QÜESTIIÓ, vol. 24, 1, p. 27-51, 2000

SPATIAL STRUCTURE ANALYSIS USING PLANAR INDICES⋆

J.M. ALBERT*
J. MATEU**
J.C. PERNÍAS*
Universitat Jaume I

Spatial planar indices have become a useful tool to analyze patterns of points. Despite that, no simulation study has been reported in literature in order to analyze the behaviour of these quantities under different pattern structures. We present here an extensive Monte Carlo simulation study focused on two important indices: the Index of Dispersion and the Index of Cluster Size, usually used to detect lack of homogeneity in a spatial point model. Finally, an application is also presented.

Keywords: Index of cluster size, index of dispersion, point processes, spatial structure

AMS Classification (MSC 2000): 62M30, 60G55

⋆ The authors wish to thank financial support by the Spanish Ministry of Education (CICYT: SEC96-1435-C03-03).

* Department of Economics. Universitat Jaume I. Campus Riu Sec. 12071 Castelló. Spain.

** Department of Mathematics. Universitat Jaume I. Campus Riu Sec. 12071 Castelló. Spain (mateu@mat.uji.es)

– Received June 1999.
– Accepted December 1999.

1. INTRODUCTION

A spatial point pattern is a collection of data $\{(x_i, y_i) \;\; i = 1, \ldots, n\}$ consisting of n locations in an essentially planar region. Examples include the locations of cell nuclei in a microscopic tissue section, trees in a forest, or cases of disease in a geographical region. A fundamental assumption in the analysis of such data is that they can usefully be regarded as a partial realisation of a stochastic point process (Cox & Isham, 1980). Many systems of individuals can also be described by attaching to the locations measurable quantities like, say, diameter of a tree. This last case is an example of marked points patterns.

There are many contexts in which the use of spatial point patterns can be very interesting, for example, in the analysis of the human population's spatial distribution, since they can give us excellent information from both the demographic and the economic points of view.

The interest in a pattern of points is found in that, with an appropriate choice of scale, even huge objects may be best represented by a point. Given suitable scales, the actual physical sizes of objects that may be represented that way are unbounded. On one extreme, microscopes are required, on the other extreme it is telescopes that are needed. The range of disciplines (pathology, geology, marine biology, zoology, physical and human geography, astronomy, economy,...) dealing with similar phenomena reflects the applicability of these techniques.

The concept of *complete spatial randomness* (CSR) is fundamental to the quantitative description of a spatial pattern. A formal definition of CSR is that the events in the region of observation A constitute a partial realisation of a homogeneous, planar Poisson process (Diggle, 1983). This process incorporates a single parameter, λ, the intensity, or mean number of events per unit area. The actual number of events in A, n say, is an observation from a Poisson distribution with mean $\lambda|A|$, where $|A|$ denotes the area of the region A. If we consider n as fixed, we arrive at the following definition of CSR: (1) each of the n events is equally likely to occur at any point within A; (2) the n events are located independently of each other. Our interest in CSR is that it represents an idealized standard which, if strictly unattainable in practice, may nevertheless be tenable as a convenient first approximation. Most of the analysis begin with a test of CSR, and there are several good reasons for this. Firstly, a pattern for which CSR is not rejected scarcely deserves any further formal statistical analysis. Secondly, tests are used as a means of exploring a set of data, rather than because rejection of CSR has an intrinsic interest. Thirdly, CSR acts as a dividing hypothesis to distinguish between patterns which are broadly classifiable as «regular» or «aggregated» (Figure 1).

A question of immediate interest is the following: Is it reasonable to expect a pattern of real data events to display randomness? Naturally, the answer depends on what these events represent. If, for example, the point pattern is defined as locations of cities, in order to evaluate the benchmark hypothesis of CSR, two contradictory forces are liable to have an effect on the observed cities locations. On one hand, the competition between neighbouring cities is likely to result in some thinning out of close neighbours, so that the pattern becomes rather more regular in appearance. On the other hand, variations in the local geography will result in some regions being more favourable for growth than are other regions. This will result in an apparent patchiness (or clustering) in the cities distribution. These two effects, and others, may well be sufficiently counterbalancing each other, so that the distribution of cities may yet retain the appearance of randomness.

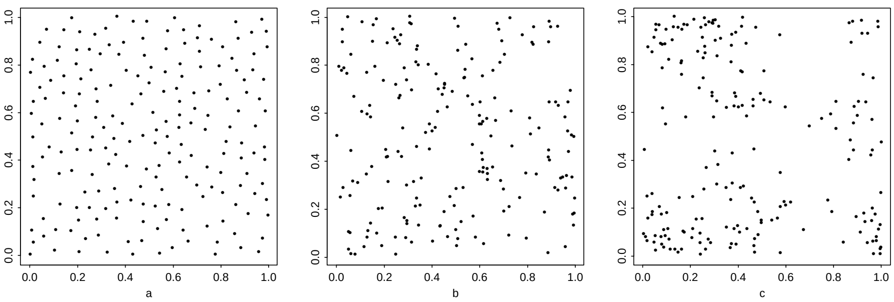

Figure 1. Examples of point patterns in the unit square showing three different spatial structures: *Left*, regular pattern; *Middle*, random pattern; *Right*, aggregated pattern.

In view of the remarkable diversity of mechanisms which may lead to an apparently random pattern, it must be stressed that a random distribution implies that the pattern has no discernible order and that its cause is undeterminable.

In literature, two major approaches have been suggested for the analysis of pattern. One involves measures of physical distances between points (Diggle 1983; Ripley 1981, 1988; Cressie 1993, Stoyan et al. 1995), the other involves analysis of the variation in the number of points in selected sub-areas of the region under study (Diggle, 1983; Upton & Fingleton, 1994). In this paper we concentrate on the second approach. All the counting methods rely on the use of quadrats. Several indices, like the *index of dispersion* and the *index of cluster size*, were proposed for scattered and contiguous quadrats. Generally, their behaviour under CSR is quite well known, though little is analyzed under departure of randomness. What is known is that the power of such indices depends in an unpredictable way on the size and shape of the individuals quadrats (Diggle, 1979, 1983; Perry & Mead, 1979; Stiteler & Patil, 1971).

This topic owes its generality as it is a general way of proceeding in detection of point pattern structures. Examples of this generality are the wide range of applications, the majority refering to contiguous quadrats. We can find applications in environmental sciences such as pattern analysis of perennial shrubs (Gulmon & Mooney, 1977), of herbs in the savanna (Hopkins, 1965) or detection of pattern in Lansing Woods trees (Diggle, 1983); in economics such as analysis of the distribution of houses (Moellering & Tobler, 1972), etc.

The goal of this paper is to analyze exhaustively the behaviour of quadrat-based indices to detect spatial structures. Particularly, we present an extensive simulation study to compare the performance of two of these indices under clear departures of CSR. In addition we present our own developed software, S.P.P.A. (1997), which can be used in this spatial context.

The plan of the paper is as follows. Section 2 presents the spatial indices. Section 3 is devoted to the simulation study presenting the whole set of results. Finally, Section 4 develops the analysis of a real application.

2. SPATIAL INDICES

Throughout this section, given an observed pattern, we attempt to deduce the nature of the process that gave rise to that pattern. Quadrat sampling involves collecting counts of events in subsets of the study region. Traditionally, these subsets are rectangular (hence the name of quadrats), although any shape is possible. Quadrats may be placed either randomly or layed out contiguously in the region.

