geographic concentration of economic agents and clusters of interconnected companies are a striking feature of virtually every national, regional, state, and even metropolitan economy (Porter, 2000), since it is of paramount importance to explain growth determinants, regional disparities and economic development. Theoretical and empirical interest in this subject has risen recently due to the New International Trade theories and to the New Economic Geography (Krugman, 1998), and to the increasing processes of production relocation. Recent research establishes that one of the most promising strategies for intra-urban job growth lies in promoting localized clusters that produce goods and services that are sold primarily within a single city, metropolitan area, or urban region (Garrocho-Rangel, Álvarez-Lobato, and Chávez, 2013).

Literature on industrial location classifies methodology for analysing industrial clusters into three generations (Combes and Overman, 2004; De Dominicis, Arbia, and De Groot, 2013; Chain *et al.*, 2019). The first generation was focused on assessing whether the concentration of a given

industry was above or below other industries or the overall activity, and used measures such as Gini, Herfindahl, or the Location Quotient indices. The second generation compared the spatial concentration in an industry with the one obtained if the location of economic units followed a random pattern, and is based on the seminal dartboard approach developed in Ellison and Glaeser (1997). First and second generations tools have usually been applied to predefined administrative geographical units of observation (regions, counties, metropolitan areas…), which raises the modifiable areal unit problem (MAUP)[1]. To overcome this problem, the third generation methods are borrowed from the point pattern analysis literature (PPA)[2]. These methods do not discretize the area of study, but take advantage of all the information contained in the geo-referenced data, therefore admitting the inspection over a range of scales (Scholl and Brenner, 2016).

The aim of this paper is to provide a guide to researchers interested in the application of point pattern methods to analyse industrial agglomeration. The most popular and promising methods based on the distance among firms are compared according to their functional form, to the selection of controls, to the significance of results, and to the treatment of border effects. Review of the state of the art in empirical applications shows that most papers are exploratory and use a case-control strategy to detect (co-)localization. The literature needs to move forward towards a regression framework in line with the estimation of Gibbs models, which have proven successful in other fields, and can shed light on the determinants of clusters.

This research adds to other papers that have focused on cluster based measures of regional concentration such as Kopczewska (2018) and on distance-based measures of spatial concentration, such as Scholl and Brenner (2016) or Marcon and Puech (2017).

The remainder of the paper is organized as follows. In the following section, different distance-based approaches applied in PPA to deal with industrial location are analyzed. Then, the empirical literature is reviewed. Finally, concluding remarks are summarized.

---

[1]   See Openshaw and Taylor (1979) for a more detailed discussion of this topic.
[2]   See Boots and Getis (1988) for more information about point pattern analysis.

2. POINT PATTERN METHODS FOR ANALYZING THE INDUSTRIAL LOCATION

This section discusses the most popular or promising distance-based measures that have been empirically tested to study industrial location in the presence of spatial inhomogeneity[3]. For this purpose, these measures are grouped according to their exploratory nature, first section, or confirmatory nature, second section.

2.1. Exploratory distance-based measures[4]

Table 1 list the most popular distance-based measures such as the *D function* (Diggle and Chetwynd, 1991), the *Duranton and Overman (D-O)* approach (Duranton and Overman, 2005), the *M function* (Marcon and Puech, 2010), and the *inhomogeneous K function* (Baddeley, Møller, and Waagepetersen, 2000). All these methods are exploratory and grounded in a case-control strategy that consist on selecting a group of controls that account for the observed inhomogeneity and compare its spatial distribution with that of the selected cases. It is then claimed that localization economies will manifest themselves as a phenomenon of extra-concentration in one industry with respect to the concentration of the firms in the control population. We evaluate these methods by focusing on four characteristics that may be relevant to researchers: The functional form used, the sampling process for the selection of controls, the procedure for testing the significance of the results, and the treatment of edge effects[5].

2.1.1. Different functional forms

Diggle and Chetwynd (1991) *D function* evaluates the difference between cases and controls estimating *K functions*. Where the *K function*

---

[3] There are more distance-based measures, such as the firm level index developed by Scholl and Brenner (2016) or the Spatial Agglomeration Index (SPAG) developed by Kopczewska (2017).

[4] For a more comprehensive review of distance-based measures of spatial concentration see Marcon and Puech (2017).

[5] For the sake of simplicity, we illustrate only univariate function expressions.

Table 1. Distance-based measures

D function*

$$D(r) = K_{case}(r) - K_{control}(r)$$

r: distance; λ: intensity of the process; N: number of points; w_{ij} = weights for correct for edge effects; $c_s(i,j,r) = 1$, if distance between two plants, i and j of industry s is less than the radius r, and = 0 otherwise.

D-O approach**

$$\hat{K}_d(r) = \frac{1}{N(N-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{r - d_{i,j}}{h}\right);$$

$$D - O = \hat{K}(r)_{cases} - \hat{K}(r)_{controls}$$

N: number of points; h = bandwidth; d_{ij}: distance between i and j; r: distance.

M function***

$$M_s(r) = \left. \frac{\sum_{j=1, i \neq j}^{N_s} c_s(i,j,r)}{\sum_{j=1, i \neq j}^{N} c(i,j,r)} \middle/ \frac{N_S}{N} \right.$$

$c_s(i,j,r) = 1$, if distance between two plants, i and j of industry s is less than the radius r, and = 0 otherwise; N: total number of firms in the area of analysis A.

