Spatial distribution of polluting facilities in Europe

José M. Albert
Departamento de Economía
Universitat Jaume I

Marta R. Casanova
Departamento de Economía Aplicada
Universitat de València
marta.roig-casanova@uv.es

Jorge Mateu
Departamento de Matemáticas
Universitat Jaume I

Vicente Orts
Departamento de Economía
Universitat Jaume I

Área Temática: *Energía, sostenibilidad, recursos naturales y medio ambiente*

In this paper we estimate the intensity function of a Cox process modelling the point pattern representing the spatial distribution of polluting facilities in Europe. The intensity of a Cox process is variable, depending on the location, therefore we can work with inhomogeneous spatial point patterns. We obtain a non-parametric spatial distribution function which allow us to know the agglomeration of polluting facilities in a chosen number of European sub-regions.

Palabras Clave: *agglomeration, Cox process, point process, polluting facilities, spatial Concentration*

Clasificación JEL: *C15, C21, L16, Q21, R30*

1. Introduction

Spatial agglomeration or clustering is one of the key stylized facts of the geographical location of individual firms and industries. Consequently, the spatial distribution of economic activity is basically inhomogeneous, and the spatial inequalities become de norm more than the exception at many economic relevant levels. The topics of economic activity location and spatial concentration of firms have attracted the attention of economists for long time ago. Thus, from the pioneering works of Von Thünen (1826), Marshall, (1890), Weber (1909), Lösch, (1940), Hoover (1948), among others, to more recent contributions of the "new economic geography" (for a survey of this literature see Duranton, (1997), Ottaviano and Puga, (1998), Fujita, Krugman and Venables, (1999), Neary, (2001), and Jovanovic, (2007)), the interest of economists has been to characterize the patterns of geographic concentration of firms and industries, and analyze the forces that can allow to explain these inequalities across space. However, in contrast to the recent advances in our understanding of economic forces that can allow to explain spatial disparities, the advances in the empirical characterization of the spatial patterns of economic activity have been comparatively small, and a lot of work must be done to provide both accurate feedbacks to theory, and strong stylized facts in this field.

The work on the characterization of the spatial distribution of economic activity has focused mainly on the calculation of indices, more or less complex, and more recently, estimating homogeneous spatial distribution functions (Sweeney and Feser (1998), Marcon and Puech (2003), Duranton and Overman (2005)). These new tools use distance based methods (see Figure 1), avoiding one of the more annoying problems of traditional indices, the modifiable areal unit problem (MAUP), that is, their dependence of the particular administrative scale chosen. Nevertheless, these methods faces at least one important shortcoming: they supposes homogeneous spatial processes[1], becoming inappropriate to analyze non-homogeneous point sets. In this paper we take a step more

[1] These authors used as method of analysis the K-function (Ripley, 1977), that means that the measures only yield valid results when they are used to asses spatial patterns where there is no large scale variation in the mean of the process (homogeneous point process).

forward, and propose a distance-based method[2], but able to detect spatial structure of inhomogeneous process, the estimation of an intensity function of a Cox process.

Finally, we illustrate the use of this tool with an application to an interesting database: polluting facilities[3] and their spatial location within the European Union.

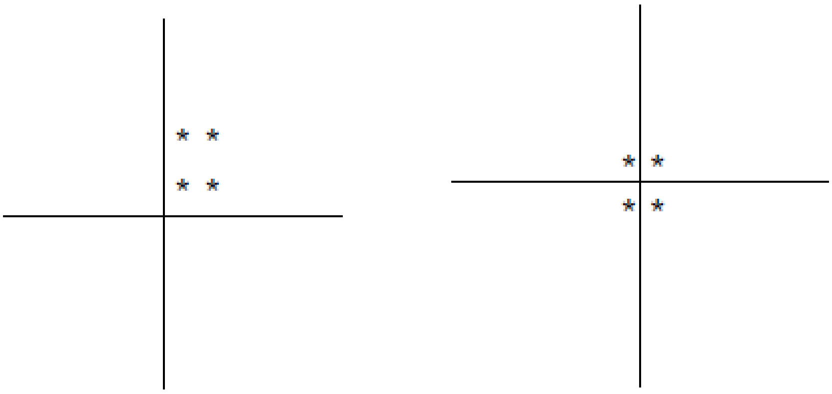

Figure 1: The cluster level of a certain points set can depend on the regional borders

The rest of de paper is organized as follows. Section 2 shows the statistical framework, section 3 outlines the methodology used for the intensity function estimation, in section 4 we develop a practical application analyzing the spatial distribution of the polluting industrial activities in Europe, and finally, in the section 5 some final considerations are made.