This is very easy to implement as it is only needed to position quadrats randomly in the study region (Figure 2) and count the numbers of events that fall in each quadrat. It is true that scattered quadrats counts provide some limited information about the nature of a point pattern and that is our present concern. Counts from scattered quadrats are, at best, a crude indicator of pattern because they take into account neither the relative positions of the points within the quadrats nor the relative positions of the quadrats themselves.

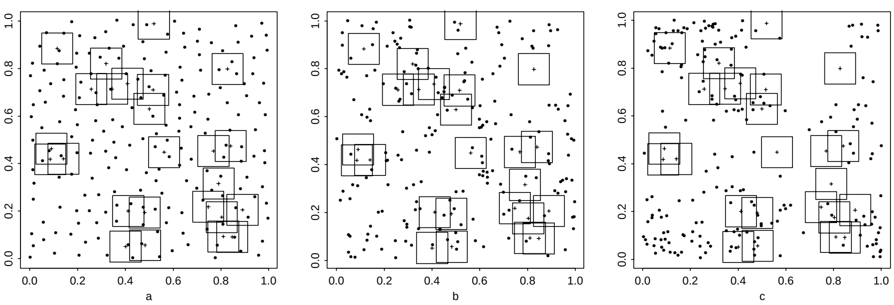

Figure 2. Equi-sized scattered quadrats superimposed over the point patterns.

The counts from equi-sized scattered quadrats (Figure 2) over a Poisson pattern will be observations from a Poisson distribution with parameter equal to the product of the intensity of the events per unit area and the area of the quadrats. A useful characteristic of the Poisson distribution is that its parameter is equal both to the mean and the variance of the distribution. If the pattern is more regular than a Poisson one, then the quadrat counts will be more uniform in size and will therefore have a relatively small variance (when compared with the size of the mean). On the other hand, if there are clusters, then some quadrats will have large counts so that the variance of the counts will be relatively large.

The analysis of grids of contiguous quadrats takes advantage of information on quadrat locations. A grid of contiguous quadrats is a spatial lattice (Figure 3). The advantage of such a grid is that neighbouring quadrats can be combined so that we may obtain information about quadrats of more than one size.

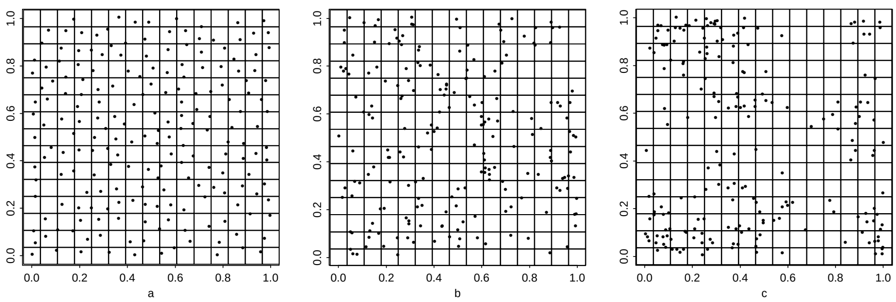

Figure 3. Spatial lattice of 15×15 contiguous quadrats superimposed over the point patterns.

A natural test of a Poisson distribution is therefore provided by examining the value of the ratio sample variance/sample mean or *index of dispersion* (ID). If we denote the observed counts in nq quadrats by $x_{1,}x_2,\ldots,x_{nq}$, then these counts have mean $\overline{x} = \sum x_i/nq$ and variance $s^2 = \dfrac{\sum(x_i-\overline{x})^2}{(nq-1)}$, where the summations are over the values of i from 1 to nq. Hoel (1943) showed that

$$ID = \frac{(nq-1)\,s^2}{\overline{x}} \tag{1}$$

has an approximate χ^2_{nq-1} distribution under CSR. This approximation is reasonable provided that $nq > 6$ and $\overline{x} > 1$. Values of $ID > \chi^2_{nq-1};(1-\alpha)$ are indicative of clustering and regularity is given by values of $ID < \chi^2_{nq-1};\alpha$, where α stands for the usual significance level.

Perry & Mead (1979) examined the behaviour of the index of dispersion test and concluded both, that it is remarkably sensitive at detecting a lack of homogeneity within a point pattern, and that the behaviour of the test was principally dependent upon the size of the mean of the quadrat counts: the larger the mean the more likely was the ID test to recognize any heterogeneity presented in the pattern.

A number of other indices have been suggested, principally for situations in which clusters may be present. All are sensitive to changes in quadrat sizes. David and Moore (1954) suggested that the quantity

$$ICS = \left(\frac{s^2}{\overline{x}}\right) - 1 \quad (2)$$

would provide an approximate index of clumping or contagiousness. This test is called *index of cluster size*. For a Poisson pattern, ICS has mean 0 and is independent of the quadrat size. An interpretation of a positive value for ICS is as the number of other events intimately associated with a randomly chosen event. A negative value for ICS indicates some regularity in the positioning of the events.

Theoretically, if we have the observed iid counts $x_{1,}x_2,\ldots,x_{nq}$ in nq quadrats following any distribution, then ICS satisfies (Serfling, 1980)

$$ICS \sim N\left(\frac{\sigma^2}{\mu} - 1, \frac{1}{n}\left(\frac{\sigma^6}{\mu^4} + \frac{\mu_4 - \sigma^4}{\mu^2} - \frac{2\mu_3\sigma^2}{\mu^3}\right)\right) \quad (3)$$

where $E(x_i) = \mu$, $Var(x_i) = \sigma^2$, $E(\{x_i - \mu\}^3) = \mu_3$ and $E(\{x_i - \mu\}^4) = \mu_4$.

Particularly, if the counts come from a Poisson distribution, i.e., if we have a CSR point pattern, then for large nq we have that

$$ICS \sim N(0, 2/n) \quad (4)$$

However, for small nq a better approximation can be found for ICS which consists of

$$ICS \sim \frac{\chi^2_{nq-1}}{nq-1} - 1 \quad (5)$$

Note that for large values of nq, we can approximate $\chi^2_{nq-1} \sim N(nq-1, 2(nq-1))$.

A number of other indices have been suggested, principally for situations where it is thought that clusters may be present. All are sensitive to changes in quadrat sizes. However, for quadrat count data, ID appears to have no serious rivals as a test statistic (Diggle, 1983). Cormack (1979) notes that alternative indices proposed by Morisita (1959) and Lloyd (1967) need to be converted to ID in order to test CSR.

It was first Greig-Smith (1952) who proposed contiguous quadrats to analyze data presented as counts by means of the index of dispersion. He suggested that a 16×16 grid of quadrats should be used. Then, Diggle (1983) used the same index of dispersion within a regular $k \times k$ grid of contiguous square sub-regions of equal area to test CSR. Particularly, significantly small values of ID indicate a tendency towards a regular spatial distribution of events, whereas significantly large values indicate aggregation.

3. SIMULATION STUDY

We present here an extensive simulation study of the two selected count-based indices, ID and ICS based on contiguous quadrats superimposed over the selected region. To consider any possible alternatives to the random (CSR) pattern, we also consider regular and aggregated point patterns. CSR patterns are generated according to a homogeneous Poisson process following the definition given in the introduction.