Inhomogeneous *K*-function****

$$\hat{K}_I(r, \lambda) = \frac{1}{|A|} \sum_{i=1}^{n} \sum_{j \neq i} \frac{w_{ij} c(i,j,r)}{\lambda(x_i)\lambda(x_j)};$$

/A/ is the total surface of the area

w_{ij}: the adjustment factor for edge effects correction; $c_s(i,j,r) = 1$, if distance between two plants, i and j of industry s is less than the radius r, and = 0 otherwise; $\lambda(x)$: intensity at location x.

Source: Own elaboration based on Diggle and Chetwynd (1991)*; Duranton and Overman (2005)**; Marcon and Puech (2010)***, and Baddeley, Møller, and Waagepetersen (2000)****.

is a cumulative function, first introduced by Ripley (1976), to detect distributions deviation from randomness in homogeneous, stationary and isotropic spatial processes. It counts events up to each distance r of the point[6]. Duranton and Overman's (2005) approach evaluates the difference, for cases and controls, of the kernels of the probability point-pair distance density functions, that counts the number of points at a distance r of each event. They compare whether the number of plants at a given distance is significantly different from the number that would have been found if the location of the firms was random. Marcon and Puech's (2010) *M function* counts neighboring points up to a chosen distance r and compares them with all industrial activities in a circle of radius r, while also accounting for the size of the sector relative to all activities in the study region. Baddeley, Møller, and Waagepetersen's (2000) *Inhomogeneous K-function* gives each point a weight that is inversely proportional to the local density of points, so more neighbors are expected where more points are located. It essentially generalizes *Ripley's K function* for non-stationary point processes in which second order intensity reweighted-stationarity is assumed.

Several methodological considerations need to be stressed in each of the approaches. First, when the area of analysis is not large, a caveat in the *D-O* approach is the presence of the mathematical problem of compensation (Marcon and Puech, 2010). The probability density function (Kd) must sum to one, so if the results show localization at short distances it will necessarily determine dispersion at longer distances which is not a consequence of real inhibition but a compensation effect. To avoid this problem Duranton and Overman (2005) proposed to analyze only a range of distances that reach the median distance in the sample. Second, a caution in *D-O* and the *Inhomogeneous K function* is the need to choose an arbitrary kernel bandwidth, although mathematical procedures can be implemented for its selection, there is no reason to assume that the bandwidth should be equal for cases and for controls. If the number of events in the sample is not large, the results would be highly dependent

---

[6] When points have marks (attributes) attached the K-function can be extended to the bivariate K_{ij} function that counts the events of type j up to a chosen distance r of the point i. When $i = j$ the bivariate K_{ij} function represents the K function for either the cases ($i = j = 1$) or the controls ($i = j = 2$).

on the arbitrary choice of the estimation kernel bandwidth (Diggle *et al.*, 2007). If the chosen bandwidth is small the intensity is highly variable, and the results will determine independency while choosing a wider band will result in more stationarity and dependence (Marcon and Puech, 2010).

Another aspect greatly debated is the fact that *D*, *M* and *Inhomogeneous K* are cumulative functions while *D-O* employs a probability density function. Marcon and Puech (2010) demonstrate that probability density functions can detect local clusters more precisely at different spatial scales, but cumulative function approaches can detect the existence of clusters up to a certain distance and spatial repulsion between the clusters. Recent papers have proven that every function can be modified in terms of obtaining its cumulative/probability density function counterpart. Behrens and Bougna (2015) construct a cumulative function based on the probability density function of *D-O*, and Lang, Marcon, and Puech (2020) propose a new relative distance-based function *m* equivalent to the *M function*, where instead of counting points up to a radius *r*, they estimate the kernel of the probability density function of point-pair distances.

Another aspect to consider is that *D-O*, *M* and the *Inhomogeneous K function* approaches can be extended to explicitly take into account the concentration of a few firms of a considerable size by incorporating the number of employees into the clustering metric instead of just the number of plants. We prefer to control for employment as a confounding factor in the control selection process, in the way that the *D function* approach matches the stratified sample as will be described in the next subsection. Establishments are the principal units among which externalities-inducing interactions are likely to occur, implying that the more enterprises in a given area, the more likely they are to enjoy positive externalities based on co-location. As the concept of localization economies deals with external economies of scale, when introducing employment weights, internal and external scale economies are conflated.

2.1.2. Selection of controls

To select the controls *D function* simulates drawing a random proportionately size matched sample of controls from the population of in-

dustries that do not belong to the target industry. The sample should match the cases in its main characteristics, such as the size of the firms or the organizational structure to properly capture the firm's interaction and not the effect that those features might have on clustering. At the same time, to avoid biased results, a procedure is implemented to detect the existence of outliers that might have an excessive impact in the reference distribution of controls ("outlier effect"). In the empirical applications, as sample size constraints generate variability in the magnitude of the *D function*, the results are usually reported averaged over multiple samples.

The selection of controls in *D-O's approach* is obtained by sampling from the sites of total manufacturing firms. This procedure is less precise because firms might have different attributes that affect their propensity to cluster, and if firms with those attributes are overrepresented in the control sample, the results will be biased. Furthermore, if the reference distribution of controls is dominated by one industry with a strong tendency to cluster (inhibition), it will be hard that when the cases are compared to the controls, the results will show cluster (inhibition). Even when both effects are softened by simulating a high number of subsamples, it is important to note that there is no reason why the stratified sample and the outlier identifier procedure could not also be applicable to *D-O's approach*.