2. Statistical framework

Definition 1. *A point process is a stochastic model governing the locations of events* $\{s_i\}$ *in some set X, a bounded region in* R^d, *Cressie (1993).* A spatial point pattern is a partial realisation of a stochastic point process (Cox and Isham (1980)). Such a realization may be visualized very easily as a cloud of points in a bounded region.

---

[2] This method is also used by Quah and Simpson (2003).
[3] Data obtained from the European Pollutant Emission Register (EPER).

Definition 2. *A homogeneous Poisson process with intensity* λ *is defined by:*

1. *The number of events* $s_1, s_2, ..., s_n$ *in any bounded region* $A \in X$ *follows a Poisson distribution with mean* ¸ $\lambda|A|$

(being $|A|$ = surface of *A*)

2. *The n events are uniform and independently distributed in A.*

In accordance with condition (2.1), the intensity λ is constant. The condition (2.2) avoids the existence of interactions among the events. In economics the presence of agglomeration phenomena is usual, and this kind of phenomena cannot be treated with homogeneous processes, we must work with inhomogeneous point processes. Any inhomogeneous process defines a variation of the inhomogeneous Poisson process in the following way.

Definition 3. *Let* $\lambda(s): X \to R_+$ *be a non-constant function on X. The set of events* $s_1, s_2, ..., s_n$ *in X is an inhomogeneous Poisson process if:*

1. *For each bounded region* $A \in X$ *the number of events N(A) = n follows a Poisson distribution with mean* $\int_A \lambda(s)ds$.

2. *Given n events in A, the localizations* $s_1, s_2, ..., s_n$ *in X form a random and independent sample of this distribution on A with a density that depends on the intensity function* $\lambda(s)$, $s \in A$, *and whose values depend on the different localizations.*

In some applications it is useful to allow that the intensity function $\lambda(s)$ of a Poisson process can vary in space, to transform itself into a stochastic process $\Lambda(s)$.

Definition 4.

1. *Let* $\Lambda(s)$, $s \in X$ *denote a non-negative stochastic process on X.*

2. *A realization of* $\Lambda(s)$ *will be a Cox process driven by* Λ *if this realization is an inhomogeneous Poisson process with intensity function* $\Lambda = \lambda$

The result of the realization inherits the properties of the process $\Lambda(s)$ in a natural way.

3. Estimation of the intensity function

In this paper we use the estimation method proposed by Berman and Diggle (1989). Our aim is to estimate the inhomogeneous version of the intensity function λ and $\lambda(s)$ respectively, starting from a data set, $\{s_i \in A : i = 1,2,...,n\}$ being A a planar region.

We can estimate λ by the usual estimator

$$\tilde{\lambda} = n / |A|$$

since λ is defined as the expected number of events by area unit.

The usual estimator for $\lambda(s)$ is

$$\tilde{\lambda}(s,h) = \left(\pi h^2\right)^{-1} N\left(\pi h^2\right),$$

the observed number of events by area unit in a disk with radius *h*, centred in a point $s_i \in s$.

Using kernel method for smoothing point process data,

$$\tilde{\lambda}(s;h) = h^{-2} \sum_{i=1}^{n} f\left\{h^{-1}(s - s_i)\right\}$$

where $(s - s_i)$ denotes distance among points, and

$$f(u) = \begin{cases} \pi^{-1} : \|u\| \le 1 \\ 0 : \|u\| > 1 \end{cases}$$

The border effects are kept in mind modifying the denominator and obtaining

$$\tilde{\lambda}(s;h) = N\left(\pi h^2\right) / A(s;h)$$

where $A(s;h)$ is the area of the intersection among the observed region *A* and the disk with radius *h* centred in *s*.

The final form of the estimator is:

$$\tilde{\lambda}(s;h) = h^{-2} \sum_{i=1}^{n} f\left\{h^{-1}(s - s_i)\right\} / A(s;h)$$

The method chooses the smoothing parameter *h* to be that value minimizing the mean square error, for $MSE(h) = E\left\{\left[\tilde{\lambda}_h(x) - \Lambda(x)\right]^2\right\}$.

4. Application

4.1. Data

We use geographical coordinates (longitude and latitude) representing the localization of all the pollutant industrial firms included in the European Pollutant Emission Register (EPER), transformed in UTM coordinates using the procedure shown in Morton (2003).