Regular or inhibitory point patterns are defined through a hard-core distance, δ, using a sequential inhibition process based on the following facts: a) x_1 is uniformly distributed in the region A; b) Given $\{x_j, j = 1,\ldots,i-1\}$, x_j is uniformly distributed on the intersection of A with $\{y : d(y,x_j) \geq \delta,\ j = 1,\ldots,i-1\}$.

Aggregated patterns are defined in terms of Poisson clustered processes as defined by Neyman & Scott (1958). These processes incorporate an explicit form of spatial clustering, and as such provide a more satisfactory basis for the modelling of aggregated spatial point patterns. They are defined based on the following three postulates: a) Parent events form a Poisson process with a fixed intensity; b) Each parent produces a random number of offsprings, realized independently and identically for each parent; c) The positions of the offsprings relative to their parents are independently and identically distributed according to a bivariate normal density function.

The S.P.P.A. computer software has been developed to generate planar coordinates of points in a fixed region with a particular spatial structure and then, among other things, calculate quadrat-based counts for a given quadrat size.

3.1. Design of the simulation study

The process of simulations has been carried out for three qualitatively different spatial structures: randomness, clustering and regularity. We have used several total number of points per pattern: for random and regular structures, $n = 400, 1000$ and for clustered structures, $n = 400, 1000, 2500$. Clustered pattern simulation is done with several number of fathers, ranging from 1 to 4. Two hard-core radius for inhibitory patterns have been used, $ir = 0{,}04295$ (with $n = 400$), and $ir = 0{,}02688$ (with $n = 1000$). For each combination of selected quantities, we simulated $r = 2000$ realizations in the unit square, $(0,1) \times (0,1)$.

For each pattern, and to apply the contiguous quadrat system, we have used several grid sizes shown in Table 1. Each line of Table 1 gives us information about the type of simulated pattern (*Pattern*), the number of points in each pattern (*Points*), the number of replications for each experiment (*Repl.*), the inhibition value used in the regular pattern simulations (*Inhib.*), and grid order values. All grid orders are magnitudes to the square, i.e., ($10 \times 10, 15 \times 15, 20 \times 20$, etc), but they are shown simplified ($10, 15, 20$, etc).

Table 1. Design of the simulation study.

Pattern	**Points**	**Repl.**	**Inhib.**	**Grid order**
Random	400	2000		2,3,...,6,8,10,11,13,...,19,20,21,23,25,30,...,100
Random	1000	2000		2,3,...,6,8,10,11,13,...,19,20,21,23,25,30,40,...,100
Regular	400	2000	.04295	5,6,8,10,11,13,...,19,20,21,23,25
Regular	1000	2000	.02688	5,6,8,10,11,13,...,19,20,21,23,25,30,35,40
1 Cluster	400	2000		4,5,6,8,10,11,13,...,19,20,21,23,25,30,35,40
1 Cluster	1000	2000		6,8,10,11,13,...,19,...,30,35,40,50,...,100,120,...,200
1 Cluster	2500	2000		10,20,30,...,100,120,140,...,200
2 Cluster	1000	2000		10,20,30,...,100,120,140,...,200
2 Cluster	2500	2000		10,20,30,...,100,120,140,...,200
3 Cluster	1000	2000		10,20,30,...,100,120,140,...,200
3 Cluster	2500	2000		10,20,30,...,100,120,140,...,200
4 Cluster	1000	2000		10,20,30,...,100,120,140,...,200
4 Cluster	2500	2000		10,20,30,...,100,120,140,...,200

Each combination of pattern, replicate, grid size, inhibition radius and number of clusters yielded 2000 estimates of both indices which are summarised by box-plots and tables.

Particularly interesting is to show the goodness of fit of the approximation given in (5) for CSR patterns with a small number of nq quadrats. As we are interested in analyzing values belonging to the tails of the chi-square distribution to safely contrast overdispersion or underdispersion, we focus upon the following percentiles, $\alpha = 1\%, 5\%, 10\%$ and $\alpha = 99\%, 95\%, 90\%$. For each percentile the *relative error* is calculated as

(6) $$re = \frac{100\,|\alpha - p|}{\alpha}$$

where p is the observed proportion of simulated values of ICS under the theoretical value obtained through (5) for $\alpha = 0{,}01, 0{,}05$ and $0{,}10$ and over the theoretical value from (5) for $\alpha = 0{,}99, 0{,}95$ and $0{,}90$.

3.2. Analysis of ID index

Table 2 shows the summarized results of the simulations with the ID index. Each entry represents the mean value of 2000 simulations. The following results are observed.

Table 2. Means of ID index. Each column entry indicates grid size and pattern structure (Ran=random, Reg=regular, c=cluster) followed by the number of points in the unit square. Blank boxes indicate that the corresponding simulation has not been analyzed.

Grid	Ran400	Ran1000	Reg400	Reg1000	1c400	1c1000	1c2500	2c1000	2c2500	3c1000	3c2500	4c1000	4c2500
2	3	3											
3	8	8											
4	15	15			257								
5	24	24	4	3	274								
6	35	35	6	5	290	677							
8	62	63	12	10	323	716							
10	99	99	21	16	363	759	1752	1390	3349	1534	3699	1031	
11	119	120	27	21	384	782							
13	167	168	44	32	433	832							
15	223	224	65	46	490	890							
17	287	288	86	62	553	955							
19	359	359	109	85	626	1028							
20	398	399	121	100	665	1067	2068	1768	3833	1981	4369	1457	3065
21	438	440	136	116	707	1107							
23	528	527	178	150	796	1197							
25	622	624	246	183	891	1293							
30	898	899		270	1167	1569	2576	2281	4369	2508	4935	1988	3628
35	1224	1224		394	1490	1894							
40	1598	1597		648	1867	2271	3275	2987	5078	3220	5658	2689	4349
45	2023												
50	2497	2499				3173	4180	3890	5985	4122	6567	3594	5258
55	3022												
60	3597	3602				4273	5273	4990	7087	5234	7685	4697	6366
65	4221												
70	4898	4902				5572	6579	6297	8391	6530	8985	5999	7664
75	5626												
80	6397	6398				7067	8077	7791	9889	8034	10481	7497	9174
85	7224												
90	8097	8102				8771	9779	9496	11593	9728	12186	9202	10875
95	9022												
100	9996	10005				10675	11682	11392	13497	11629	14096	11107	12769
120						15071	16077	15797	17895	16035	18491	15509	17177
140						20278	21281	20998	23095	21234	23704	20711	22377
160						26269	27287	26989	29095	27235	29685	26708	28375
180						33072	34079	33806	35891	34033	36502	33503	35173
200						40663	41696	41392	43501	41630	44067	41121	42772

3.2.1. *CSR point pattern distribution*

- In patterns with 400 points, ID follows a perfect χ^2_{nq-1} distribution (pval= 0.98) for grid sizes of 2×2 up to 35×35. For bigger grid sizes, ID still follows a chi-square distribution (pval= 0.39) but with a slightly smaller mean (Figure 4). Then, theory is generally satisfied except for very large grid sizes.

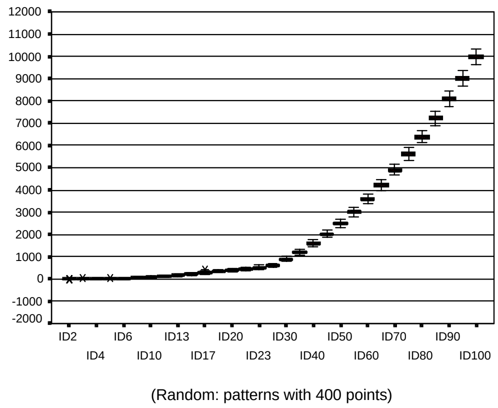

Figure 4. Boxplots of ID values under CSR pattern structure with 400 points.