M function presents two refinements regarding *D function* and *D-O's approach*. First, the need to select a sample of controls is avoided, as the whole manufacture is used as the reference distribution. However, this does not control for the existence of one industry in the area with a strong tendency to cluster/inhibition, or for the existence of the effects of a certain firm's attributes on clustering. Second, if there are data for more than one sector pattern, they explicitly consider the relative size of the industry in the denominator, allowing the direct comparison of results between industries and regions of different size. Nonetheless the approach still assumes that all industries location choices are affected by the same covariates.

Inhomogeneous K function needs to select controls only when the first-order intensity functional form is not known and needs to be estimated. It cannot be estimated non-parametrically from the same observed pattern from which the function is estimated. Without any other information or assumption about the underlying process, it is not

possible to distinguish spatial inhomogeneity from a spatial dependence phenomenon (Diggle *et al.*, 2007).

2.1.3. Significance of the results

In order to formally assess the significance of the empirically observed values of the results the only approach in which a formal test of significance can be utilized is *D function*[7]. Because the exact distribution of *D* function is known, its variance can be evaluated theoretically and proper confidence bands can be constructed. This advantage is limited, as the rate of convergence to the asymptotic multivariate normal distribution is unclear (Diggle and Chetwynd, 1991). To implement an exact test and construct the confidence bands, Monte Carlo techniques are still required as in *D-O's approach* and *M function* empirical functions.

2.1.4. Treatment of edge effects

Edge effects arise because the theoretical distributions for most spatial point statistics assume an unbounded area, yet observed distributions are estimated from delineated regions. Edge effects will tend to distort the estimated function for the points that are close to the boundary because the possibility of having neighbors outside the boundary is denied. While *D function* and *K inhomogeneous function* correct for edge effects, *D-O's approach* and *M function* do not account for their existence implying bias for large *r*. Our intuition is that the *D-O* approach does not correct for edge effects because its seminal empirical application is for the United Kingdom, therefore, the probability of finding an event out of the boundary is truly zero. The case of the *M function* is different as it is a quotient of two quotients. Comparing the number of neighbors in a certain industry to the total number of neighbouring establishments in the same area (in a disc of radius r in the numerator and in the whole region in the denominator), the edge effects cancel out.

---

[7] In Scholl and Brenner (2016) a new index has been proposed which allows the use of significance tests.

2.1.5. Relevant theoretical and practical issues on the application of distance based measures

Overall, the main limitation in analyzing the existence of agglomeration under a case-control strategy is that the robustness of the results heavily depends on an appropriate specification of the properties of the spatial process represented by the control group. Two important aspects of the description of agglomeration/dispersion are spatial inhomogeneity (first-order intensity), and spatial interaction (second-order variation). Spatial inhomogeneity relates to the fact that some regions may have a mean number of points higher than others; for example, in the Central Business District we might have a higher concentration of firms, while the suburbs might show a lower density of firms. Besides that, spatial interaction relates to the dependence between points in pairs of locations. Case-control strategies are very useful for determining the spatial scale of any cluster of firms, but as long as there are influences that are clearly unique to the industry under evaluation, even when the controls are a correct representation of first order intensity these approaches are not appropriate to capture the nature of agglomeration. The *Inhomogeneous K function* approach is the only one that could identify this pattern by introducing a proper covariate to estimate the first-order intensity. Case control analysis basically yields a series of isolated univariate comparisons that determine if "case" industries exhibit more co-location than "control" industries.

Summarizing, though most methods analyzed are useful to measure agglomeration as a whole, most of them do not allow for the identification of the location of clusters. While it is possible to claim that first-order effects are controlled, most methods cannot be used in regression models to learn about its reationship with the analysed industry.

2.2. Gibbs regression model approach

Gibbs models for point processes (or Markov point processes) provide a regression framework that constitutes a fruitful approach to empirically explore cluster's determinants. We show in Table 2 that a Gibbs (Markov) process, *X*, can be expressed as exponential family densities and allow for separate estimation of effect sizes on components of the

Table 2. Gibbs model approach

Gibbs point process density (with respect to Poisson unit density)

$$f(x) = a \prod_{y \subseteq x} b(u) S(u,x)$$

Møller and Waagepetersen (2007)
a: normalizing constant; $S(u,x)b(u)$: density at location u; $S(u,x)$: Interactions among points (in pairs, triples or higher orders).

Papangelou conditional intensity:

$$\lambda(u,x) = \exp\left\langle \Phi^T b(u) + \theta^T S(u,x) \right\rangle$$

$\Phi\theta$: estimated canonical parameters.

trend (first-order effects) and a specific representation of the interaction (second-order effects). The trend component $b(u)$ depends only on the spatial location u, and reflects spatial inhomogeneity that affects the location decision of firms. It captures aspects related to the natural or built environments that are likely to impact a firm's location choice. History might affect certain locations' population density, public service endowments and geographical characteristics also play an important role for certain industries. The interaction component $S(u,x)$ provides another set of rich specification choices. Several alternative specifications of the interaction can be applied to firm location modeling but only three of the functional forms of interaction have appeared in applied work: Strauss hard core, Geyer saturation, and Area/penetrable spheres.

