

# Spatial location patterns of Spanish manufacturing firms

José M. Albert, Marta R. Casanova and Vicente Orts\*

**Preliminary Version.** Please do not quote without permission.

## Abstract

In this paper, we evaluate the spatial location patterns of Spanish manufacturing firms and we assess the different tendencies to cluster in each industry. To do this, we use a distance-based method (Marcon and Puech, 2003, Duranton and Overman, 2005), more concretely the Ripley's  $K$  function, which allows us to treat space as continuous, measuring concentration by counting each firm's average number of neighbours within a circle of a given radius. In this way,  $K(r)$  function is used to describe characteristics of the point patterns at different geographical scales at the same time, letting us detect the statistical significance of departures from randomness.

Our approach incorporates an additional improvement. By means of the software employed, 'R', we can apply border corrections adequately in any irregular polygonal shape; therefore, we introduce a polygonal envelope, which allows us to improve the delimitation of our area of study, Spain, avoiding the nuisance of empty spaces.

We apply this method to Spanish manufacturing sectors at two-digit level and we realise that results depend on the benchmark employed. In fact, whether we use 'complete spatial randomness' as benchmark, every sector analysed presents significant concentration whatever the distance of the radius we consider. However, as is well known, this type of analysis is sensitive to the benchmark considered, thus we also use as benchmark the whole of manufacturing sectors. In this case, we find patterns of localization completely different to the previous ones; appearing dispersion in some sectors relative to all manufacturing firms. Therefore, by means of this method we can know characteristic features of the pattern of localization of every manufacturing sector, like whether concentration or dispersion exists, which is its intensity and at which distance, or geographical scale, we obtain its highest level.

**Keywords:** distance-based method, polygonal envelope, Ripley's  $K$  function, Spanish manufacturing firms, spatial location patterns.

**JEL classification:** C15, C40, C60, R12.

---

\* Department of Economics, Universitat Jaume I.  
Contact: Marta R. Casanova, Department of Economics, Universitat Jaume I, Avda. Sos Baynat s/n, 12071 Castellón, (mroig@eco.uji.es).

## 1. Introduction

Economic activity is heterogeneously distributed across space and this distribution does not take into account, necessarily, administrative frontiers or country boundaries. Some traditional examples of this industries' geographic concentration are high-tech firms in Silicon Valley, the auto industry in Detroit, the carpet industry in Dalton or, dating back, the textile industry in Lancashire. We as well find, in Spain, the tile industry in Castellón and the leather and footwear industry in Elche.

Heterogeneity of spatial distribution of activity can be caused by multiple and very different factors and substantial literature has focused on this topic. The causes of concentration of economic activity have been evolving along time and the propensity of firms to agglomerate in space has changed as regions have become more integrated. Marshall (1890) introduced the notions of three positive externalities because of the fact that firms were located next to other firms: benefits coming from localized information and knowledge spillovers, from a specialised local labour market, as well as lower costs coming from a closer supply and demand (backward and forward linkages associated with large local markets). The 'new economic geography', driven by Krugman<sup>1</sup>, stressed the role of historical accidents and "agglomerative forces", emphasizing the increasing returns to scale. Moreover, it is necessary to highlight that without increasing returns in individual firm's production, the paper of the linkages would not be the same, not being necessary to concentrate production in the largest market, but rather establish a firm in each market. New economic geography as well emphasized the importance of the role of trade costs towards the spatial concentration, the theoretic relationship between market integration and industrial concentration and introduced the concept of 'multiple equilibriums', as we can see in Fujita *et al.* (1999), Puga (1999, 2002) and Ottaviano and Puga (1998). Mobile factors, as firms or workers, locate for profiting from a higher productivity and this creates a positive feedback. Firms and workers go where productivity is high and tend to raise it, creating an uneven distribution of activity.<sup>2</sup>

Not only the determinants of geographic concentration worry the economists, but also *how to measure* this mentioned spatial distribution. Therefore, this paper is not going to analyse which are the determinants of concentration, or the relative importance of them, but is going to measure the spatial location patterns of Spanish manufacturing sectors.

---

<sup>1</sup> See, for example, Krugman (1991a, 1991b).

<sup>2</sup> For further details and to go into the forces that are changing the spatial distribution of activity, see Overman, Redding and Venables (2003) and Venables (1995, 2006).

Many are the employed techniques to measure geographic concentration in the literature, but the most usual indices are Herfindahl, Gini or Ellison and Glaeser (1997).

*Herfindahl index* is a measure of industry concentration equal to the sum of the squared market shares of the firms in the industry. The *Gini index* is a measure that studies whether distribution is or not concentrated, in other words, measures the uniformity of a distribution. It compares the geographic patterns of employment, income or wealth distribution for one industry and for the aggregate. This index cannot distinguish whether concentration of activity is due to a few huge firms in a specific area or due to many small firms in the same specific area. Krugman (1991a), Brülhart (2001) or Amiti (1997), among others, used this index to measure the intensity of economic activity spatial distribution. Finally, *Ellison and Glaeser (1997)* tried to solve this inconvenient comparing concentration resulting from a random and independent location of firms with the real geographic concentration of an industry, taking into account, this time, the size of firms. Thus, this index let us compare concentration between industries or concentration of a given industry in different countries and, many authors as Devereux *et al.* (2004), Rosenthal and Strange (2001), Maurel and Sédillot (1999) or Callejón (1997) used it to measure geographic concentration of activity in their respective countries, UK, US, France and Spain.

Revising the literature about the spatial distribution of activity in Spain, we realise that many other authors, apart from Callejón, are interested in this topic, emphasizing Viladecans (2001), Paluzie *et al.* (2004) and Alonso-Villar *et al.* (2001, 2003 and 2004), among others. Other studies, as Paluzie *et al.* (2001) and Tirado *et al.* (2002), try to analyse, as well, the determinants of the localization of the industrial activity in Spain. Should be known that most of this Spanish studies use the Gini index to measure geographic concentration of activity, but Alonso-Villar *et al.* (2001) introduce in their analysis the index proposed in Maurel and Sédillot (1999)<sup>3</sup> and Viladecans (2001) uses an econometric spatial index, called Moran's I statistic of spatial association.

These methods used until now, in Spain and in other countries, to measure geographic distribution of activity, have a common characteristic. Their conclusions may differ considerably according to the scale chosen; that is to say, empirical results remain, in some degree, inconclusive with patterns detected depending on the geographical scale considered. Thus, Viladecans (2001) uses more than one geographic level in her analysis<sup>4</sup>, the same as Alonso-Villar *et al.*

---

<sup>3</sup> Moreover, compares it with other two concentration indices to obtain results that are more robust.

<sup>4</sup> She uses municipal and provincial level, corresponding to NUTS 5 and 3.

(2001) do<sup>5</sup>, to assess which would be the administrative unit more suitable in every analysis.

Moreover, all these traditional concentration indices *treat space as discrete*. In fact, these indices restrict the spatial distribution just to *one scale*, analysing the distribution of activity over discrete geographic units, not having why to coincide, these ones, with the relevant scale from the economic point of view.

The main purpose of this paper is to avoid the inconvenience of *geographic scale*, as Marcon and Puech (2003) or Duranton and Overman (2005) did. In this way, we will *treat space as continuous* to obtain a proper analysis of spatial location patterns, instead of sticking to administrative scale data, arbitrary from the economic point of view. To do this, we must employ a specific method that satisfies these essential requirements, a method that avoids the scale problem, letting us know and compare the concentration intensity for every spatial scale. Thus, just a *distance-based method* possesses the necessary requirements to achieve our aim, being as well essential a suitable data to consider space as continuous. Therefore, we employ UTM<sup>6</sup> coordinates to situate firms, as a dot in the map, obviating administrative frontiers.

The distance-based method we are going to use to measure spatial distribution of activity in Spain is Ripley's *K* function<sup>7</sup> and presents great advantages compared to traditional concentration indices. In fact, by means of the method employed, we can know whether concentration exists, which is its intensity and at which distance we obtain its highest level. Besides, as already said in the above paragraph, this method treats space as continuous, being able to avoid the administrative scale problem.<sup>8</sup>

Finally, before showing how is structured our paper, we must emphasize other two aspects that differ our empirical analysis from others. First, our localization of firms is precise, because we know the UTM coordinates of every establishment. On the contrary, Marcon and Puech (2003) and Duranton and Overman (2005) know the postcode of each manufacturing firm and associate geographic coordinates for all postcodes, obtaining a location error between 100 m and about 2 km. Second, we incorporate an innovative technique to improve the delimitation of our area of study, substituting the rectangular shape, used by other authors, for a polygonal shape. In this case, Marcon and Puech (2003) limited their empirical analysis to rectangular areas because they affirmed that '*complexity depends on*

---

<sup>5</sup> They use provincial and regional level, corresponding to NUTS 3 and 2.

<sup>6</sup> Universal Transverse Mercator.

<sup>7</sup> For further details, see Ripley (1976, 1977, 1979).

<sup>8</sup> Marcon and Puech (2003) have previously used this concentration index, in France, to measure the spatial distribution of economic activity.

the shape of the area under study'. However, in our case, by means of the software employed, 'R'<sup>9</sup>, we can apply border corrections adequately in any irregular polygonal shape.

The paper is organised as follows. In Section 2, we outline the methodology employed, Ripley's  $K$  function. In Section 3, we describe our data and the area of study used. In Section 4, we present the main results achieved, discuss them and outline the key improvements incorporated into the measuring method of spatial concentration used. Finally, Section 5 contains the main conclusions reached.

## 2. Methodology

The Ripley's  $K$  function is a distance-based method that measures concentration by counting each firm's average number of neighbours within a circle of a given radius, calling 'neighbours' all the points situated at a distance equal or lower than the radius ( $r$ ).

$K(r)$  function describes characteristics of the point patterns at many and different scales, depending the value of " $r$ " we take into account.

$$K(r) = \frac{1}{\lambda N} \sum_{i=1}^N \sum_{j=1, i \neq j}^N I(d_{ij})$$

Being  $r$  the different values of the radius we consider,  $d_{ij}$  the distance between two firms and  $N$  the total number of points observed in the area of the study region. Moreover, as we have said before, a point is considered 'neighbour' if the distance between  $i^{th}$  and  $j^{th}$  points is lower than  $r$ ; then, the indicator function  $I(d_{ij})$  takes value 1. Whereas if this distance, ( $d_{ij}$ ), is higher than  $r$  the point is not considered neighbour and  $I(d_{ij})$  takes value 0.