- Again, in patterns with 1000 points, ID follows a perfect χ^2_{nq-1} distribution (pval= 0.97) for grid sizes from 2×2 up to 70×70. This distribution is still observed for bigger grid sizes (pval= 0.57) but with some little changes in the mean value (Figure 5). Generally, there seems to be no clear distinction in the ID performance between CSR patterns with different number of points.

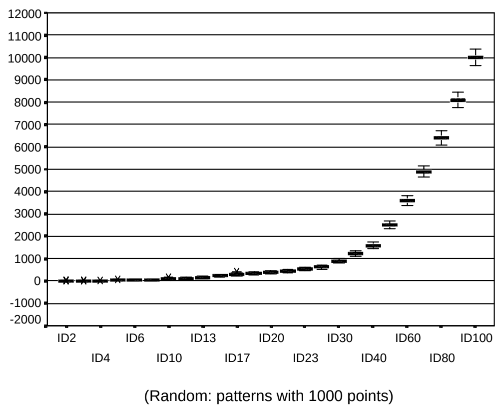

Figure 5. Boxplots of ID values under CSR pattern structure with 1000 points.

3.2.2. Regular point pattern distribution

- For patterns with 400 points, we observe in any case that $ID < \chi^2_{nq-1}; \alpha = 0{,}01$ (Figure 6). The magnitude of the ID index is increased with the grid size, though they are significantly smaller (pval= 0.01) compared to the corresponding ID value under CSR structure. The chi-square distribution is no longer satisfied.

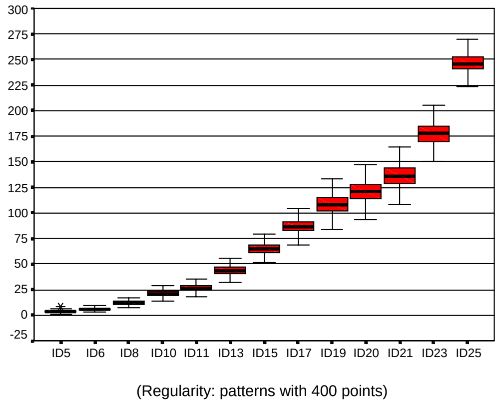

Figure 6. Boxplots of ID values under regular pattern structure with 400 points and inhibition radius $ir = 0{,}04295$

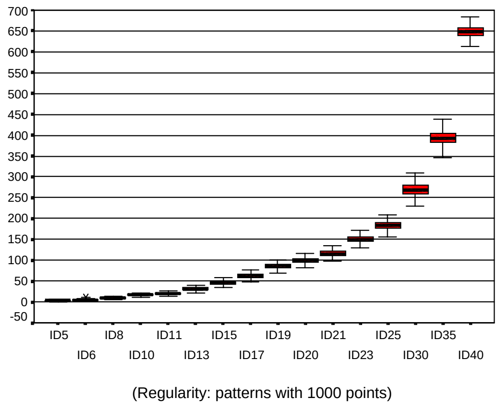

Figure 7. Boxplots of ID values under regular pattern structure with 1000 points and inhibition radius $ir = 0{,}02688$.

- For patterns with 1000 points, again it is generally observed that $ID < \chi^2_{nq-1}; \alpha = 0{,}01$ (Figure 7). ID values increase with the grid size though still significantly smaller (pval= 0.01) than the corresponding values under CSR condition. It is also observed that for equal grid size, ID values with 1000 points are smaller than ID values for patterns with 400 points. This is clearly due to the fact that we are using different values for the inhibition radius and this is detected by the ID index.

3.2.3. Cluster point pattern distribution

In this section we comment the results for the four different cluster pattern structures, though, for shortness, we only show the corresponding boxplots for 1 cluster patterns (see Figures 8, 9 and 10).

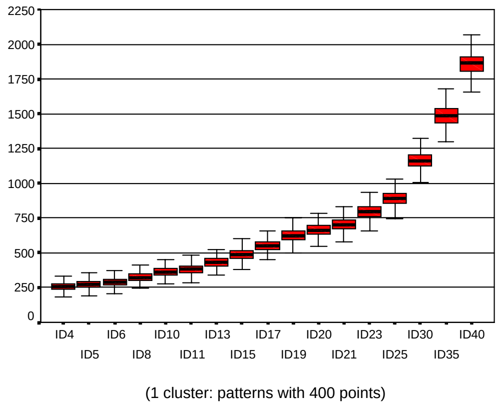

Figure 8. Boxplots of ID values under 1 cluster pattern structure with 400 points.

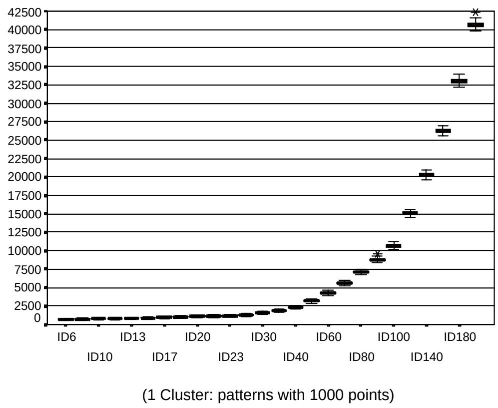

Figure 9. Boxplots of ID values under 1 cluster pattern structure with 1000 points.

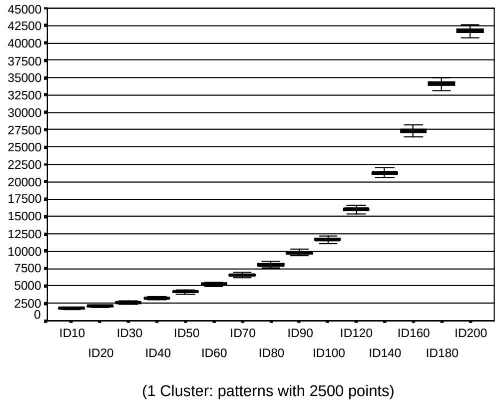

Figure 10. Boxplots of ID values under 1 cluster pattern structure with 2500 points.

- In any case, we observe that $ID > \chi^2_{nq-1}; (1-\alpha = 0{,}99)$ and its value increases with the grid size. ID index is a very sensitive index to detect clustering in that not only the ratio standard deviation/mean decreases as the grid size increases, but also ID increases with the number of clusters, except for the four cluster case. In fact, the pattern with four clusters may appear to be less aggregated that those patterns with a smaller number of clusters. Then, ID index is good in detecting scales of aggregation in point patterns.
- Within the same scale of aggregation, ID values show less variability in those patterns with a larger number of points.
- The statistical distribution of ID under clustering depends on the grid size and the number of points. The chi-squared distribution is always rejected though ID values seem to follow a gaussian distribution in several grid sizes. This behaviour is independent of the scale of aggregation.
- Note that grid size depends on the number of points in the pattern. Using a large grid size so that most of our cells or quadrats have zero counts causes the results to be biased. This comment is also clearly true for any other kind of pattern structure.