Because of the normalizing constant a, it is difficult to work directly with the density and the model is made tractable by working instead with the *Papangelou* conditional intensity function, which is the probability of observing a point u of the process in a small neighborhood du of u, conditional upon the rest of the process X. Because of the exponential form of the Gibbs model, standard software implementations for generalized linear (additive) models can be used to estimate parameters of the conditional intensity function.

The main topics in this framework are the estimation of the actual characteristics (or parameters) of location potential, and its comparison within different economic/geographic situations. This framework is useful to isolate the different sources of agglomeration economies and

determine which covariates are most effective in explaining the observed agglomeration patterns. Gibbs models can yield a richer set of results than case-control methods and move from the simple hypothesis testing to complete model specification and validation that forms the basis for most empirical research in regional science. Details on the Gibbs process formulation can be found in Møller and Waagepetersen (2007) with extensions and interpretation for the firm location choice in Sweeney and Gómez-Antonio (2016).

3. EMPIRICAL LITERATURE REVIEW OF POINT PATTERN ANALYSIS TECHNIQUES TO DETECT LOCALIZATION ECONOMIES

The first to introduce these techniques to examine clustering of manufacturers was Barff (1987), who implemented *Ripley's K function* for the Cincinnati metropolitan area. A limitation of this analysis is that the null hypothesis is a completely random spatial distribution of establishments. Case-control methodologies were developed to avoid this assumption. Review of empirical application allows to determine the impact of each methodology, and to highlight how substantial recent research has focused on developing new methods or extensions, mainly under the case-control strategy. Besides, it shows that the literature is orphan in order to identify the determinants of industrial clusters and how certain papers cover this drawback by estimating area-based linear regression models. Finally, we hope it is helpful in order to guide and attract new researchers to this field.

Table 3 summarizes the main research questions and the methodological deviations, if any, of each paper. Most of the papers that has focused on a single exploratory distance-based measure are devoted to the D-O approach[8], or to the D function.

In addition to the application of Ripley's K function and its modifications, while there are numerous papers comparing *D-O's approach, D function* with other area-based indexes, there are few papers comparing

---

[8] D-O approach is one of most cited articles according to the World of Science horizontal criteria (Chain *et al.*, 2019) and it still inspires new indices, such as the new multisectorial co-location index developed by Pablo-Martí and Arauzo-Carod (2020).

two or more distance-based approaches to the same dataset. Overall the reviewed papers are very accurate at detecting and identifying the relevant distance for cluster existence, but are purely descriptive and say nothing about the causes of the departure from randomness.

These approaches, either explicitly or implicitly, aim to distinguish join localization from colocation (first-order from second-order concentration). To isolate the potential factors determining concentration, several papers analyzed different subgroups of firms according to their characteristics, others considered the dynamic dimension, proposed new cumulative or density counterpart functions, or extended the *Inhomogeneous K function* approach to marked weighted patterns. Recently, several papers have estimated linear regression models of distance-based indexes on covariates to explicitly identify industrial cluster determinants.

Nevertheless, in order to identify cluster's determinants, point process models can be defined and fitted to data. Although this approach has proven to be successful in other fields, they do not have attracted a similar interest for the analysis of industrial location determinants.

Sweeney and Gómez-Antonio (2016) were the first to fit explicit Gibbs models to point pattern data incorporating both spatial inhomogeneity and inter-point interactions to explain the observed pattern of industrial establishments. The inhomogeneity is modeled with the covariates distance to the city center and distances to certain type of roads. The estimated interaction effects of the Strauss hard core, of the Geyer saturation and of the Area interation models can be interpreted as evidence of the strength and scope of localization economies. Subsequently, Gómez-Antonio and Sweeney (2018) estimated a Gibbs model to test the role of local public goods on attracting establishments to the city. A complete model of location choices is estimated detangling first-order from second-order interaction effects. Their results challenged some of the outcomes of the inter-urban industrial location literature.

Published work estimating Gibbs models is very limited and to date has focused on unmarked Gibbs models. The next step in this analysis is extending the methods and results for qualitative or quantitative marked spatial point process in Euclidean space or on a road network. The extension allows for different levels of attraction or repulsion among different categories of industry, employment size or both.

Table 3. Distance based measures and Gibbs models applications

Diggle and Chetwynd's (1991) D function and similar approaches	
Sweeney and Feser (1998)	Clustering and firm size
Feser and Sweeney (2002a)	Collocation in different product value chains, U.S. metropolitan areas
Feser and Sweeney (2002b)	Collocation in Printing and Publishing, responsibility of value chain in clustering or dispersion
Feser and Sweeney (2000)	Collocation in 12 product value chains and 14 metropolitan areas
Sweeney and Konty (2005)	Robust estimation for spatially censored data
D function without a stratified sample procedure to construct the counterfactuals	
Marcon and Puech (2003)	14 manufacturing industries in greater Paris and France
Kosfeld, Eckey, and Lauridsen (2011)	Mining and manufacturing in Germany, application of the subsample similarity to reduce computational requirements
Albert, Casanova, and Orts (2012)	Manufacturing in Spain, first and second nature advantages
Buzard *et al.* (2017)	Technological spillovers in R&D labs in the Northeast corridor of the U.S.; Monte Carlo procedure to use manufacturing employment distribution as a counterfactual; multiscale core-cluster approach based on local K functions to determine the size and shape of clusters
Helbich and Leitner (2010)	Clustering of higher-order services in the urban fringe of Vienna
Lotwick and Silverman (1982)'s extension of Rypley's K (colocation between two industries)	
Arbia, Espa, and Quah (2008)	Patent innovations for six industrial sectors in Italy
Arbia *et al.* (2010)	Information and Communication Technology firms in Rome 1920-2005 period; space time K function
The Mark-weighted K function by Penttinen (2006)	
Giuliani, Arbia, and Espa (2014)	Hi tech firms in Italy