This function can be defined as the average number of neighbours in a radius ( $r$ ), divided by the density ( $\lambda$ ); which is defined as follows, being  $A$  the area of the study region.

$$\lambda = \frac{N}{A}$$

If we substitute the value of  $\lambda$  in the previous  $K(r)$  function, we obtain the definitive expression of  $K(r)$  value.

$$K(r) = \frac{A}{N^2} \sum_{i=1}^N \sum_{j=1, i \neq j}^N I(d_{ij})$$

With reference to the *benchmark*, we are going to analyse two scenarios. On the one hand, in first scenario we use *Complete Spatial Randomness* (CSR) as a

---

<sup>9</sup> This software is downloadable from the following website: <http://www.r-project.org/>.

benchmark; however, we realise that it is neither optimal nor realistic because economic activity does not locate in a random and independent way. Economic activities are spatially concentrated because of dissimilarities in such natural features as mountains, rivers or harbours. In this way, CSR as a benchmark does not take into account the first nature or the second nature, understanding second nature as the endogenous mechanisms leading to agglomeration.

Therefore, our first analysis consists in calculating the  $K$  function of the real point pattern (empirical  $K$  value) and comparing it with the  $K$  function of a Poisson point pattern with the same intensity, or number of points, (theoretical  $K$  value).<sup>10</sup> The theoretical value of  $K(r)$  will be  $\pi r^2$  as long as we assume CSR. In this way, we define  $M_{CSR}(r)$  as the value that quantifies the difference between the empirical  $K$  value and the theoretical  $K$  value, and is characterized as follows,

$$M_{CSR}(r) = K(r) - \pi r^2$$

Whether the empirical  $K$  value,  $K(r)$ , is higher than the theoretical  $K$  value,  $\pi r^2$ , this indicates concentration of our point pattern distribution, since the real density is greater than the benchmark's, in the aggregates. Lower values indicate dispersion and whether  $K(r)$  is equal to  $\pi r^2$ , it means that our points are independently distributed.

On the other hand, in the second scenario we use the whole of manufacturing industries as a benchmark. As we already know, this benchmark takes into account the first nature and can explain some aspects of the second nature, thus it is a more appropriate benchmark than the CSR.

$$M_{TM}(r) = K(r) - K_{TM}(r)$$

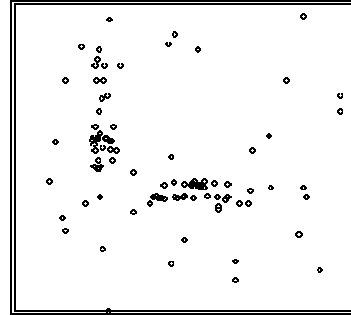
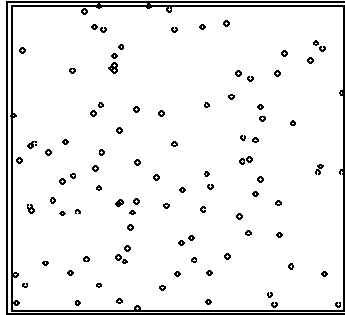
In this way,  $M_{TM}(r)$  is the difference between the  $K$ -value of every considered sector and the  $K$ -value of the total manufacturing sectors, appearing concentration in a particular sector whether its  $K$ -value is higher than the total manufacture  $K$ -value. In such a case, we affirm that this sector is concentrated relative to all other manufacturing firms.

Once explained both scenarios, the following figures try to illustrate the above-mentioned affirmations with CSR benchmark. On the one hand, Figure 1 shows an independent distribution and Figure 2 shows a concentrated distribution, both having one hundred points and the same area. On the other hand, Figure 3 and

---

<sup>10</sup> Marcon and Puech (2003) used the normalised function,  $L(r)$ , because they considered that 'a practical limitation of Ripley's  $K$  function is the need to compare any value to  $\pi r^2$ '. However, this is not an inconvenient for us, since our statistical software 'R' calculates this theoretical value ( $\pi r^2$ ) for each radius and we can compare it directly with real  $k$ -value.

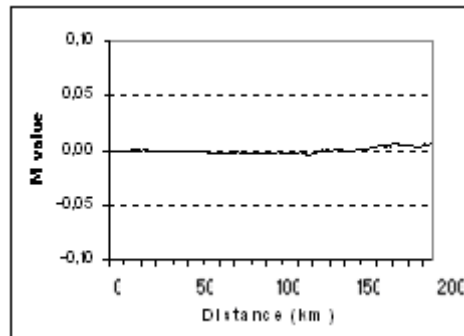
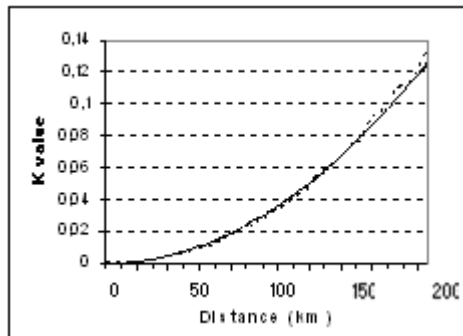
Figure 4 give us useful information about the spatial location patterns of the distribution shown in Figure 1 and 2.



**Figure 1.** Independent distribution.

**Figure 2.** Concentrated distribution.

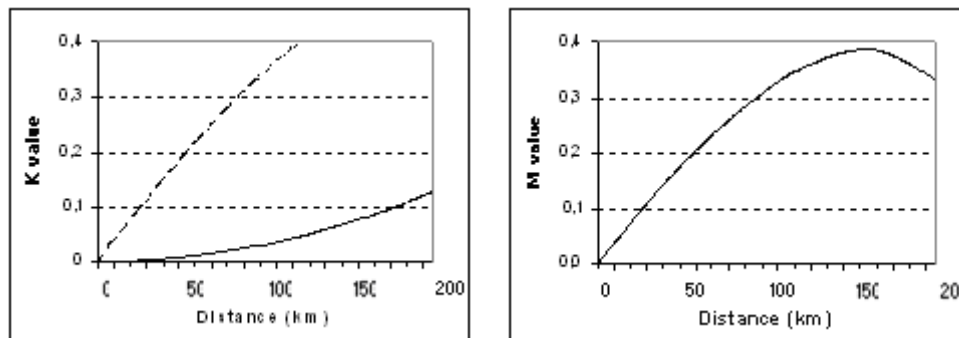
As we can see, two lines appear in Figure 3 (the graphic on the left). The dashed line represents the empirical  $K$  value, that is to say, the  $K$  value of the observed point pattern in Figure 1, and the continuous line takes a value of  $\pi r^2$ , since this represents the CSR benchmark. We realise that the values of these two lines are almost the same, whatever the distance of the radius we take into account ( $K(r) \approx \pi r^2$ ). The graphic on the right of Figure 3 shows us the  $M$  curve, which is the difference between the empirical and the theoretical  $K$  value. Therefore, we can affirm that an independent and random distribution of points produces an almost flat  $M$  curve, around the zero value.



**Figure 3.** Ripley's  $K$  function (theoretical and empirical) and  $M$  function corresponding to the point pattern from Figure 1.

Graphs in Figure 4 give us the same information that the previous ones, the  $K$  values and the difference between them, the  $M$  value. However, the interpretation of them is completely different, since they are the result of a concentrated distribution of points. We observe, on the left graph, that  $K(r) > \pi r^2$  at all distances of " $r$ " considered, meaning this that the point pattern in Figure 2

presents concentration. Moreover, viewing  $M$  curve we can know at which distance the highest level of concentration is reached. In this way,  $M$  value indicates us whether the point pattern is concentrated or dispersed, depending its positive or negative value, adding information about at which distance positive or negative peaks are found, being possible to find more than one. Therefore, owing to the  $M$ -value graph provides us information that is more detailed, we will use this graph along the paper to analyse the spatial location patterns of Spanish manufacturing firms.



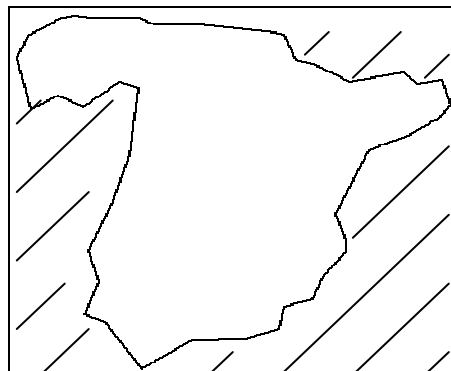
**Figure 4.** Ripley's  $K$  function (theoretical and empirical) and  $M$  function corresponding to the point pattern from Figure 2.

Now, to evaluate the statistical significance of departures from randomness, in a robust way, we should construct a confidence interval. The traditional technique to construct this confidence interval is the Monte Carlo method, generating a large number of independent random simulations. Should be highlighted that the construction of the confidence interval will be different in both scenarios. In first scenario, we simulate Poisson patterns with the same number of points than the real distribution. In addition, these simulations use the same area than the observed point pattern, being executed inside the above-mentioned polygonal area. In second scenario, we as well simulate random distributions of points but, this time, location of firms is fixed. Thus, to obtain the confidence interval in the second scenario we should take away the total manufacture  $K$ -value from the simulations  $K$ -value. With reference to the confidence interval, we just can add that both of them are generated by using 100 simulations and both allow us to reject the non-significant values, choosing a 95% confidence interval.

We should append that this methodology does not allow us to obtain robust results when the value of " $r$ " is, approximately, higher than 25% of the largest distance of our area of study. This inconvenient is overcome in our analysis, since we restrict it to 200 km, a value of the radius that represents a quarter of the largest distance of the area of study.

Regarding to the area of study, Marcon and Puech (2003) did not analyse the French whole country, but an industrial area of 40 x 40 km around Paris and a larger French rectangular area of 550 x 630 km. Their reason by not using the whole country was the increasing complexity when simulating random points inside the area and when correcting the border-effects on convex shapes. Therefore, these inconveniences limited their empirical analysis to rectangular areas. Nevertheless, Duranton and Overman (2005) did not have this shortcoming, since border-effect corrections were not necessary in their test, simplifying this fact their empirical analysis.

In our analysis, we improve the delimitation of our area of study, substituting the rectangular shape, used by other authors, for a polygonal shape. Thanks to the statistical software employed, 'R', we can apply border corrections, adequately, in any irregular polygonal shape, simplifying the treatment of border effects. In Figure 5, we can observe the polygonal shape that delimits perfectly our territory and envelopes the area of study, avoiding the nuisance of empty spaces where no firms are found, represented by the oblique lines.



**Figure 5.** Polygonal shape envelope.

Finally, should be emphasized that just Marcon and Puech (2003) and Duranton and Overman (2005) have used, until now, distance-based methods to assess the geographical concentration of activity. Methodology of both tests is similar, but not identical, and we can find some differences whether we analyse them in depth.

On the one hand, Duranton and Overman (2005) emphasized that their test fulfil all the five requirements that any concentration method should satisfy<sup>11</sup>. However, Ripley's  $K$  function does not satisfy two of these five requirements, since it does

---

<sup>11</sup> As underlined by D&O (2005), (1) it is comparable across industries, (2) controls for the overall agglomeration of manufacturing, (3) controls for industrial concentration, (4) is unbiased with respect to scale and aggregation and (5) gives an indication of the significance of the results.

not control for the overall agglomeration of manufacturing or for the industrial concentration and D function does not respect requirements 1 and 3.