3.3. Analysis of ICS index

Table 3 presents the summarized results of the corresponding simulations with the ICS index. Each entry in the table represents the mean value of 2000 simulations. The following results are obtained.

Table 3. Means of ICS index. Each column entry indicates grid size and pattern structure (Ran=random, Reg=regular, c=cluster) followed by the number of points in the unit square. Blank boxes indicate that the corresponding simulation has not been analyzed.

Grid	Ran400	Ran1000	Reg400	Reg1000	1c400	1c1000	1c2500	2c1000	2c2500	3c1000	3c2500	4c1000	4c2500
2	-0,01509	0,0250											
3	-0,02333	-0,0062											
4	0,00239	-0,0003			16,11								
5	-0,00448	0,0005	-0,843	-0,867	10,44								
6	-0,00608	-0,0073	-0,832	-0,862	7,30	18,34							
8	-0,01696	0,0012	-0,807	-0,848	4,13	10,37							
10	0,00039	0,0030	-0,788	-0,835	2,66	6,67	16,697	13,0392	32,8244	14,493	36,361	9,4166	
11	-0,00718	0,0005	-0,778	-0,827	2,20	5,52							
13	-0,00426	0,0021	-0,739	-0,812	1,58	3,95							
15	-0,00281	0,0002	-0,709	-0,796	1,19	2,97							
17	-0,00192	-0,0011	-0,700	-0,784	0,92	2,32							
19	-0,00201	-0,0026	-0,698	-0,763	0,74	1,86							
20	-0,00217	0,0004	-0,696	-0,749	0,67	1,67	4,1821	3,4315	8,60691	3,9654	9,9505	2,6509	6,68246
21	-0,00439	-0,0002	-0,690	-0,736	0,61	1,52							
23	-0,00092	-0,0011	-0,664	-0,717	0,51	1,27							
25	-0,00273	0,0001	-0,605	-0,707	0,43	1,07							
30	-0,00069	-0,0002		-0,700	0,30	0,75	1,8657	1,5372	3,85966	1,7893	4,4896	1,2118	3,03538
35	-0,00038	-0,0001		-0,678	0,22	0,55							
40	-0,00078	-0,0010		-0,595	0,17	0,42	1,0482	0,8679	2,17584	1,0139	2,5382	0,6819	1,71952
45	-0,00033												
50	-0,0009	0,0001				0,270	0,6725	0,5567	1,39478	0,6494	1,6277	0,4381	1,10421
55	-0,00052												
60	-0,00045	0,0009				0,187	0,4652	0,3865	0,96915	0,4542	1,1353	0,3051	0,76895
65	-0,0006												
70	-0,00011	0,0005				0,137	0,343	0,2854	0,71275	0,3329	0,834	0,2245	0,56445
75	0,00028												
80	-0,00028	-0,0001				0,105	0,2623	0,2176	0,54534	0,2556	0,6379	0,1717	0,43366
85	-8,9E-06												
90	-0,00026	0,0003				0,083	0,2074	0,1725	0,43137	0,2012	0,5047	0,1362	0,34277
95	-0,00018												
100	-0,00032	0,0006				0,068	0,1683	0,1394	0,34981	0,163	0,4097	0,1108	0,27705
120						0,0467	0,1165	0,09711	0,2428	0,1136	0,2842	0,0771	0,19292
140						0,0346	0,0858	0,07139	0,17839	0,0834	0,2095	0,0567	0,14175
160						0,0262	0,0659	0,0543	0,13656	0,0639	0,1596	0,0433	0,10846
180						0,0208	0,0519	0,04342	0,10779	0,0504	0,1267	0,0341	0,08562
200						0,0166	0,0424	0,03483	0,08755	0,0408	0,1017	0,0281	0,06933

3.3.1. CSR point pattern distribution

- For CSR patterns with 400 points, ICS values oscillate around zero. However, for small to medium grid sizes, there exist quite large standard deviations and also some outliers are present (Figure 11). Standard deviations rapidly decrease when grid size increases, which is natural in terms of the theoretical expression of the variance in formula (4) for large samples. The results show that for small samples, the chi-square distribution given by (5) is satisfied and also ICS values follow the gaussian distribution (4) when large samples are considered.

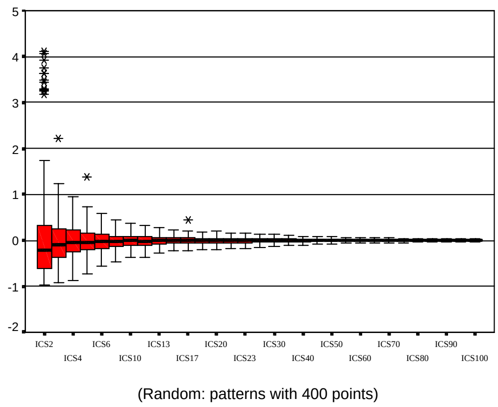

Figure 11. Boxplots of ICS values under CSR pattern structure with 400 points.

- Generally, the same results can be found for CSR patterns with 1000 points. The difference is that the larger the number of points, we need smaller grid sizes to get the same or smaller standard deviations (Figure 12).

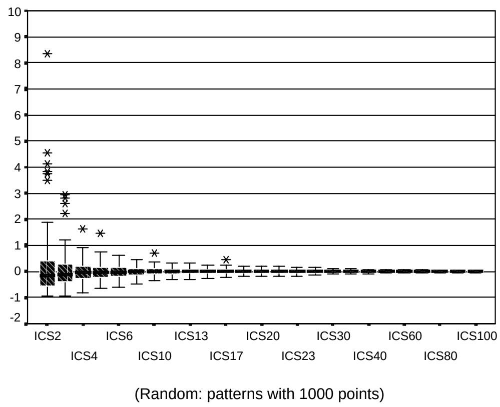

Figure 12. Boxplots of ICS values under CSR pattern structure with 1000 points.

- Concerning the evaluation of the goodness of fit of the approximation given in (5) for CSR patterns with a small number of nq quadrats in terms of the relative error, we calculated the relative errors following (6) for six percentiles, $\alpha = 0{,}01, 0{,}05, 0{,}10, 0{,}90, 0{,}95, 0{,}99$, and for several grid sizes up to $nq = 100$ (10×10). The results are shown in Table 4. In general, and independently of the number of points per pattern, the relative errors are below 20%, which can be considered as small enough to trust on the results. However, we find relative errors bigger that 20% at the very end of the tails, i.e., for $\alpha = 0{,}01$ and $0{,}99$, which means that under very severe conditions the results of a CSR contrast might be misleading.