Table 3. Distance based measures and Gibbs models applications (continued…)

Duranton and Overman's (2005) approach, D-O	
Duranton and Overman (2005)	Manufacturing industries in United Kingdom
Duranton and Overman (2008)	Co-localization between British industries and location patterns according to firm characteristics. Comparison between co-localization forces and own-industry clustering
Klier and McMillen (2008)	New auto supplier plants in the U.S.
Nakajima, Saito, and Uesugi (2012)	Duranton and Overman (2008) replicated for manufacture and service industries in Japan
Kerr and Kominers (2015)	Patent technology clusters in Sillicon Valley
Barlet, Briant, and Crusson (2013)	Index of divergence in the space of density distributions that is fully comparable across industries, service and manufactures in France
Murata *et al.* (2014)	U.S. patent citation and localization of knowledge spill-overs
Alfaro and Chen (2014)	Index to text driving forces of multinationals vs domestic firms; data for plants in over 100 countries
Behrens (2016)	Coagglomeration patterns in Canadian automotive industry
Marcon and Puech (2010) M function approach	
Jensen and Michel (2011)	Shops in Lyon (France)
Lang, Marcon, and Puech (2020)	Density function of cumulative M function, m function approach
Moreno-Monroy and García-Cruz (2016)	Agglomeration a co-agglomeration patterns of informal and formal manufacturing industries in Cali (Colombia)
Coll-Martínez, Moreno-Monroy, and Arauzo-Carod (2019)	Agglomeration and co-agglomeration of creatives industries in Barcelona metropolitan area (Spain) using M and m functions
Baddeley, Møller, and Waagepetersen's (2000) Inhomogeneous K function approach	
Arbia *et al.* (2012)	High Tech industries in Milan
Espa, Arbia, and Giuliani (2013)	High Tech industries in Milan, trend surface cuadratic model

Table 3. Distance based measures and Gibbs models applications (continued…)

Comparison papers	
Cao *et al.* (2017)	Location patterns in Shangai, Ripley's K, Leslie and Kronenfeld co-location index, kernel density function and nearest neighbourhood analysis
Bonneu and Thomas-Agnan (2015)	Extension of Inhomogeneous K function, D-O and M
Funderburg and Zhou (2013)	D-function *versus* D-O, 20 manufacturing industry clusters in California
Billings and Johnson (2012)	D-O *versus* an Index of specialization for service and manufacturing industries in Denver
Scholl and Brenner (2016)	A Firm Level Cluster Index, D-O and M function for German microsystem technology industry.
Sweeney and Feser (2004)	D-function, Localization quotient, EG, and G(s) (Gettis and Ord's G) for six manufacturing industries in Los Angeles and Atlanta
Duranton and Overman's (2005) *versus* Ellison and Glaeser (EG)	
Vitali, Napoletano, and Fagiolo (2013)	Manufacturing in six European countries
Koh and Riedel (2014)	Four digit industries and industrial services in Germany
Behrens and Bougna (2015)	Location in Canada; 2001-2009 period, spatially weighted EG

4. CONCLUDING REMARKS

We have documented the main warnings worth considering when utilizing distance-based methods to detect agglomeration. Some of these caveats could be softened by incorporating some characteristics of one methodology into the others. The stratified sample of the controls that implements the *D function* could be easily translated to the *D-O's* approach and conversely the procedure to construct global envelopes in *D-O* and the *M function* could be implemented for the *D function* approach.

D function and D-O's (2005) approaches have been more frequently employed than the others. Substantial recent research has been focused

Table 3. Distance based measures and Gibbs models applications (continued…)

Regression models	
Klier and McMillen (2008)	D-O, conditional logit model for American auto supplier plants
Ellison, Glaeser, and Kerr, (2010)	D-O and EG and Marshallian agglomeration forces in US industries
Alfaro and Chen (2014)	D-O and EG and agglomeration forces in multinationals firms
Arbia (2001)	Krugman's (1991) firm demography model for San Marino Republic
Arbia *et al.* (2015)	Firm demography processes and spatial interactions in Trento (Italy)
Sweeney and Gómez-Antonio (2016)	Gibbs models for electronic industry in Madrid (Spain)
Gómez-Antonio and Sweeney (2018)	Gibbs models and the role of public goods in electronic industry in Madrid (Spain)

on developing new functions or extensions that share the limitations of case-control strategies. Some of the approaches have been extended to explicitly take into account the concentration of a few firms of a considerable size by incorporating the number of employees into the clustering metric instead of just the number of plants. Establishments are the principal units among which externalities-inducing interactions are likely to occur, implying that the more enterprises in a given area, the more likely they are to enjoy positive externalities based on co-location. As the concept of localization economies deals with external economies of scale, when introducing employment weights, internal and external scale economies are conflated.