The left hand side of *Figure 2* shows the point pattern which represents the location of the EU[4] pollutant firms. A simple visual inspection of this graph makes us think that we are faced with a point pattern which is a realization of an inhomogeneous spatial point process. Next, we have estimated the inhomogeneous intensity function of this point pattern using the method of Berman and Diggle, by means of the free software R and CRAN (2004),this software allows us to calculate the equation (2) for a certain number of sub-regions inside the study region. In all the cases, the study region was divided in 400 equal square sub-regions, with a side of 160 km of longitude, and the calculation of the intensity has been performed taking an interaction distance of h = 0.01 (32 km), since this was the value that minimized the mean square error. That is to say, we considered the distance between any pollutant facility and all those located in their surroundings in a radius of 32 km.

4.2. Results

The main results are shown graphically. The surface variations on the right hand side of *Figure 2* represent the values of the 400 estimated local intensities, whose magnitude[5] are shown in the chart of the *Figure 3*. Each one of these 400 sub-regions is identified by their UTM coordinates, therefore, for each sub-region we know both, the location and the intensity in pollutants industrial facilities. In this paper we take a further step more and present the Cox process as a useful instrument to measure the cluster or concentration of inhomogeneous point processes. Since the results obtained with the

[4] The old European Union of 15 countries, plus Hungary and Norway.

[5] This magnitude is function of both, the number of pollutant facilities inside the sub-region, and the number of pollutant facilities located around this sub-region in a distance equal or smaller than 32 kilometres.

method of Berman and Diggle are sensitive both to the number of points and to the distance among them, to make comparable the estimates obtained on points patterns with different number of points, we act in the following way: for each different number of points, and through Monte Carlo simulations, we obtain enough random points patterns to build confidence intervals with the values obtained when fitting the intensity of a Cox process; using the measures of intensity that surpass those of the confidence interval, we can make comparisons of concentration measures among different point patterns.

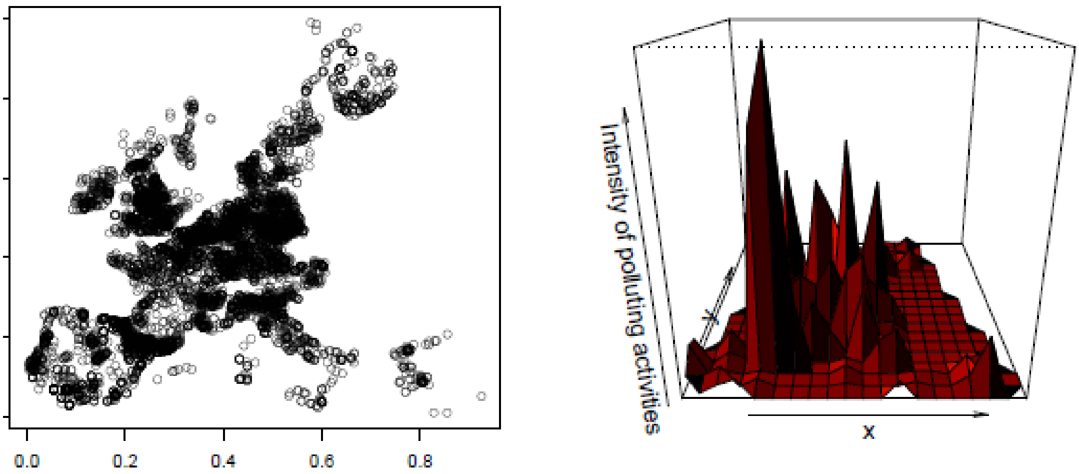

Figure 2: Spatial location and intensity of pollutant facilities in UE

Acting this way with the point pattern which represent the spatial location of pollutant facilities in EU we obtain the graph of the Figure 4. In Figure 4, the axis of ordinates measures the fitted values of local intensity, and we use these values as measures of pollutant industrial facilities concentration in EU regions; in this Figure the confidence interval (95% of confidence) is represented by the distance among the two parallel lines in the low side of the graph ranging from 0 to 17.256,00. These values above the confidence limit indicate cluster, and the more far away they are from the interval, the larger is the cluster. The axis of abscissas represents each one of the 400 sub-regions in which we have divided the European Union. Figure 4 indicates that exists a strong spatial concentration of pollutant industrial facilities in the studied European regions,

mainly in the northeast of Spain (north of Catalonia), around the territory among Belgium, Luxembourg and Germany, in the southeast of England, in the area between France and Belgium, and in the centre of Germany.