On the other hand, the Duranton and Overman's method has a great disadvantage, with regard to Ripley's  $K$  function, not being possible to quantify the concentration or dispersion, but only detecting the proportion of sectors that are concentrated.

Comparing now our test, we realise that  $M_{TM}(r)$  function fulfils all the requirements established by Duranton and Overman, with the exception of the third one.

### **3. Data and Area of Study**

Our empirical analysis uses current establishment level data, for the year 2007, from SABI (Analysis System of Iberian Balances), which contains detailed information about Spanish and Portuguese companies. We restrict our database to manufacturing establishments, using the National Classification of Economic Activities<sup>12</sup> and analysing sectors at two-digit level. Besides, we add another two requirements to our database. First, we make that our database contains only intercontinental Spanish manufacturing firms, without including firms from Canary and Balearic Islands, Ceuta and Melilla. Second, we restrict our analysis just to firms employing at least ten workers. Finally, once applied these requirements, our database contains exactly 43.087 observations, or firms.

The comparison between our restriction, related to the number of employees, and Marcon and Puech (2003)'s restriction must be considered. Actually, they use French manufacturing firms employing at least twenty workers. This difference, concerning the number of employees, is because of the fact that SME (small and medium-sized enterprises) predominate in Spain. Therefore, a too elevated percentage of firms would be left out if we just considered those with twenty or more than twenty workers, as Marcon and Puech did.

Spanish manufacturing activities are classified into 23 sectors according to 'NACE 93 - Rev. 1' and these are the following ones: (15) Food products and beverages, (16) Tobacco products, (17) Textiles, (18) Wearing apparel and dressing, (19) Tanning and dressing of leather, (20) Wood and products of wood, (21) Pulp, paper and paper products, (22) Publishing, printing and recorded media, (23) Coke, refined petroleum products, (24) Chemical and chemical products, (25) Rubber and plastic products, (26) Other non-metallic mineral products, (27) Basic metals, (28) Fabricated metal products, (29) Other machinery and equipment, (30)

---

<sup>12</sup> NACE 93 - Rev. 1

Office machinery and computers, (31) Electrical machinery, (32) Radio, televisions and other appliances, (33) Instruments, (34) Motor vehicles and trailers, (35) Other transport equipment, (36) Furniture and other products, (37) Recycling.

Table A1, situated in *Appendix 1*, shows us a brief descriptive analysis about the above-mentioned sectors, including additional information as the number of firms or the technological intensity of every sector. Therefore, as we can see in ‘Table A1’, there are great differences in the number of firms, depending on the sector we refer. Therefore, three of the twenty-three sectors considered, (16) Tobacco products, (23) Coke, refined petroleum products and (30) Office machinery and computers, have less than 200 firms and will not be analysed, since its results may be altered when using our methodology.

For every establishment we know its *precise location* through geographic coordinates (longitude and latitude). Every firm has been localized exactly in the same place as in reality; in this way, our margin of error for any firm’s location is non-existent and we perfectly know every firm’s precise spatial location. These geographic coordinates are transformed into UTM coordinates or, also named, flat coordinates. This transformation has been carried out by means of the method proposed by Morton (2003). This procedure converts Latitude and Longitude coordinates to Easting and Northing coordinates, on a Transverse Mercator projection; being the UTM coordinates expressed in metres. We should highlight that the construction of this system let us move away from the Equator with hardly distortions, because any point is far away from the central meridian of its zone. For this reason, it appears the great advantage of the UTM system, being able to localize any firm with absolute precision.

Given that the spatial location of Spanish manufacturing firms is the key information to construct our database, it is obvious that our area of study should be quite similar, or identical, to the Spanish shape. It is as well obvious that will be unproductive to analyse a rectangular shape, as in previous studies, because Spanish shape is not regular as a rectangle. Therefore, will be necessary the construction of a polygonal area of study by means of a polygonal envelope, already mentioned before. The area of this envelope has similar measures to Spanish territory, since has been built by means of the union of thirty-five points of the Spanish territory perimeter.<sup>13</sup>

---

<sup>13</sup> Thanks to this technique and the statistical software used, we can treat the ‘border-effects’ adequately in any polygonal shape, aspect that so much worried Marcon and Puech.

#### 4. Empirical Results and Discussion<sup>14</sup>

In *Appendix 2*, we can find the resulting graphs coming from our empirical analysis. On the one hand, graphs situated on the left show the spatial distribution of firms, in the area of study, of every Spanish manufacturing sector. On the other hand, graphs on the right-up show the spatial location patterns of these sectors measured by the  $M_{CSR}$  function and right-down graphics show the spatial location patterns measured by means of the  $M_{TM}$  function.

**Table 1.** Results of the  $M_{CSR}$  function for every Spanish manufacturing sector

Sectors (NACE 93 - Rev. 1)	Significant concentration	Significant dispersion	Type of cluster <sup>15</sup>	Significant peak <sup>16</sup> Distance ( $M$ value)
15 Food products and beverages	All distances	---	2	100 km (0,05)
17 Textiles	All distances	---	1	87 km (0,27)
18 Wearing apparel and dressing	All distances	---	2	87 km (0,09)
19 Tanning and dressing of leather	All distances	---	2	156 km (0,50)
20 Wood and products of wood	All distances	---	2	200 km (0,08)
21 Pulp, paper and paper products	All distances	---	1	99 km (0,16)
22 Publishing, printing & recorded media	All distances	---	1	85 km (0,20)
24 Chemical and chemical products	All distances	---	1	90 km (0,17)
25 Rubber and plastic products	All distances	---	1	97 km (0,13)
26 Other non-metallic mineral products	All distances	---	2	200 km (0,08)
27 Basic metals	All distances	---	1	80 km (0,10)
28 Fabricated metal products	All distances	---	1	86 km (0,09)
29 Other machinery and equipment	All distances	---	1	96 km (0,13)
31 Electrical machinery	All distances	---	1	90 km (0,16)
32 Radio, televisions & other appliances	All distances	---	1	91 km (0,21)
33 Instruments	All distances	---	1	90 km (0,21)
34 Motor vehicles and trailers	All distances	---	1	100 km (0,11)
35 Other transport equipment	0-186 km	189-200 km	1	64 km (0,07)
36 Furniture and other products	All distances	---	2	183 km (0,12)
37 Recycling	All distances	---	1	86 km (0,09)

<sup>14</sup> Discussion about  $M_{TM}$  results is coming briefly.

<sup>15</sup> By means of this information we know, roughly, the *persistence* of the concentration.

<sup>16</sup> This peak shows us the distance at which the maximum significant concentration is reached in every Spanish manufacturing sector. Besides, the ( $M$  value) informs us about the highest *intensity* of concentration.

The information obtained thanks to the graphics constructed by means of the  $M_{CSR}$  value appears summarized in 'Table 1'. Thus, first of all and paying attention to the second and third column, we must highlight that every Spanish manufacturing sector analysed, excepting (35) 'Other transport equipment', present concentration whatever the distance of " $r$ " we consider. Nevertheless, we should not forget that the highest value of the radius considered is 200 km.

Secondly, we should follow our analysis of results valuing the *intensity* reached by the different sectors, that is to say, introducing which are the highest and the lowest concentrated manufacturing sectors in Spain. In this way, the Spanish manufacturing sectors that reach a *highest concentration level* are, in the following order, (19) Tanning and dressing of leather, (17) Textiles, (32) Radio, televisions & other appliances, (33) Instruments, (22) Publishing, printing and recorded media, (24) Chemical and chemical products and (31) Electrical machinery. We should highlight that this result is independent from the  $M$  value taken into account,  $M_{CSR}$  or  $M_{TM}$ . On the contrary, the sectors with the *lowest concentration level* are (15) Food products and beverages, (20) Wood and products of wood, (26) Other non-metallic mineral products and (35) Other transport equipment.

It may be interesting to compare the above-mentioned results with those obtained by Duranton and Overman (2005) in UK. Surprisingly, they are very similar. On the one hand, they find that the localised sectors in UK are 17, 18, 19, 22, 30, 31, 32 and 33, which almost coincide with the most concentrated sectors in Spain. On the other hand, the lowest concentrated sectors in Spain coincide perfectly with those non-localized sectors in UK, 15, 20 and 26. Therefore, we notice that manufacturing sectors tend to follow similar patterns of localization between countries. At least, it seems to be between Spain and UK. Besides, the most concentrated sectors in France and US are Textile (17) and Leather products (19)<sup>17</sup>, coinciding as well these results with the most concentrated sectors in Spain and UK.

Alonso-Villar *et al.* (2004) as well studied the geographical concentration of Spanish industry, between 1993 and 1999<sup>18</sup>, and concluded that the most highly concentrated industries, according to M-S, are 19, 17, 32, 22, 33 and 24. As we see, these results coincide perfectly with ours. Besides, thanks to the results obtained by them, we as well can deduce that the spatial location patterns of

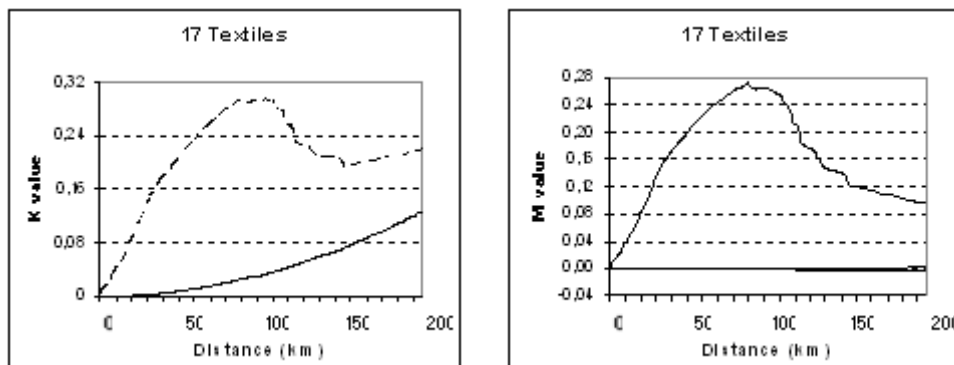
---

<sup>17</sup> As can be deduced from Maurel and Sédillot (1999) and Ellison and Glaeser (1997) results.

<sup>18</sup> Although they just presented results for 1999, because not many differences were observed throughout the whole period.

Spanish manufacturing sectors have not varied significantly in the last years, since the highest concentrated sectors are still being the same as in 1999.