Case control methods are very useful to detect the cluster scope but say nothing about the causes of the departure from randomness. Understanding cluster determinants remains crucial and case-control approaches, either explicitly or implicitly, aim to distinguish join localization from colocation (first-order from second-order concentration). This distinction is of extreme importance because allows to better understanding location process and, therefore, to implement policies that

help firms providing the type of environment that is needed (spatial characteristics, specialized services, access to the same type of inputs, infrastructures…). Some of the research questions in the analysis of behavioral factors and relocation remain unresolved.

Public policy evaluation processes can benefit from the use of measures such as Gibbs models constitute a promising research line. Gibbs models allow the identification of the sources that explain clusters and are flexible enough to avoid confounding results. Their specific advantage is that they provide a regression framework that takes point-referenced data as unit of observation. Therefore, Gibbs models yield a far richer set of results than prior methods.

Summing up, the choice among the methods may well depend on the research question to be addressed, since as shown in Kopczewska (2017) and in Scholl and Brenner (2016) these measures are complementary and combined applications of them are meaningful. ◄

REFERENCES

Albert, J.M., Casanova, M.R., and Orts, V. (2012). Spatial location patterns of Spanish manufacturing firms. *Papers in Regional Science*, 91(1), pp. 107-136. https://doi.org/10.1111/j.1435-5957.2011.00375.x

Alfaro, L., and Chen, M.X. (2014). The global agglomeration of multinational firms. *Journal of International Economics,* 94(2), pp. 263-276. https://doi.org/10.1016/j.jinteco.2014.09.001

Arbia, G. (2001). Modelling the geography of economic activities on a continuous space. *Papers in Regional Science*, 80(4), pp. 411-424. https://doi.org/10.1111/j.1435-5597.2001.tb01211.x

Arbia, G., Cella, P., Espa, G., and Giuliani, D. (2015). A micro spatial analysis of firm demography: The case of food stores in the area of Trento (Italy). *Empirical Economics*, 48(3), pp. 923-937. https://doi.org/10.1007/s00181-014-0834-6

Arbia, G., Espa, G., Giuliani, D., and Mazzitelli, A. (2010). Detecting the existence of space-time clustering of firms. *Regional Science and Urban Economics*, 40(5), pp. 311-323. https://doi.org/10.1016/j.regsciurbeco.2009.10.004

Arbia, G., Espa, G., Giuliani, D., and Mazzitelli, A. (2012). Clusters of firms in an inhomogeneous space: The high-tech industries in Milan. *Economic Modelling*, 29(1), pp. 3-11. https://doi.org/10.1016/j.econmod.2011.01.012

Arbia, G., Espa, G., and Quah, D. (2008). A class of spatial econometric methods in the empirical analysis of clusters of firms in the space. *Empirical Economics,* 34(1), pp. 81-103. https://doi.org/10.1007/s00181-007-0154-1

Baddeley, A.J., Møller, J., and Waagepetersen, R. (2000). Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3), pp. 329-350. https://doi.org/10.1111/1467-9574.00144

Barff, R.A. (1987). Industrial clustering and the organization of production: A point pattern analysis of manufacturing in Cincinnati, Ohio. *Annals of the Association of American Geographers*, 77(1), pp. 89-103. https://doi.org/10.1111/j.1467-8306.1987.tb00147.x

Barlet, M., Briant, A., and Crusson, L. (2013). Location patterns of service industries in France: A distance-based approach. *Regional Science and Urban Economics*, 43(2), pp. 338-351. https://doi.org/10.1016/j.regsciurbeco.2012.08.004

Behrens, K. (2016). Agglomeration and clusters: Tools and insights from co-agglomeration patterns. *Canadian Journal of Economics/Revue canadienne d'économique*, 49(4), pp. 1293-1339. https://doi.org/10.1111/caje.12235

Behrens, K., and Bougna, T. (2015). An anatomy of the geographical concentration of Canadian manufacturing industries. *Regional Science and Urban Economics*, 51, pp. 47-69. https://doi.org/10.1016/j.regsciurbeco.2015.01.002

Billings, S.B., and Johnson, E.B. (2012). A non-parametric test for industrial specialization. *Journal of Urban Economics*, 71(3), pp. 312-331. https://doi.org/10.1016/j.jue.2011.12.001

Bonneu, F., and Thomas-Agnan, C. (2015). Measuring and testing spatial mass concentration with micro-geographic data. *Spatial Economic Analysis*, 10(3), pp. 289-316. https://doi.org/10.1080/17421772.2015.1062124

Boots, B.N., and Getis, A. (1988). *Point Pattern Analysis*. Newbury Park, CA: SAGE Publications.