5. Final considerations

In this work, we have presented an appropriate new method to analyze the inhomogeneous distribution of the polluting facilities in Europe. This method consists in estimating the intensity function of a Cox process, which allows us to obtain a measure of the polluting facilities concentration in any number of sub-regions (we have chosen 400 sub-regions) of same size inside the European Union, for any chosen (we have chosen 32 kilometres) distance. Since spatial agglomeration is one of the key stylized facts of geographical location of economic activity, the inhomogeneous spatial point processes are the more appropriate to represent the location of economic variables, of there the interest of the Cox Processes.

20	0	0	0	0	0	0	0	0	0	0	0	0	1161	0	499	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	232	11639	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	3687	0	6846	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	3352	6561	6755	2459	0	0	0	0
16	0	0	0	0	0	0	10	2415	0	0	0	4201	8229	575	18384	0	0	0	0	0
15	0	0	0	0	0	4002	0	0	0	0	0	27006	27028	0	0	0	0	0	0	0
14	0	0	0	0	3589	20635	5637	0	0	0	19078	10568	1582	0	0	0	0	0	0	0
13	0	0	0	6963	63667	9832	0	0	0	73765	2349	21096	1718	0	0	0	0	0	0	0
12	0	0	123	36330	0	41695	159	0	0	3600	19531	428	0	0	0	0	0	0	0	0
11	0	0	82408	188	8111	119735	39417	0	5882	95297	26880	38092	0	0	0	0	0	0	0	0
10	0	0	0	0	29849	43813	8424	116204	88823	51839	48467	114773	0	0	0	0	0	0	0	0
9	0	0	0	0	0	1628	15698	11899	9602	151932	31355	0	0	0	0	0	0	0	0	0
8	0	0	0	0	27558	32284	44242	36140	59409	28681	46802	34212	25430	0	0	0	0	0	0	0
7	0	0	0	0	566	11374	6864	17405	3708	0	5732	9353	10433	0	0	0	0	0	0	0
6	0	315	0	0	0	10810	2734	30388	14579	63280	75462	808	0	0	0	0	0	0	0	0
5	0	2706	19490	2118	22162	6343	10103	43943	20032	0	54009	18575	0	0	0	0	0	0	0	0
4	20708	0	9389	3274	177330	231026	50426	0	0	4341	0	19776	29321	0	0	0	1304	7747	0	0
3	24593	19	1557	0	2763	0	5	0	0	7446	0	0	429	2684	2711	0	16526	0	0	0
2	8160	9250	16045	0	6095	39	0	0	0	2257	0	0	0	3722	0	0	8553	40386	0	0
1	0	4702	11276	16109	0	0	0	0	0	0	0	0	14865	0	0	0	0	0	0	0
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Figure 3: Intensidad de instalaciones industriales contaminantes en U.E.