Third, we should evaluate the *persistence* of this concentration in the space, in other words, the spatial scale dimensions of the cluster. If we pay attention to this characteristic of the spatial distribution, we can find two differentiated kinds of sectors, named 1 or 2 depending the characteristic features of their localization patterns.<sup>19</sup> Sectors named with number 1 present, in the beginning, an increasing in the activity concentration, reaching an absolute maximum point or peak and, finally, decreasing just as the distance of the radius makes higher. We should add that, although the evolution of these sectors is similar between them, the intensity and the distance at which is reached the maximum concentration are different; being similar, but not identical, the distribution patterns of these ones. Most of the Spanish manufacturing sectors belong to this ‘first type’ of localization patterns<sup>20</sup> and this is a clear example.

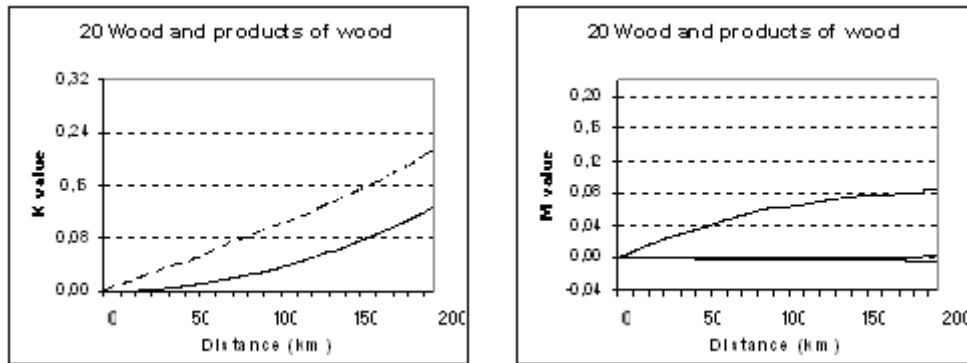


**Figure 6.** Sector with ‘first type of cluster’.

However, just six sectors belong to the second classification (15, 18, 19, 20, 26 and 36). These sectors keep a constant or growing concentration, whatever the value of “*r*” considered, without reaching an absolute maximum peak. In this second kind of sectors, we assume that this maximum peak will be reached in a value of the radius higher than 200 km, and to exemplify it we use the ‘*Wood and products of wood*’ sector.

<sup>19</sup> We can find this information in the fourth column, ‘Type of cluster’.

<sup>20</sup> Sectors 17, 21, 22, 24, 25, 27, 28, 29, 31, 32, 33, 34, 35 and 37.



**Figure 7.** Sector with ‘second type of cluster’.

We should emphasize that this second type of sectors<sup>21</sup> have a common characteristic, related to technological intensity. In fact, the whole of these sectors are traditional and low-tech.

Once analysed the intensity and the persistence of concentration of every Spanish manufacturing sector, we should think some questions related to the *distance* (*r*-value) at which the highest concentration is reached. Then, *should we understand that every sector reaches the highest concentration level at the same distance?* Moreover, whether it would happen, *would this fact mean that these sectors have similar localization patterns?* If we ask ourselves these two questions, we realise that the intensity and the distance at which the highest concentration<sup>22</sup> is reached differs between sectors. That is to say, roughly speaking, each Spanish manufacturing sector seems to present different concentration patterns, although, we can find some similarities between them.

Now, paying attention to the *distance* at which the highest concentration is reached or, in other words, to the spatial scale at which the cluster is produced, we realise that this distance coincides in some sectors. In fact, as we can see in ‘Table 1’, nine of the twenty sectors analysed obtain their maximum concentration peak at a distance comprised between 85 and 91 km<sup>23</sup>. Once appreciated this coincidence, we should find out whether there is any common characteristic between these sectors. For example, sectors 17 and 18 tend, historically, to follow similar patterns of distribution, or sectors 31, 32 and 33 may depend on similar technological spillovers, distributing, like this, in a similar way. However, we cannot find common characteristics to all of them, confirming, like this, that are sectors with widely varying characteristics. Therefore, we see that a common behaviour exists in the distance at which appears the highest concentration, but

<sup>21</sup> Those that keep a constant or growing concentration whatever the “*r*” we take into account.

<sup>22</sup> As well named ‘maximum concentration peak’ or ‘significant peak’.

<sup>23</sup> These sectors are 17, 18, 22, 24, 28, 31, 32, 33 and 37.

this coincidence cannot be attributed to common characteristics of the sectors. In this way, we should look for another reason that could explain this fact, being able to be possible that our 'benchmark' can create regularities in the maximum concentration distance of every sector. Indeed, by means of a random and independent distribution (CSR benchmark) is not taken into account the "first nature" factors. This fact can create a coincidence in the distance at which the highest concentration is reached, creating a fictitious regularity, non-attributable to economic factors, but to "first nature" factors.

The patterns of location that presents Spanish manufacturing sectors are not arbitrary and the literature has already demonstrated it. Many are the agglomeration externalities that induce activity to concentrate and firms to locate next to other firms. The role of '*knowledge spillovers*' seems evident, whether we refer to them as a source of concentration, although in our case it is not so obvious. In fact, the two first sectors that have a higher level of concentration are clearly low-tech. This indicates us that not only the 'knowledge spillovers' determine the concentration of activity, but also other factors as local labour pooling, natural advantages, tradition, transport costs or the linkages<sup>24</sup>, play an important role in the patterns of localization. Therefore, among the most highly concentrated Spanish manufacturing sectors we find those for which the geographic concentration is completely determined by *historical trends*<sup>25</sup> (17 and 19), those for which *technological spillovers* seem to be the main reason of their localization (31, 32 and 33) and, finally, those for which the search of *skilled labour* is determinant in their decision to concentrate.

Spanish manufacturing sectors do not follow location patterns appropriate with economic theory, in relation to their technological intensity, since according to literature, high-tech sectors must be the most highly concentrated; although, neither occurs in UK or France. As underlined by Devereux *et al.* (2004) and according to their results, '*the most geographically concentrated industries appear to be relatively low-tech*' and Maurel and Sédillot (1999) obtain that the most concentrated sectors are textile and leather products, two of the most traditional and low-tech sectors.

Therefore, according to spatial location patterns, we can affirm that the low-tech sectors do not carry out the theory, since these are amongst the most highly concentrated; however, we must emphasize that the whole of the sectors that keep a constant or growing spatial distribution when "*r*" makes higher, without reaching an absolute maximum peak, are low-tech.

---

<sup>24</sup> So emphasized by Krugman (1991a).

<sup>25</sup> They are probably settled and clustered in the same area since the Industrial Revolution.

Finally, we should notice an obvious aspect of our analysis. Therefore, considering that our data is aggregated at two-digit level, there will be a clear effect of compensation between the different sub-sectors of activity, since the most aggregated and the most dispersed ones will compensate between them.

## 5. Conclusions

This paper has investigated the spatial location patterns of manufacturing firms in Spain. To do this, we have used a distance-based method, which allows us to treat space as continuous and avoids the inconvenience of using just one administrative scale. By means of this method, we can know the intensity of concentration of every Spanish manufacturing sector, at which distance the maximum concentration peak appears and the persistence of every cluster.

Our empirical analysis incorporates an advantage, which simplifies the treatment of border effects. By means of the software employed, R, we can apply border corrections, adequately, in any irregular polygonal shape. Therefore, we introduce a new envelope technique that delimits perfectly the Spanish territory, without leaving any firm out. This new envelope has polygonal shape and allows us to analyse the whole country, treating, at the same time, the border effects. In this way, we incorporate an improvement to Marcon and Puech (2003) paper, since they limit their empirical analysis to rectangular areas.

According to our results<sup>26</sup>, we conclude that every Spanish manufacturing sector analysed, excepting (35), presents concentration whatever the distance of “ $r$ ” we consider. We as well obtain that the spatial location patterns of Spanish manufacturing firms do not appear determined by a unique factor. Nevertheless, we can find a common characteristic in the distribution of six sectors, which keep a constant or growing concentration whatever the “ $r$ ” we consider. In fact, the whole of these sectors are traditional and low-tech and their distribution in space is more homogeneous. According to the coincidence in the distance at which the highest concentration is reached, we conclude that, may be, a fictitious regularity is created, which is non-attributable to economic factors, but to “first nature” factors. Finally, we can add that appear coincidences between our results and Duranton and Overman ones, being possible to conclude that manufacturing sectors tend to follow similar patterns of localization between countries, at least between Spain and UK.

---

<sup>26</sup> We should realise that only results coming from  $M_{CSR}$  function have been analysed, being well known that this type of analysis is sensitive to the benchmark considered. Therefore, it will be fruitful the comparison between  $M_{CSR}$  results and  $M_{TM}$  results.

Our paper has introduced some improvements, since the data, the area of study and the computing analysis are not a limitation now. However, there are some aspects to follow improving in next studies. We can add firm size in next analysis, being able to control the industrial concentration, in this way.

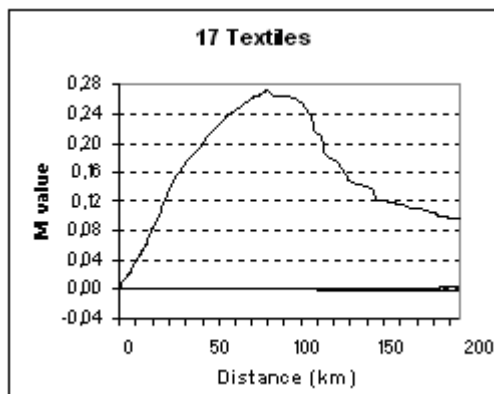
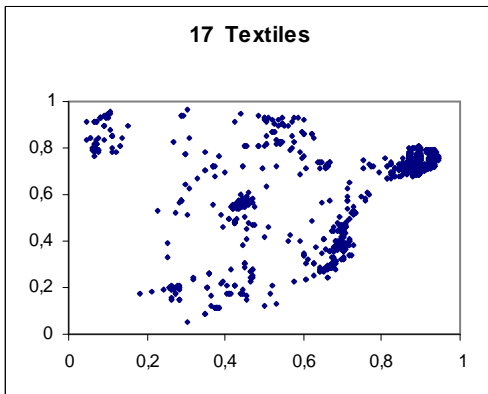
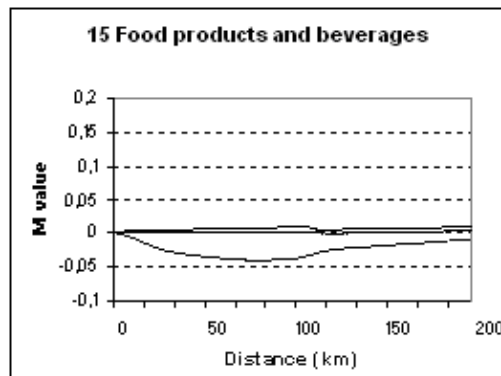
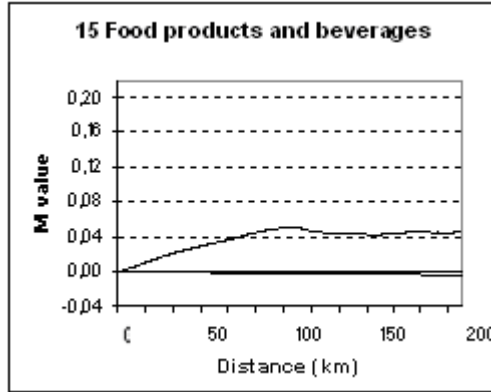
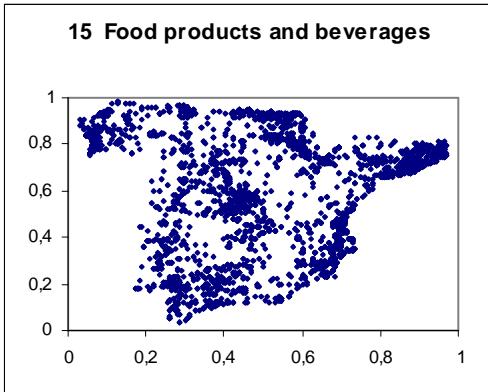
## Appendix 1