Buzard, K., Carlino, G.A., Hunt, R.M., Carr, J.K., and Smith, T.E. (2017). The agglomeration of American R&D labs. *Journal of Urban Economics*, 101, pp. 14-26. https://doi.org/10.1016/j.jue.2017.05.007

Cao, W., Li, Y., Cheng, J., and Millington, S. (2017). Location patterns of urban industry in Shanghai and implications for sustainability. *Journal of Geographical Sciences*, 27(7), pp. 857-878. https://doi.org/10.1007/s11442-017-1410-8

Chain, C.P., Santos, A.C.d., Castro Júnior, L.G.d., and Prado, J.W.d. (2019). Bibliometric analysis of the quantitative methods applied to the measurement of industrial clusters. *Journal of Economic Surveys*, 33(1), pp. 60-84. https://doi.org/10.1111/joes.12267

Coll-Martínez, E., Moreno-Monroy, A., and Arauzo-Carod, J. (2019). Agglomeration of creative industries: An intra-metropolitan analysis for Barcelona. *Papers in Regional Science,* 98(1), pp. 409-431. https://doi.org/10.1111/pirs.12330

Combes, P., and Overman, H.G. (2004). The spatial distribution of economic activities in the European Union. In: J.V. Henderson and J.-F. Thisse, *Handbook of Regional and Urban Economics*, Vol. 4 (pp. 2845-2909). The Netherlands: Elsevier.

De Dominicis, L., Arbia, G., and De Groot, H.L. (2013). Concentration of manufacturing and service sector activities in Italy: Accounting for spatial dependence and firm size distribution. *Regional Studies,* 47(3), pp. 405-418. https://doi.org/10.1080/00343404.2011.579593

Diggle, P.J., and Chetwynd, A.G. (1991). Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics,* 47(3), pp. 1155-1163. https://doi.org/10.2307/2532668

Diggle, P.J., Gómez-Rubio, V., Brown, P.E., Chetwynd, A.G., and Gooding, S. (2007). Second-order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics*, 63(2), pp. 550-557. https://doi.org/10.1111/j.1541-0420.2006.00683.x

Duranton, G., and Overman, H.G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4), pp. 1077-1106. https://doi.org/10.1111/0034-6527.00362

Duranton, G., and Overman, H.G. (2008). Exploring the detailed location patterns of UK manufacturing industries using microgeographic data. *Journal of Regional Science*, 48(1), pp. 213-243. https://doi.org/10.1111/j.1365-2966.2006.0547.x

Ellison, G., and Glaeser, E.L. (1997). Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105(5), pp. 889-927. https://doi.org/10.1086/262098

Ellison, G., Glaeser, E.L., and Kerr, W.R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3), pp. 1195-1213. 10.1257/aer.100.3.1195

Espa, G., Arbia, G., and Giuliani, D. (2013). Conditional *versus* unconditional industrial agglomeration: Disentangling spatial dependence and spatial heterogeneity in the analysis of ICT firms' distribution in Milan. *Journal of Geographical Systems*, 15(1), pp. 31-50. https://doi.org/10.1007/s10109-012-0163-2

Feser, E., and Sweeney, S.H. (2002a). Spatially binding linkages in manufacturing product chains. *Global Competition and Local Networks* (pp. 111-130). Londres: Routledge.

Feser, E.J., and Sweeney, S.H. (2002b). Theory, methods and a cross-metropolitan comparison of business clustering. In: P. McCann (ed.), *Industrial Location Economics* (pp. 222-259). Cheltenham Glos: Edward Elgar Publishing.

Feser, E.J., and Sweeney, S.H. (2000). A test for the coincident economic and spatial clustering of business enterprises. *Journal of Geographical Systems*, 2(4), pp. 349-373. https://doi.org/10.1007/PL00011462

Funderburg, R.G., and Zhou, X. (2013). Trading industry clusters amid the legacy of industrial land-use planning in southern California. *Environment and Planning A*, 45(11), pp. 2752-2770. https://doi.org/10.1068%2Fa45393

Garrocho-Rangel, C., Álvarez-Lobato, J.A., and Chávez, T. (2013). Calculating intraurban agglomeration of economic units with planar and network K-functions: A comparative analysis. *Urban Geography*, 34(2), pp. 261-286. https://doi.org/10.1080/02723638.2013.778655

Giuliani, D., Arbia, G., and Espa, G. (2014). Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration. *International Regional Science Review*, 37(3), pp. 251-272. https://doi.org/10.1177%2F0160017612461357

Gómez-Antonio, M., and Sweeney, S. (2018). Firm location, interaction, and local characteristics: A case study for Madrid's electronics sector. *Papers in Regional Science*, 97(3), pp. 663-685. https://doi.org/10.1111/pirs.12274

Helbich, M., and Leitner, M. (2010). Postsuburban spatial evolution of Vienna's urban fringe: Evidence from point process modeling. *Urban Geography*, 31(8), pp. 1100-1117. https://doi.org/10.2747/0272-3638.31.8.1100

Jensen, P., and Michel, J. (2011). Measuring spatial dispersion: Exact results on the variance of random spatial distributions. *The Annals of Regional Science*, 47(1), pp. 81-110. https://doi.org/10.1007/s00168-009-0342-3

Kerr, W.R., and Kominers, S.D. (2015). Agglomerative Forces and Cluster Shapes. *The Review of Economics and Statistics*, 97(4), pp. 877-899. https://doi.org/10.1162/REST_a_00471

Klier, T., and McMillen, D.P. (2008). Evolving agglomeration in the US auto supplier industry. *Journal of Regional Science*, 48(1), pp. 245-267. https://doi.org/10.1111/j.1467-9787.2008.00549.x

Koh, H., and Riedel, N. (2014). Assessing the localization pattern of German manufacturing and service industries: A distance-based approach.