Table 4. Relative errors of the ICS index under CSR pattern structure for several percentiles and for patterns with $n = 400$ points and with $n = 1000$ points (in parenthesis) in the unit square.

		grid size						
		2×2	3×3	4×4	5×5	6×6	8×8	10×10
α	0.99	5 (30)	40 (10)	20 (20)	25 (5)	25 (10)	15 (15)	20 (0)
	0.95	5 (3)	12 (3)	8 (1)	4 (6)	4 (19)	14 (0)	7 (12)
	0.90	1.5 (3)	8.5 (3)	3 (3)	5.5 (0)	10 (16)	19 (1.5)	5 (4.5)
	0.10	1.5 (15)	16 (1.5)	4.5 (3.5)	5.5 (9)	9.5 (2)	16.5 (4.5)	5.5 (15)
	0.05	6 (30)	22 (1)	12 (1)	1 (9)	16 (10)	1 (4)	15 (10)
	0.01	40 (95)	5 (40)	30 (15)	25 (0)	0 (45)	10 (55)	25 (15)

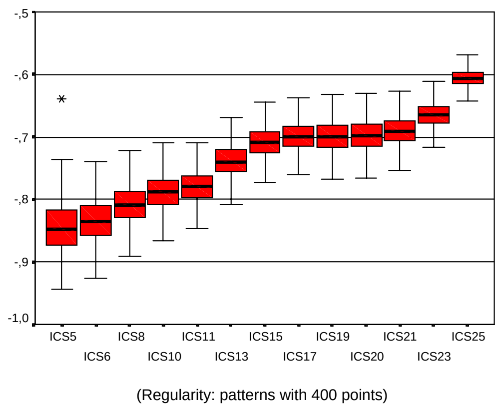

Figure 13. Boxplots of ICS values under regular pattern structure with 400 points and inhibition radius $ir = 0{,}04295$.

3.3.2. Regular point pattern distribution

- We find that ICS values for regular patterns are significantly smaller than those under CSR (pval= 0.01), showing always clearly negative values. However, there seems to be an increasing tendency towards zero as the grid size increases (Figures 13 and 14). This shows that there should be an optimum grid size, for example around 20×20 in patterns with 400 points or 30×30 when the number of points is 1000. If we surpass them, then ICS values are biased. The chi-squared distribution is no longer satisfied in favour of the Gaussian distribution which is only observed for those grid sizes near the optimum (Figures 13 and 14).

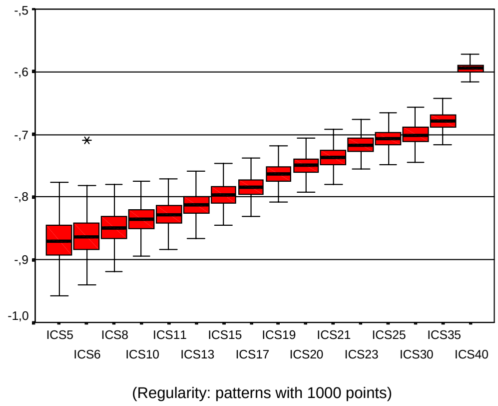

Figure 14. Boxplots of ICS values under regular pattern structure with 1000 points and inhibition radius $ir = 0{,}02688$.

3.3.3. Cluster point pattern distribution

Again we comment the results for all cluster structures though, for shortness, we only show those boxplots corresponding to one cluster with 400, 1000 and 2500 points (Figures 15, 16 and 17).

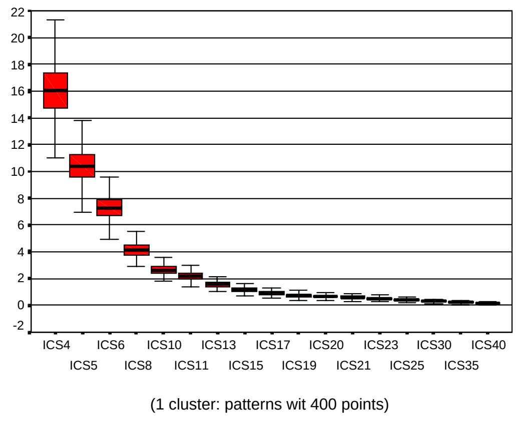

Figure 15. Boxplots of ICS values under 1 cluster pattern structure with 400 points.

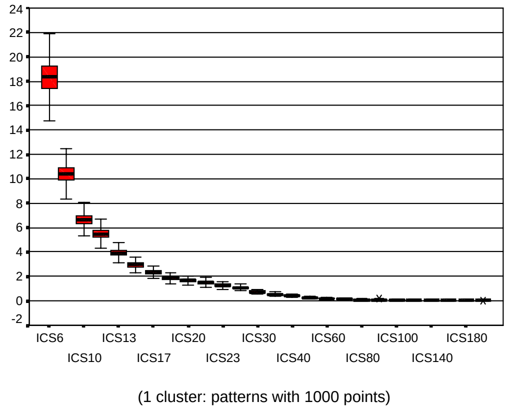

Figure 16. Boxplots of ICS values under 1 cluster pattern structure with 1000 points.

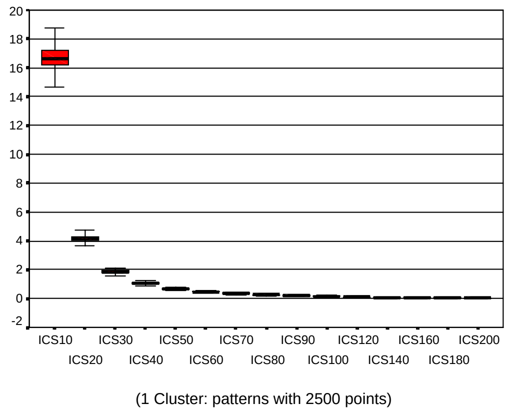

Figure 17. Boxplots of ICS values under 1 cluster pattern structure with 2500 points.

Generally speaking, under cluster spatial structures, ICS takes positive values, significantly different from zero (pval= 0.01, which is characteristic of ICS under CSR patterns). However, ICS tends to decrease to zero as the grid size increases. The relative standard deviation of this indicator (standard deviation/mean) is increased when the grid size increases.

We have also found that ICS index is very sensitive to both, the scale of aggregation (given by the number of clusters) and the number of points. In fact, the mean values of the ICS index for patterns with 1000 points can be obtained by multiplying those obtained with 400 points times 2.5. This is equally true if we compare patterns with 400 points with those patterns with 2500 points. In this case we have to multiply times 6.25 the corresponding values of ICS.

- The distribution of ICS under clustering is somewhat complicated and depends on the number of points and the degree of aggregation. For example, when we consider patterns with one cluster, gaussianity is satisfied only within a few grid sizes (from 4×4 to 13×13) for patterns with 400 points. The number of grid sizes in which the gaussian assumption is satisfied, increases when the number of points increases (from 6×6 to 60×60, for n=1000, and in most of grid sizes for n=2500).
- For those patterns with 2 clusters, values of ICS are more indicative of spatial clustering compared to patterns with only one cluster. The gaussian behaviour of ICS is similar as commented above.
- In general, ICS values clearly indicate the degree of clustering by taking larger values. The gaussian distribution is comfortable reached when the number of points is large.

3.4. General conclusions

After analyzing step by step the results obtained in the simulation study, the following general results can be outlined.

- The results confirm that the *index of dispersion* follows an approximate χ^2_{nq-1} distribution under CSR when $nq > 6$ and $\bar{x} > 1$. Moreover, we can enlarged this condition to $nq < 6$ (grid 2×2) and $\bar{x} < 1$. Equally, our results show that values of $ID > \chi^2_{nq-1;(1-\alpha)}$ are indicative of clustering and regularity is given by values of $ID < \chi^2_{nq-1;\alpha}$ (see Figures 18 and 19). Therefore, we confirm by simulation Hoel's (Hoel, 1943) and Diggle's results (Diggle, 1983).