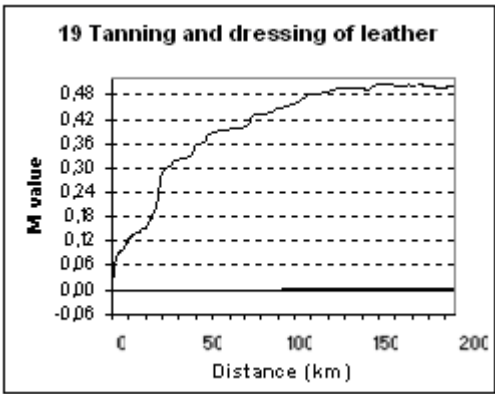
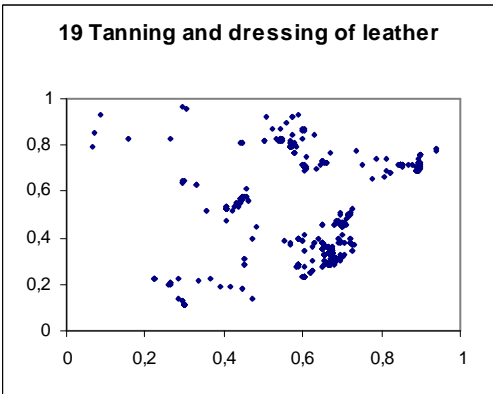
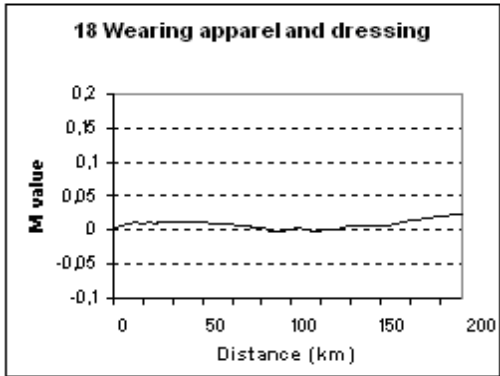
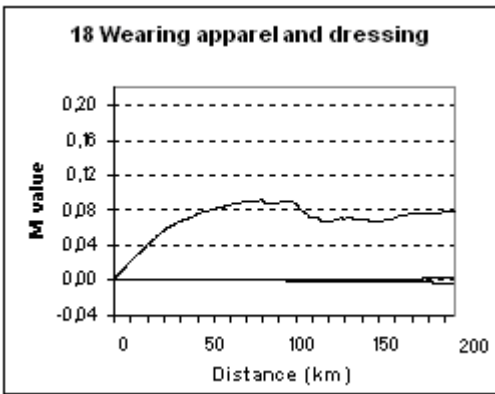
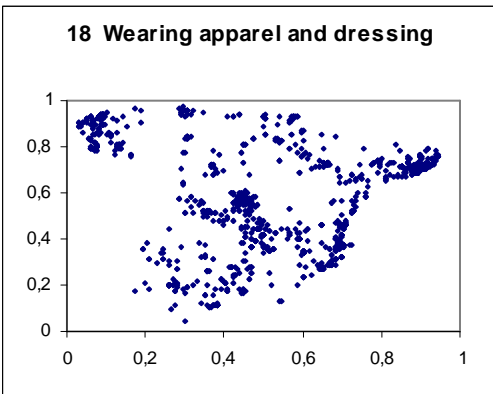
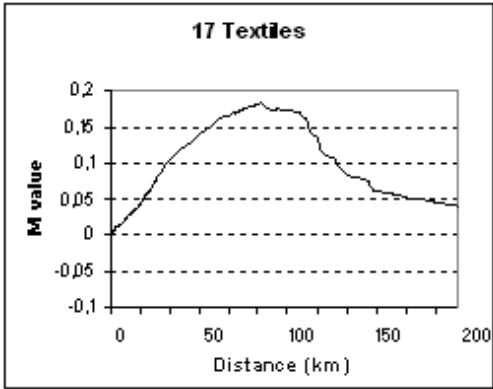
**Table A1.** Additional descriptive information about Spanish manufacturing sectors

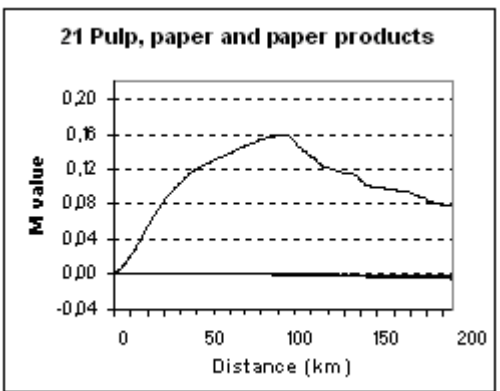
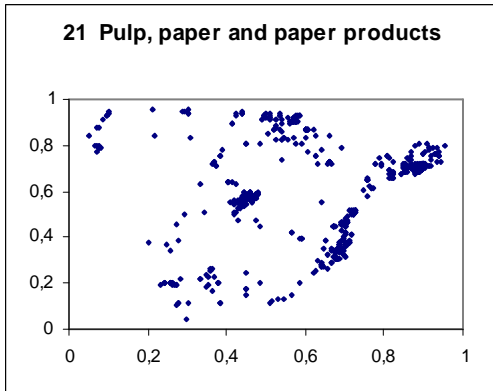
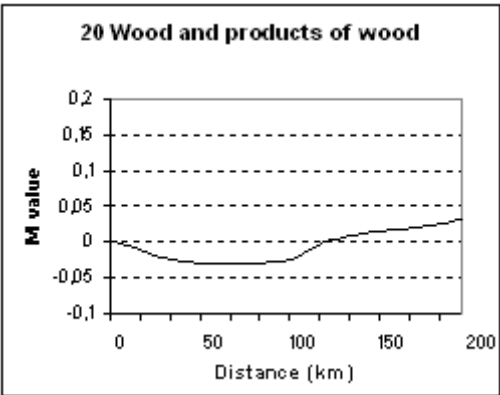
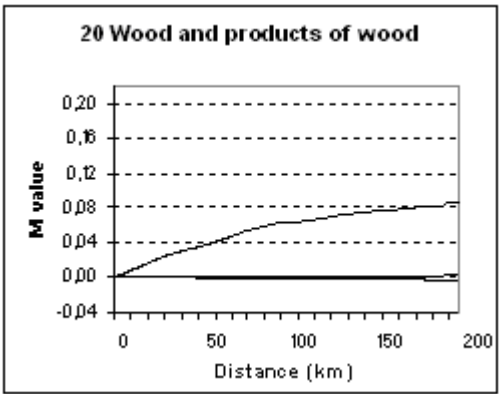
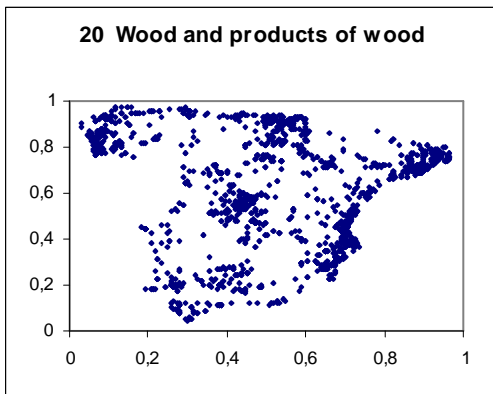
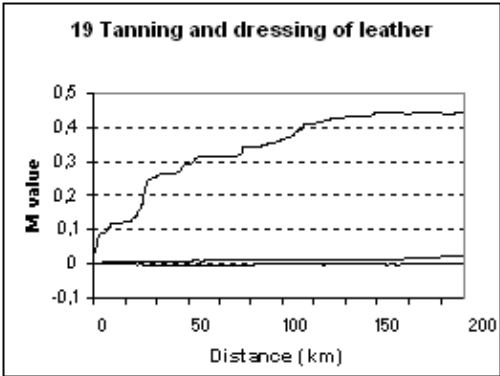
Sectors (NACE 93 - Rev. 1)	Number of firms	Technological intensity <sup>27</sup>
15 Food products and beverages	5761	L
16 Tobacco products	6	L
17 Textiles	1949	L
18 Wearing apparel and dressing	1710	L
19 Tanning and dressing of leather	1698	L
20 Wood and products of wood	2340	L
21 Pulp, paper and paper products	837	L
22 Publishing, printing & recorded media	3004	L
23 Coke, refined petroleum products	12	M-L
24 Chemical and chemical products	1722	H
25 Rubber and plastic products	2165	M-L
26 Other non-metallic mineral products	3413	M-L
27 Basic metals	986	M-L
28 Fabricated metal products	8094	M-L
29 Other machinery and equipment	3015	M-H
30 Office machinery and computers	77	H
31 Electrical machinery	1099	M-H
32 Radio, televisions & other appliances	344	H
33 Instruments	376	H
34 Motor vehicles and trailers	876	M-H
35 Other transport equipment	451	M-H
36 Furniture and other products	2924	L
37 Recycling	228	L