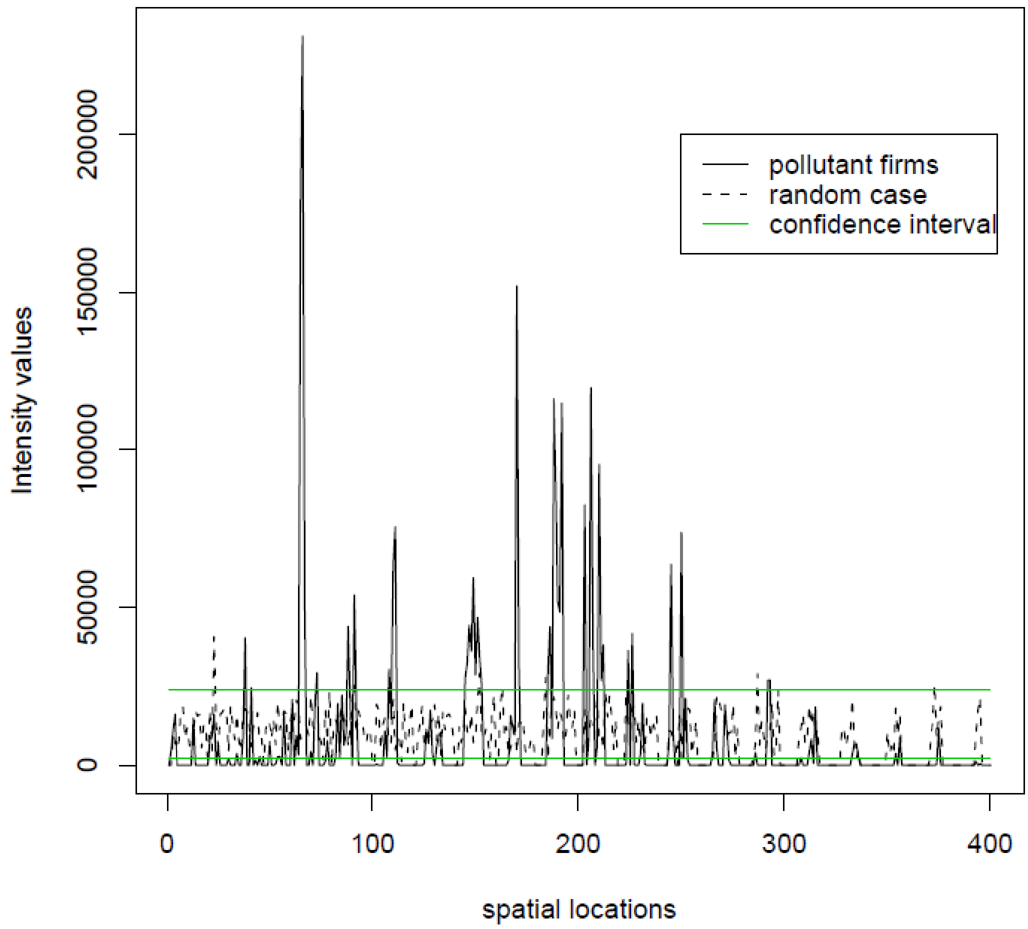

Figure 4: Medida de concentración de instalaciones industriales contaminantes en U.E.

References

M. Berman and P.J. Diggle (1989) Estimating weighted integrals of the second order intensity of a spatial point process. *Journal of the Royal Statistical Society*, B:81-92.

European Commission (2001) European pollutant emission register. Electronic publication, European Commission, http://www.eper.cec.eu.int.

D.R. Cox and V. Isham (1980) *Point processes*. Chapman & Hall, London.

N. Cressie (1993) *Statistics for Spatial Data*. John Wiley and Sons, New York.

G. Duranton (1997) La nouvelle economie géographique: agglomération et dispersion.

Economie et Prévision, 131(5):1-24.

Gilles Duranton and Henry G. Overman (2005) Testing for localization using micro-geographic data. *Review of Economics Studies*, 72:1077-1106.

Krugman P. Fujita, M. and A.J. Venables (1999) *The Spatial Economy*. MIT Press, Cambridge, Massachusetts.

E.M. Hoover (1948) *The Location of Economic Activity*. New York: MacGraw-Hill.

M.N. Jovanovic (2007) *Economic Integration and Spatial Location of Firms and Industries*, volume I, II of *The International Library of Critical Writings in Econonomics*. Edward Elgar Publishing Limited, Cheltenham, UK, Northampton.

Friedrich Leisch (2004) Precompiled binary distributions. Technical report, http://cran.r-project.org/, June.

A. Lösch (1940) *Die RaÄumliche Ordnung der Wirtschaft*. Iena, Gustav Fisher. Englishtraslation: The Economics of Location, Yale University Press(1954).

Eric Marcon and Florence Puech (2003) Evaluating the geographic concentration of industries using distance-based methods. *Journal of Economic Geography*, 3:409-428.

A. Marshall (1890) *Principles of Economics*. London: Macmillan.

Alan Morton (2003) Tm ll.xls workbook. Electronic publication, Distribution mapping software (DMAP), www.dmap.co.uk.

P Neary (2001) Of hype and hyperbolas: Introducing the new economic geography. *Journal of Economic Literature*, 39:536-61, June.

Puga D. Ottaviano, G.I.P. (1998) Agglomeration in the global economy: a survey of the new economic geography. *The World Economy*, 21(6):707-731.

D.T. Quah and H. Simpson (2003) Spatial cluster empirics. Working Paper,Economics Department, LSE, London, U.K., June.

S.H. Sweeney and E.J. Feser (1998) Plant size and clustering of manufacturing activity.*Geographical Analysis*, 30(1):45-64.

J.H. Von Thünen (1826) *Der Isolierte Staad in Beziehung auf Landwirtschaft un NationalÄokonomie*. Hamburg: Perthes. English traslation: The Isolated state, Oxford, Pergammon Press(1966),.

A. Weber (1909) *Ueber den Standort der Industrien*. TÄubingen: J.C.B. Mohr. English traslation: The Theory of the Location of Industries, Chicago University Press (1929).