Regional Studies, 48(5), pp. 823-843. https://doi.org/10.1080/00343404.2012.677024

Kosfeld, R., Eckey, H., and Lauridsen, J. (2011). Spatial point pattern analysis and industry concentration. *The Annals of Regional Science*, 47(2), pp. 311-328. https://doi.org/10.1007/s00168-010-0385-5

Kopczewska, K. (2017). Distance-based measurement of agglomeration, concentration and specialisation. In: K. Kopczewska, P. Churski, A. Ochojski, and A. Polko (eds.), *Measuring Regional Specialisation. A New Approach* (pp. 173-216). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-51505-2_3

Kopczewska, K. (2018). Cluster-based measures of regional concentration. Critical overview. *Spatial Statistics,* 27, pp. 31-57. https://doi.org/10.1016/j.spasta.2018.07.008

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy,* 99(3), pp. 483-499. https://doi.org/10.1086/261763

Krugman, P. (1998). What's new about the new economic geography? *Oxford Review of Economic Policy,* 14(2), pp. 7-17. https://doi.org/10.1093/oxrep/14.2.7

Lang, G., Marcon, E., and Puech, F. (2020). Distance-based measures of spatial concentration: Introducing a relative density function. *The Annals of Regional Science*, 64, pp. 243-265. https://doi.org/10.1007/s00168-019-00946-7

Lotwick, H., and Silverman, B. (1982). Methods for analysing spatial processes of several types of points. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(3), pp. 406-413. https://doi.org/10.1111/j.2517-6161.1982.tb01221.x

Marcon, E., and Puech, F. (2003). Evaluating the geographic concentration of industries using distance-based methods. *Journal of Economic Geography*, 3(4), pp. 409-428. https://doi.org/10.1093/jeg/lbg016

Marcon, E., and Puech, F. (2010). Measures of the geographic concentration of industries: Improving distance-based methods. *Journal of Economic Geography*, 10(5), pp. 745-762. https://doi.org/10.1093/jeg/lbp056

Marcon, E., and Puech, F. (2017). A typology of distance-based measures of spatial concentration. *Regional Science and Urban Economics*, 62, pp. 56-67. https://doi.org/10.1016/j.regsciurbeco.2016.10.004

Møller, J., and Waagepetersen, R.P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34(4), pp. 643-684. https://doi.org/10.1111/j.1467-9469.2007.00569.x

Moreno-Monroy, A.I., and García-Cruz, G.A. (2016). Intra-metropolitan agglomeration of formal and informal manufacturing activity: Evidence from Cali, Colombia. *Tijdschrift voor economische en sociale geografie*, 107(4), pp. 389-406. https://doi.org/10.1111/tesg.12163

Murata, Y., Nakajima, R., Okamoto, R., and Tamura, R. (2014). Localized knowledge spillovers and patent citations: A distance-based approach. *Review of Economics and Statistics*, 96(5), pp. 967-985. https://doi.org/10.1162/REST_a_00422

Nakajima, K., Saito, Y.U., and Uesugi, I. (2012). Measuring economic localization: Evidence from Japanese firm-level data. *Journal of the Japanese and International Economies*, 26(2), pp. 201-220. https://doi.org/10.1016/j.jjie.2012.02.002

Openshaw, S., and Taylor, P. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In: N. Wrigley and R. Bennett (eds.), *Statistical Applications in the Spatial Sciences*. London: Pion Press.

Pablo-Martí, F., and Arauzo-Carod, J. (2020). Spatial distribution of economic activities: A network approach. *Journal of Economic Interaction and Coordination*, 15, pp. 441–470. https://doi.org/10.1007/s11403-018-0225-8

Penttinen, A. (2006). Statistics for marked point patterns. In: *The Yearbook of the Finnish Statistical Society* (pp. 70-91). Helsinki: The Finnish Statistical Society.

Porter, M.E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic Development Quarterly*, 14(1), pp. 15-34. https://doi.org/10.1177%2F089124240001400105

Ripley, B.D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2), pp. 255-266. https://doi.org/10.2307/3212829

Scholl, T., and Brenner, T. (2016). Detecting Spatial Clustering Using a Firm-Level Cluster Index. *Regional Studies*, 50(6), pp. 1054-1068.

Sweeney, S.H., and Feser, E.J. (1998). Plant size and clustering of manufacturing activity. *Geographical Analysis,* 30(1), pp. 45-64. https://doi.org/10.1111/j.1538-4632.1998.tb00388.x

Sweeney, S.H., and Feser, E.J. (2004). Business location and spatial externalities: Tying concepts to measures. In: M.F. Goodchild and D.G. Janelle (eds.), *Spatially Integrated Social Science* (pp. 239-262). New York: Oxford University Press.

Sweeney, S.H., and Konty, K.J. (2005). Robust point-pattern inference from spatially censored data. *Environment and Planning A*, 37(1), pp. 141-159. https://doi.org/10.1068%2Fa35318

Sweeney, S., and Gómez-Antonio, M. (2016). Localization and industry clustering econometrics: An assessment of Gibbs models for spatial point processes. *Journal of Regional Science*, 56(2), pp. 257-287. https://doi.org/10.1111/jors.12238

Vitali, S., Napoletano, M., and Fagiolo, G. (2013). Spatial localization in manufacturing: A cross-country analysis. *Regional Studies*, 47(9), pp. 1534-1554. https://doi.org/10.1080/00343404.2011.625006