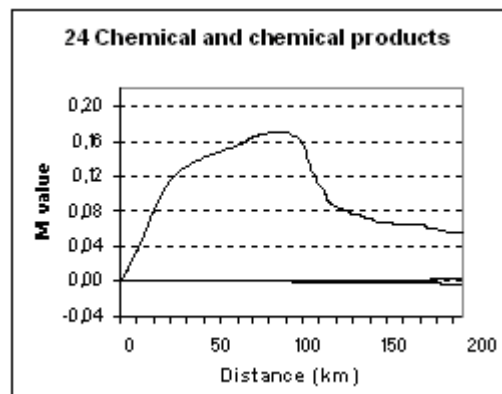
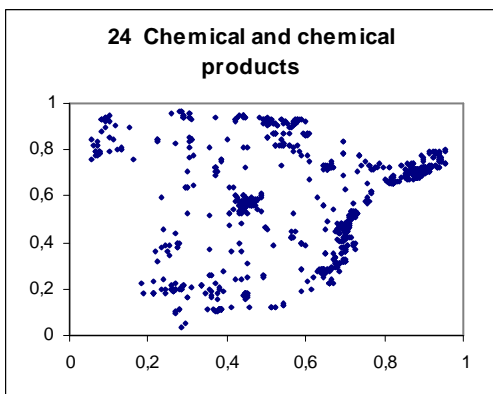
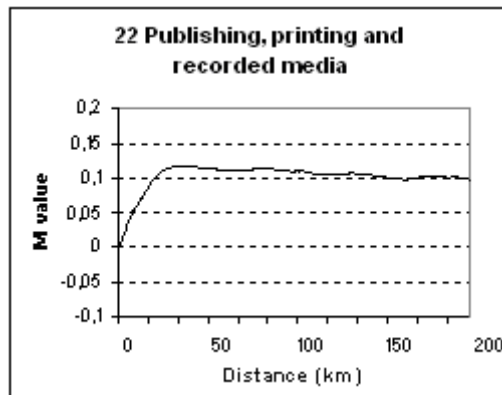
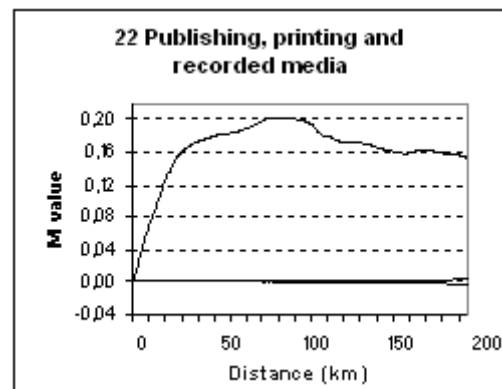
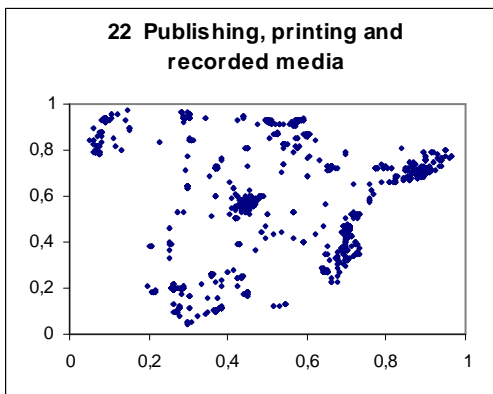
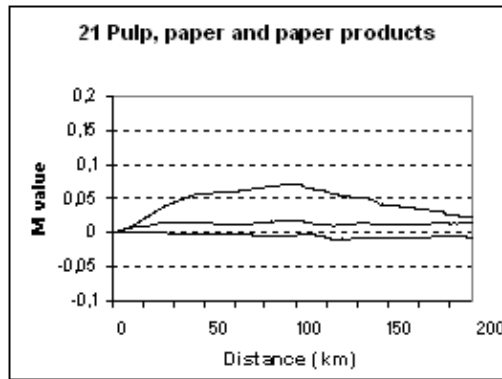
<sup>27</sup> This classification of sectors according to the technological intensity belongs to the National Statistics Institute, meaning H = high, M-H = medium high, M-L = medium low and L = low.

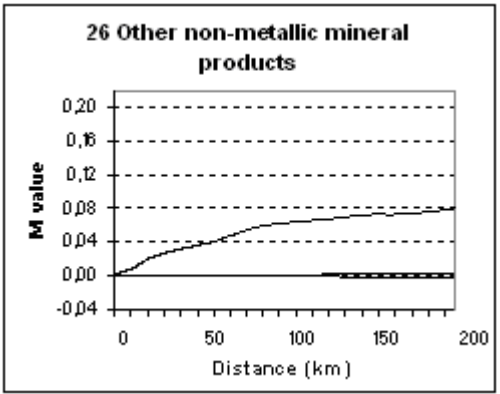
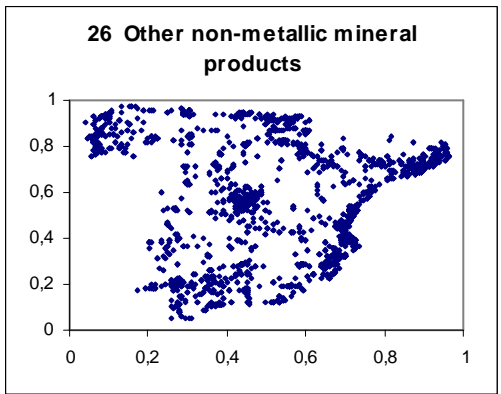
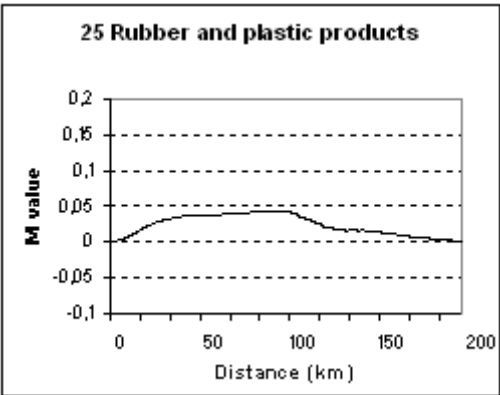
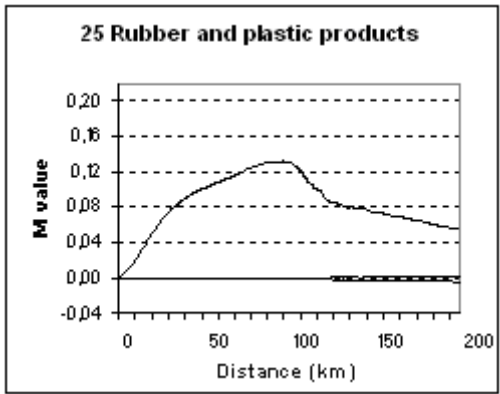
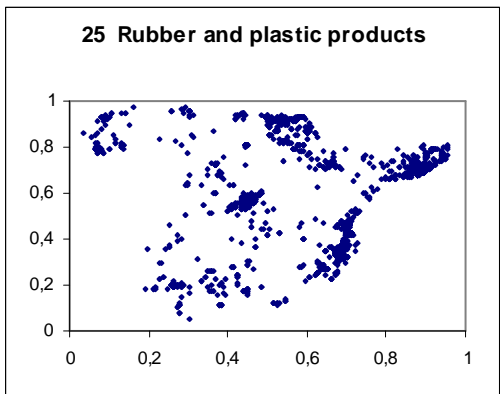
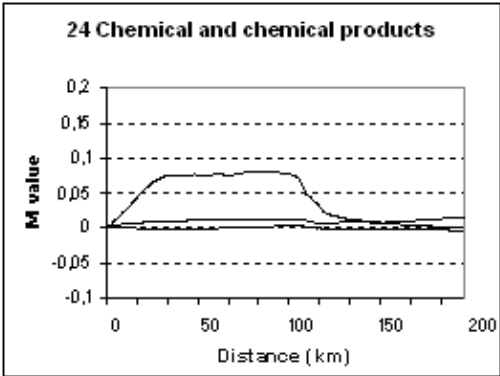
*Appendix 2*

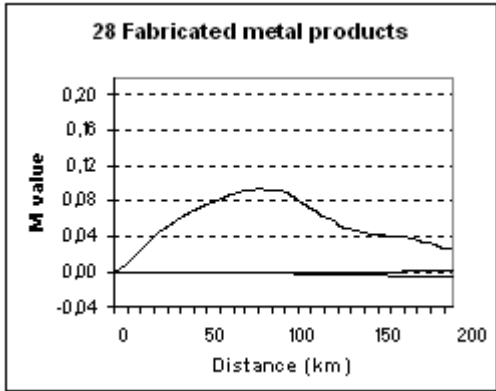
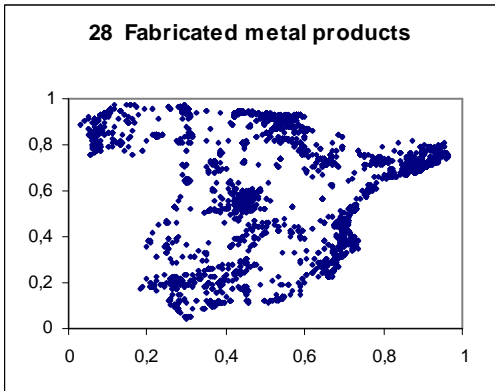
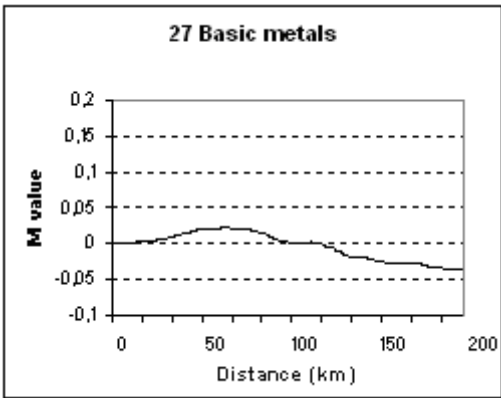
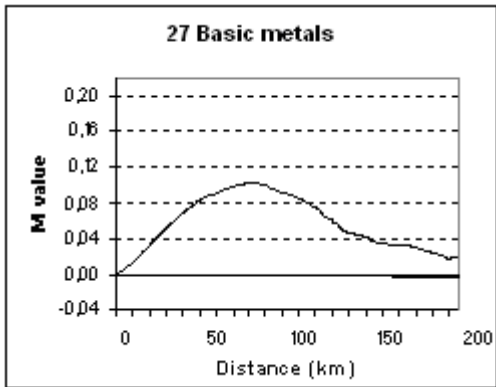
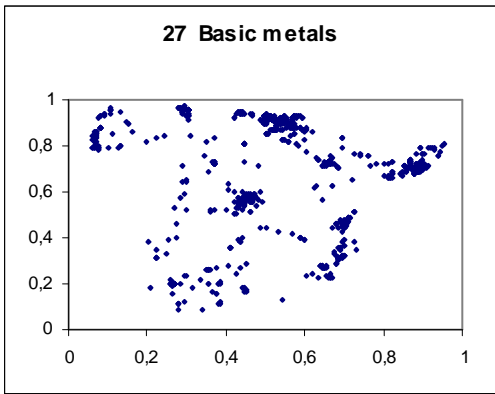
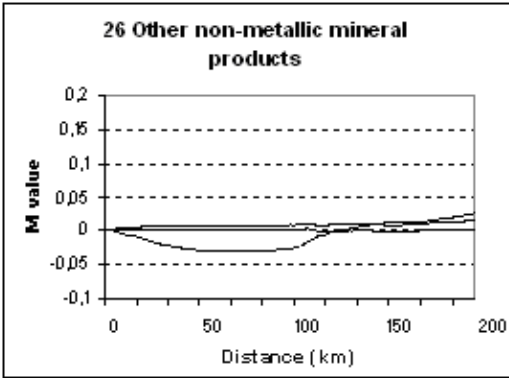