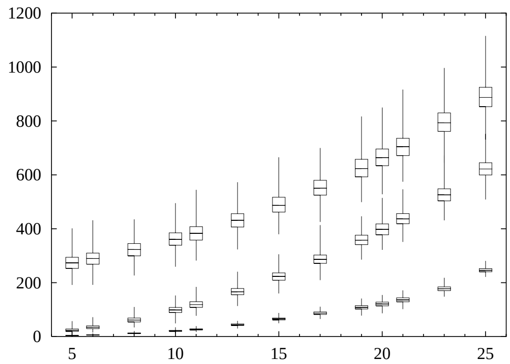

Figure 18. Comparison of boxplots for the ID index under the three pattern structures (1 cluster, CSR and regularity) for patterns with n=400 points in the unit square.

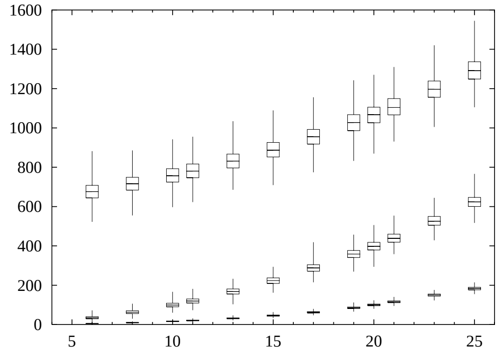

Figure 19. Comparison of boxplots for the ID index under the three pattern structures (1 cluster, CSR and regularity) for patterns with n=1000 points in the unit square.

- Similarly, we have also confirmed Douglas theory adapted for contiguous quadrats (Douglas, 1975) for the *index of cluster size*: for a Poisson pattern, ICS has mean 0 and is independent of the quadrat size; ICS has a positive mean for clustered patterns and a negative mean for regular patterns (see Figures 20 and 21).
- We have checked the distribution of both indices under alternative spatial patterns. ID index is no longer chi-squared distributed if CSR is rejected in favour to regular or aggregated patterns. However, under these alternatives, it seems that ID follows a gaussian distribution, particularly when the number of points is large enough (at least 1000 points). On the other hand, ICS index follows a chi-squared distribution under small grid sizes and a gaussian distribution for bigger grid sizes as theory states.

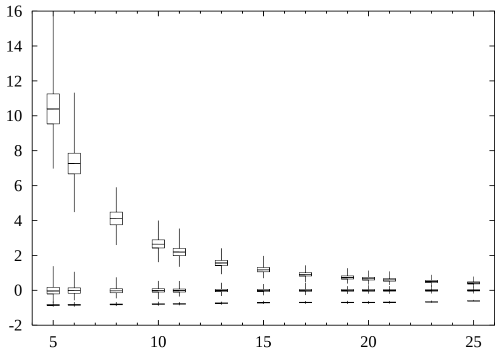

Figure 20. Comparison of boxplots for the ICS index under the three pattern structures (1 cluster, CSR and regularity) for patterns with n=400 points in the unit square.

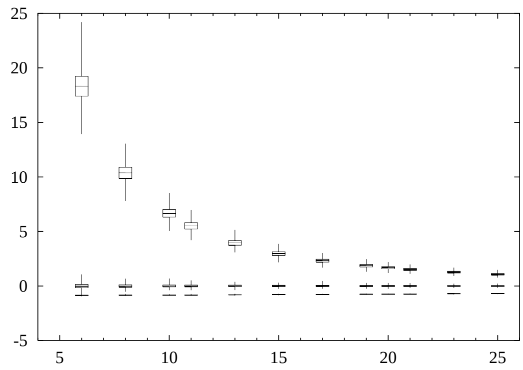

Figure 21. Comparison of boxplots for the ICS index under the three pattern structures (1 cluster, CSR and regularity) for patterns with n=1000 points in the unit square.

- Both indices are very sensitive at detecting not only the degree of clustering but also the number of points. The latter element is crucial to get lower standard deviation in the index values.
- To detect randomness, the larger the grid size, the better indication of CSR pattern (in terms of bias and standard deviation) we get from both indices. This is not generally true when detecting regularity or aggregation. There seems to be an optimum grid size, from which onwards the results are clearly biased, even confusing. The optimum grid size depends on the number of points. Generally speaking, and according to our simulations, the number of contiguous quadrats should not exceed the total number of points. A low grid order means that we may have insufficient information. But also a too large grid order means that we are introducing irrelevant information to our data structure, introducing bias to the estimates. This was not previously found in literature.

4. APPLICATION

It is known that information about the spatial distribution of the population may give us insights of interesting economic phenomena (Richardson, 1986; Hudson & Fowler, 1966; Lösch, 1954). This is due to the fact that human settlements have a history conditioned by economy which has been developed in specific geographical areas. Consequently, we present here a practical use of both, the ID and ICS indices, to describe qualitative and quantitatively the spatial structure of two important spanish peninsular provinces, Madrid and Barcelona as this analysis will provide economists and geographers with relevant information.

4.1. Data and results

The data set represents coordinates, longitude (eastings) and latitude (northings) of points, where each point defines the location of a 20,000 inhabitants crowd. The data set was transformed adequately to have planar coordinates. The origin of the coordinate axis is set up as the point of the meridian at 9^o-west longitude and the parallel at 36^o-north latitude. Let each measurement unit be equal to 1500 metres (approximately). Then, we assign p_i points to $i-th$ city as follows

$$p_i = \frac{h_i}{20000}, \tag{7}$$

where h_i denotes the number of inhabitants in each city, obtained from INE (1994), and the coordinates from Dirección General I.G.N. (1994). The spatial locations of Madrid and Barcelona are shown in Figure 22.

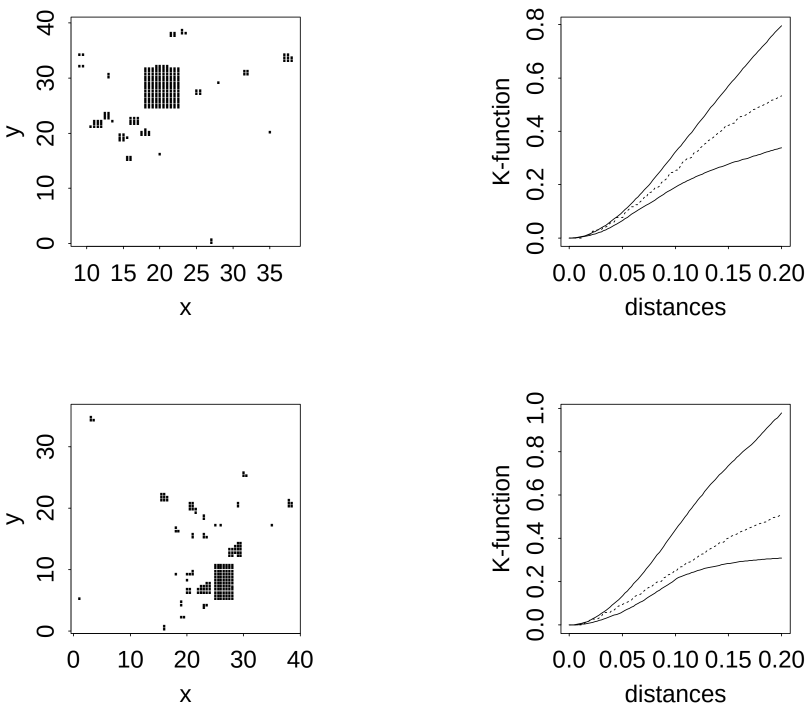

Figure 22. *Upper row*: Spatial locations of Madrid and empirical (dotted line) and confidence intervals (solid lines) of the K-function under a Neyman-Scott process with two clusters. *Lower row*: Spatial locations of Barcelona and empirical (dotted line) and confidence intervals (solid lines) of the K-function under a Neyman-Scott process with two clusters.