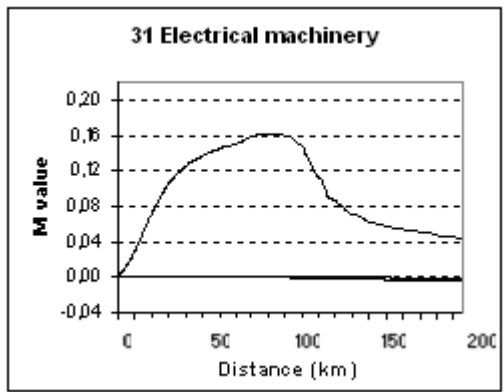
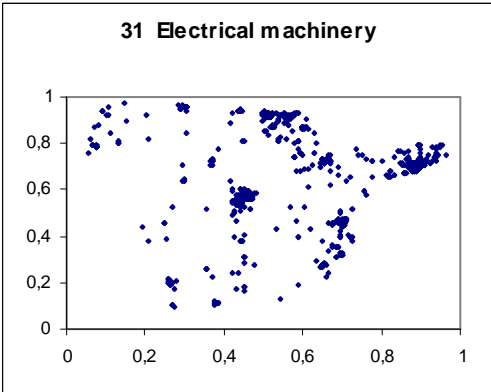
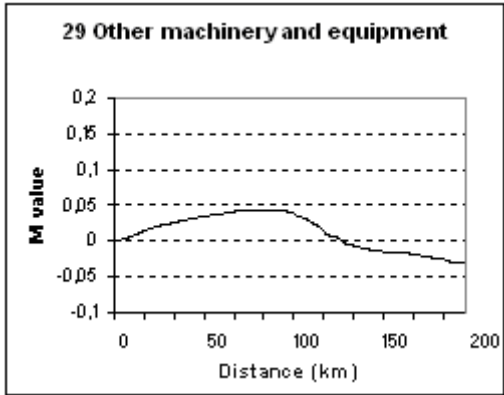
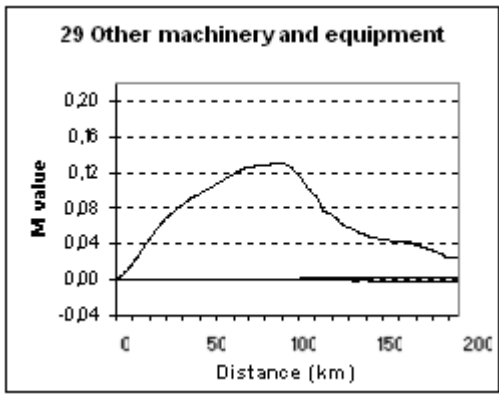
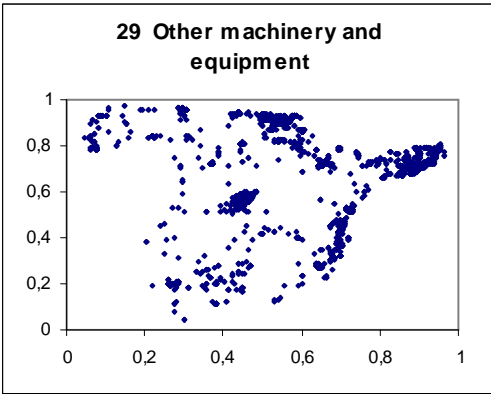
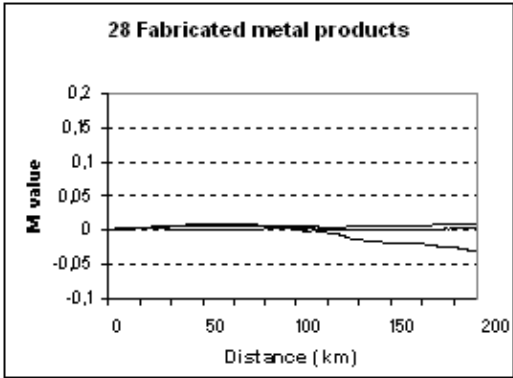


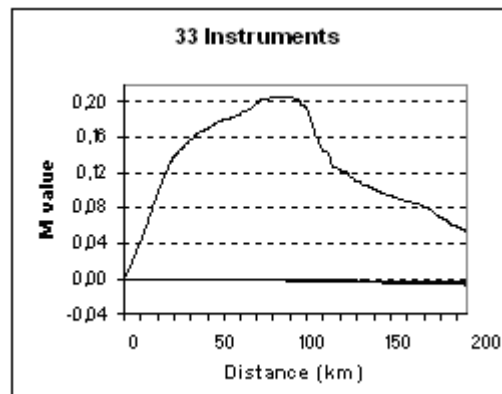
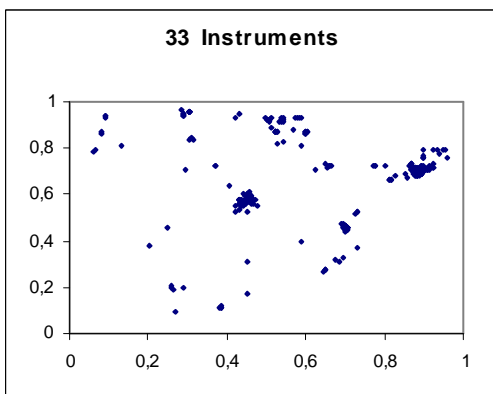
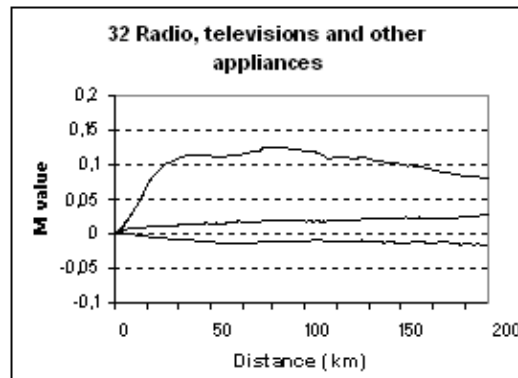
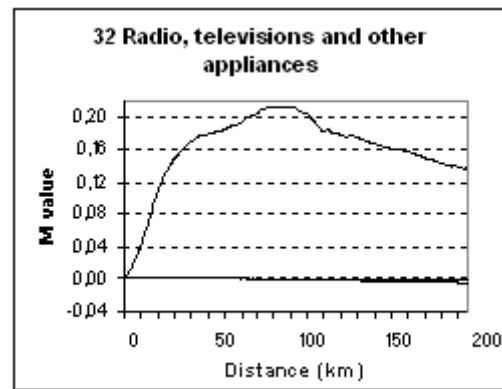
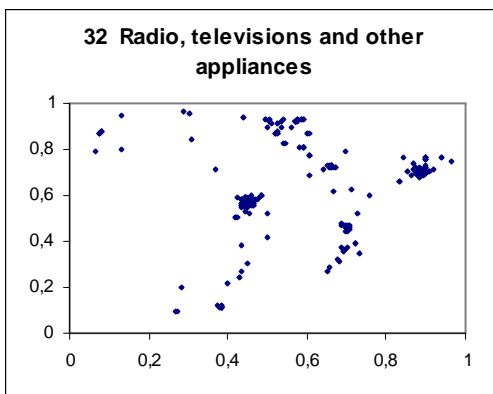
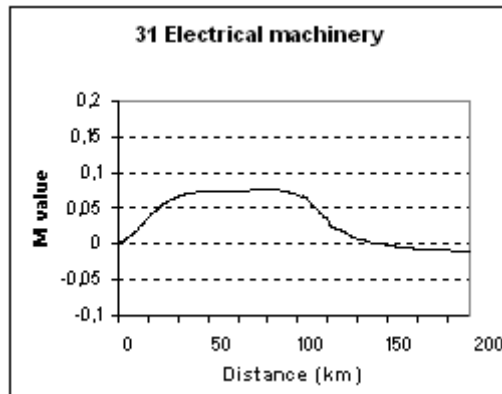


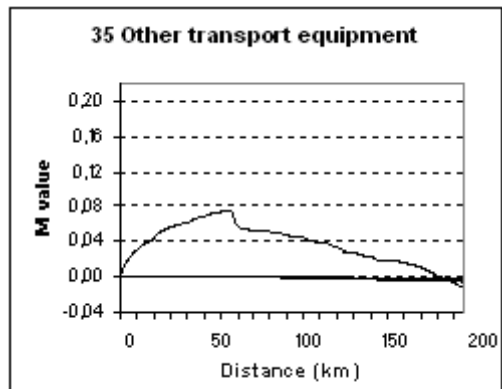
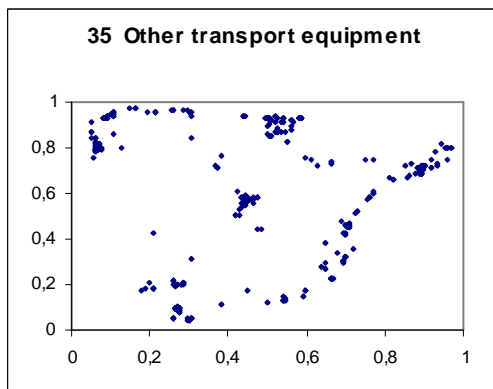
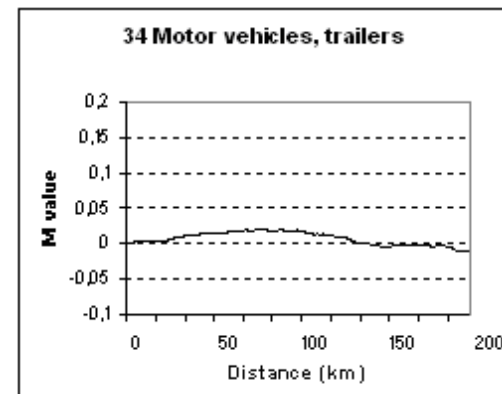
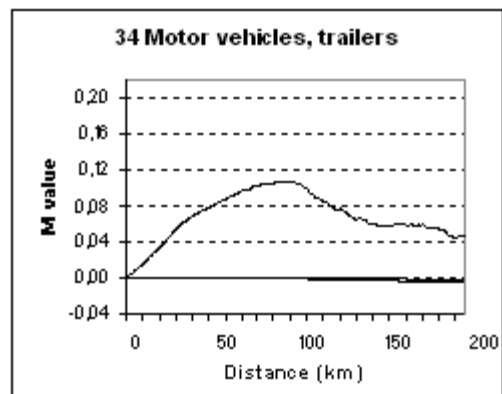
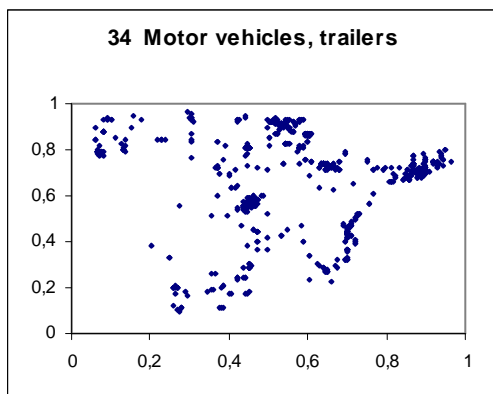
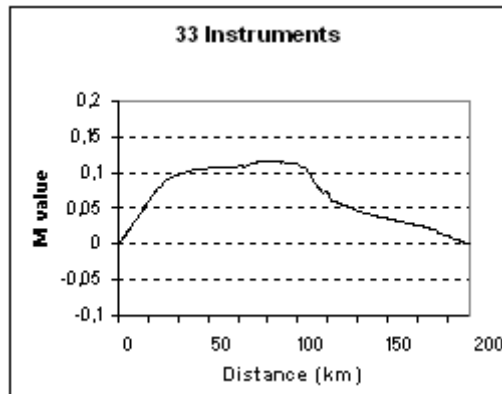


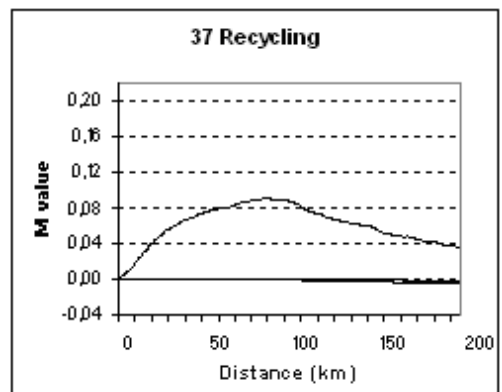
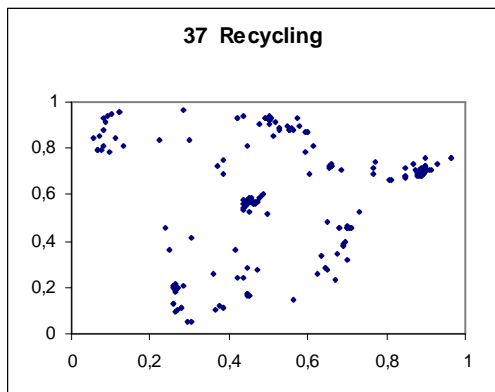
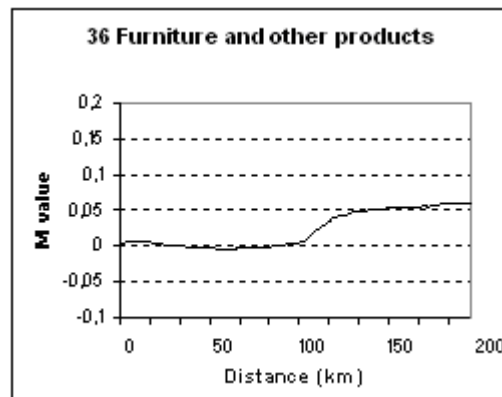
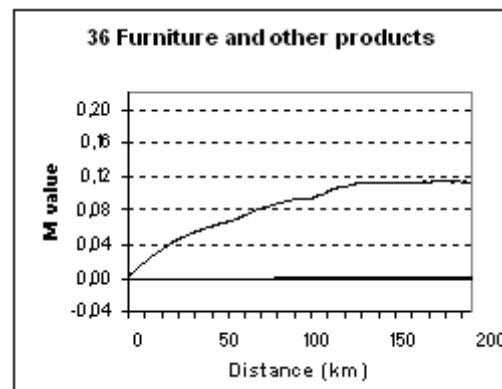
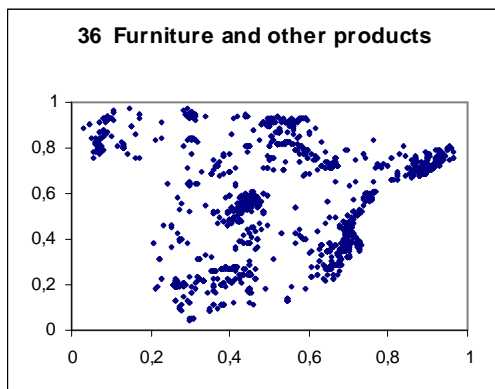
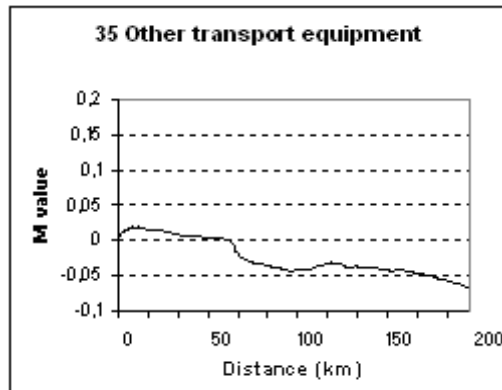


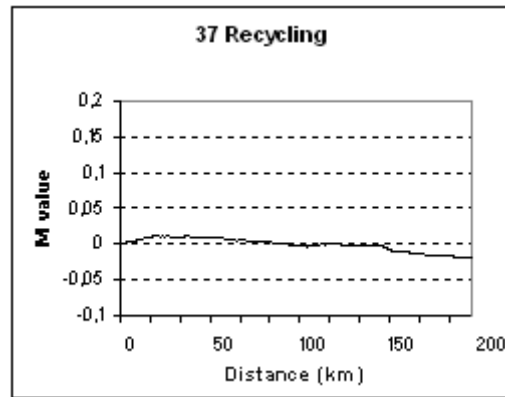












## References

- Alonso-Villar, O., Chamorro-Rivas, J.M. and González-Cerdeira, X., (2001) An Analysis of the Geographic Concentration of Industry in Spain. Documento de Trabajo 0103, Departamento de Economía Aplicada, Universidade de Vigo.
- Alonso-Villar, O., Chamorro-Rivas, J.M. and González-Cerdeira, X., (2003) Spillovers geográficos y sectoriales de la industria española. *Revista de Economía Aplicada*, 32, 77-95.
- Alonso-Villar, O., Chamorro-Rivas, J.M. and González-Cerdeira, X., (2004) Agglomeration economies in manufacturing industries: the case of Spain. *Applied Economics*, 36, 2103-2116.
- Amiti, M., (1997) Specialisation patterns in Europe. Discussion paper 363, Centre for Economic Performance, London
- Brülhart, M., (2001) Evolving geographical concentration of European manufacturing industries. *Weltwirtschaftliches Archiv*, 137 (2), 215-243.
- Callejón, M., (1997) Concentración geográfica de la industria y economías de aglomeración. *Economía Industrial*, 317, 61-68.
- Devereux, M.P., Griffith, R. and Simpson, H., (2004) The geographic distribution of production activity in the UK. *Regional Science and Urban Economics*, 34, 533-564.
- Duranton, G., Overman, H.G., (2005) Testing for Localisation using Micro-Geographic Data. *Review of Economic Studies*, 72, 1077-1106.
- Ellison, G., Glaeser, E., (1997) Geographic concentration in U.S. manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105 (5), 889-927.
- Ellison, G., Glaeser, E., (1999) The geographic concentration of industry: does natural advantage explain agglomeration? *American Economic Review*, 89 (2), 311-316.
- Fujita, M., Krugman P. and Venables A.J., (1999) *The Spatial Economy: Cities, Regions and International Trade*. MIT Press, Cambridge, MA.
- Krugman, P., (1980) Scale Economies, Product Differentiation and the Pattern of Trade. *American Economic Review*, 70, 950-959.
- Krugman, P., (1991a) *Geography and Trade*. MIT Press, Cambridge, USA.
- Krugman, P., (1991b) Increasing returns and economic geography. *Journal of Political Economy*, 99 (3), 413-499.
- Marcon, E. and Puech, F., (2003a) Evaluating the Geographic Concentration of Industries using Distance-Based Methods. *Journal of Economic Geography*, 3 (4), 409-428.

- Marcon, E. and Puech, F., (2003b) Measures of the Geographic Concentration of Industries: Improving Distance-Based Methods. *Cahiers de la MSE*, 2003.18. 22p.
- Marshall, A., (1890) *Principles of Economics*. MacMillan, London.
- Maurel, F., Sédillot, B., (1999) A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics*, 29 (5), 575-604.
- Morton, A., (2003) Workbook from Alan Morton. Electronic publication. Distribution mapping software (DMAP), <http://www.dmap.co.uk>.
- Ottaviano, G.I.P., Puga, D., (1998) Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'. *The World Economy*
- Ottaviano, G., Thisse, J.F., (2004) Agglomeration and Economic Geography. *Handbook of Regional and Urban Economics*, Volume 4, 2563-2608.
- Overman, H.G., Redding S. and Venables A.J., (2003) The Economic Geography of Trade Production and Income: a Survey of Empirics. *Handbook of International Trade*. Choi, E.K. and Harrigan, J., Blackwell Publishing Ltd., 353-387.
- Paluzie, E., Pons, J. and Tirado, D.A., (2001) Regional Integration and Specialization Patterns in Spain. *Regional Studies*, 35 (4), 285-296.
- Paluzie, E., Pons, J. and Tirado, D.A., (2004) The geographical concentration of industry across Spanish regions, 1856-1995. *Jahrbuch für Regionalwissenschaft (Review of Regional Research)*, 24, 2, 143-160.
- Puga, D., (1999) The rise and fall of regional inequalities. *European Economic Review*, 43 (2), 303-334.
- Puga, D., (2002) European regional policies in light of recent location theories. *Journal of Economic Geography*, 2, 373-406.
- Quah, D., (1996) Regional convergence clusters across Europe. *European Economic Review*, 40, 951-958.
- R Development Core Team (2007) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ripley, B.D., (1976) The second-order analysis of stationary point processes. *Journal of Applied Probability* 13, 255-266.
- Ripley, B.D., (1977) Modelling Spatial Patterns. *Journal of the Royal Statistical Society - Series B (Methodological)*, 39 (2), 172-192.

Ripley, B.D., (1979) Test of 'randomness' for spatial patterns. *Journal of the Royal Statistical Society - Series B (Methodological)*, 41, 368-374.

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

SABI. System of Iberian Balances Analysis.

Tirado, D.A., Paluzie, E. and Pons, J., (2002) Economic integration and industrial location: the case of Spain before World War I. *Journal of Economic Geography*, 2, 343-363.

Venables, A.J., (1995) Economic Integration and the Location of Firms. *American Economic Review*, 85(2): 296-300.

Venables, A.J., (2006) Shifts in economic geography and their causes. Discussion Paper. Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Viladecans, E., (2001) La concentración territorial de las empresas industriales: un estudio sobre la unidad geográfica de análisis mediante técnicas de econometría espacial. Document de treball 2001/2, Institut d'Economia de Barcelona.