In Tables 4 and 5 the results are shown for the two selected provinces and compared to the total spanish peninsular territory. The column entries give us the number of points included in each territorial pattern, and the ID and ICS values for the following grid orders 10×10, 20×20 and 30×30.

Table 4. ID and ICS values for two spanish provinces

	Points	**ID_{10}**	**ID_{20}**	**ID_{30}**	**ICS_{10}**	**ICS_{20}**	**ICS_{30}**
Spain	1183	8393	27491	59763	83,78	67,90	65,48
Barcelona	185	1961	3020	3367	18,81	6,57	2,74
Madrid	235	2587	3843	4974	25,14	8,63	4,53

Table 5. Quantitative comparison of ID and ICS values for two spanish provinces

	Points	**ICS_{10}**	**ICS_{20}**
Barcelona	185	18,81	6,57
Madrid	$\frac{235}{1{,}2702703} = 185$	19,79	6,79

The indices values indicate that we have demographic structures characterized by cluster spatial processes (pval= 0.01 for both indices and grid order when testing CSR). But if we want to use more quantitatively the indices information, let us limit ourselves to compare the provinces of Barcelona and Madrid, as they are set on surfaces of (approximately) the same magnitude. Keeping in mind the total number of points in each pattern structure (185 and 235), we concentrate on grid orders of 10×10 and 20×20, as 30×30 will add irrelevant information as commented previously on the paper. Also, though both indices are sensitive at detecting the spatial structure depending on the number of points and degree of clustering, if there exists, the ICS values present the interesting characteristic of picking up the proportion among the number of points of different patterns within the same cluster degree. Due to this characteristic we have considered more appropriate to use the ICS index in order to make the comparisons in quantitative terms.

In order to remove from Madrid ICS index the component due to the increment in the number of points with regard to Barcelona, the values of the ICS index are divided by 1,27 (note that 235/185=1,2702703). As a result (see Table 5) we can conclude that the demographic spatial structure of Madrid province presents bigger intensity and bigger cluster degree than that of Barcelona.

Knowing that both, Madrid and Barcelona spatial patterns are clustered spatial structures, we tried to fit Neyman-Scott processes (as defined in section 3) to both patterns. As a result, Madrid corresponds to a clustered pattern with two parents (pval= 0.89), one with 155 offsprings and a dispersion from the parent of 0.06 and other parent with 80 offsprings dispersed 0.10 from the parent. On the other hand, Barcelona corresponds to a clustered pattern with again two parents (pval= 0.83), one with 90 offsprings and a dispersion parameter of 0.06 and other parent with 100 offsprings and a dispersion parameter of 0.068.

The goodness of fit of both processes (see Figure 22) has been measured by means of the K-function, a second-order property defined as (Diggle, 1983)

$$K(t) = \lambda^{-1} E(NFE(t))) \tag{8}$$

where λ stands for the first-order intensity function and NFE(t) represents the number of further events within distance t of an arbitrary event (Diggle, 1983).

Then we have showed an application of the use of spatial indices in detecting a pattern structure and also in making quantitative comparisons between point patterns.

ACKNOWLEDGEMENTS

The referees are gratefully acknowledged for their helpful comments that have substantially improved an earlier version of the paper.

REFERENCES

Cormack, [1979]. «Spatial aspects of competition between individuals». In *Spatial and Temporal Analysis in Ecology* (Cormack, R.M. and Ord, J.K., eds.), Internatiuonal Co-operative Publishing House, Maryland, 151-212.

Cox, D.R. & Isham, V. [1980]. *Point processes*, Chapman & Hall, London.

Cressie, N. [1993]. *Statistics for Spatial Data*. John Wiley & Sons, 2nd Edition, New York.

David, F.N. & Moore, P.G. [1954]. «Notes on contagious distributions in plant populations». *Annals of Botany of London*, 18, 47-53.

Diggle, P.J. [1979]. «Statistical methods for spatial point patterns in ecology». In *Spatial and Temporal Analysis in Ecology* (Cormack, R.M. and Ord, J.K., eds.), International Co-operative Publishing House, Maryland, 95-150.

Diggle, P.J. [1983]. *Statistical Analysis of Spatial Point Patterns*, Academic Press, London.

Dirección General del Instituto Geográfico Nacional [1994]. *Atlas Nacional de España*, MOPTMA.

Greig-Smith, P. [1952]. «The use of random and contiguous quadrats in the study of the structure of plant communities». *Annals of Botany*, 16, 293-316.

Gulmon, S.L. & Mooney, H.A. [1977]. «Spatial and temporal relationships between two desert shrubs *Atriplex hymnelytra* and *Tidestromia oblongifolia* in Death Valley, California». *Journal of Ecology*, 65, 831-838.

Hoel, P.G. [1943]. «On indices of dispersion». *Annals of Mathematical Statistics*, 14, 155-162.

Hopkins, B. [1965]. «Observations on savanna burning in the Olokemeji Forest Reserve». *Journal of Applied Ecology*, 2, 367-381.

Hudson, J.C. & Fowler, P.M. [1966]. *The concept of Pattern in Geography*. Department of Geography, University of Iowa, Discussion paper: Series 1.

INE, [1994]. *Anuario Estadístico de España de 1993, 1994*.

Lloyd, M. [1967]. «Mean crowding». *Journal of Animal Ecology*, 36, 1-30.

Lösch, A. [1954]. *The economics of location*. Yale University Press, New Haven.

Moellering, H. & Tobler, W.R. [1972]. «Geographical variances». *Geographical Analysis*, 4, 34-50.

Morisita, M. [1959]. «Measuring of the dispersion and analysis of distribution patterns. *Memoires of the Faculty of Science, Kyushu University*», *series E. Biology*, 2(4), 215-235.

Neyman,J. & Scott, E.L. [1958]. «Statistical approach to problems of cosmology (with discussion)». *Journal of the Royal Statistical Society*, B 20, 1-43.

Perry, J.N. & Mead, R. [1979]. «On the power of the index of dispersion test to detect spatial pattern». *Biometrics*, 35, 613-622.

Richardson, H.W. [1986]. *Economía regional y urbana*. Alianza Editorial, Madrid.

Ripley, B.D. [1981]. *Spatial Statistics*. Wiley, New York.

Ripley, B.D. [1988]. *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.

S.P.P.A. [1997]. *Spatial Point Pattern Analysis*. Computer Software developed by Albert, Albert, Mateu & Pernias, Universitat Jaume I ,Castellón.

Serfling, R.J. [1980]. *Approximation Theorems of Mathematical Statistics.* Wiley & Sons, New York.

Stiteler, W.M. & Patil, G.P. [1971]. «Variance to mean ratio and Morisita's index as measures of spatial pattern in ecological populations». In *Statistical Ecology*, Vol. 1, (Patil, Pielou and Waters, eds.), University Park: Pennsylvania State University Press, 423-459.

Stoyan, D., Kendall, W.S. & Mecke, J. [1995]. *Stochastic Geometry and its Applications*. Akademie-Verlag, 2nd Edition, Berlin.

Upton, G.J.G. & Fingleton, B. [1994]. *Spatial Data Analysis by Example*. **Vol. 1**, John Wiley & Sons, New York.