KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No. 794

Analysis of Industrial Agglomeration Patterns:
An Application to Manufacturing Industries in Japan

Tomoya Mori and Tony E. Smith

November 2011

KYOTO UNIVERSITY

KYOTO, JAPAN

Analysis of Industrial Agglomeration Patterns: An Application to Manufacturing Industries in Japan

Tomoya Mori*†and Tony E. Smith‡

November 5, 2011

Abstract

The standard approach to studying industrial agglomeration is to construct summary measures of the "degree of agglomeration" within each industry and to test for significant agglomeration with respect to some appropriate reference measure. But such summary measures often fail to distinguish between industries that exhibit substantially different spatial patterns of agglomeration. In a previous paper, a cluster-detection procedure was developed that yields a more detailed spatial representation of agglomeration patterns (Mori and Smith [28]). This methodology is here applied to the case of manufacturing industries in Japan, and is shown to yield a rich variety of agglomeration patterns. In addition, to analyze such patterns in a more quantitative way, a new set of measures is developed that focus on both the global extent and local density of agglomeration patterns. Here it is shown for the case of Japan that these measures provide a useful classification of pattern types that reflect a number of theoretical findings in the New Economic Geography.

JEL Classifications : C49, L60, R12, R14

Keywords : Industrial Agglomeration, Cluster Analysis, Spatial Patterns of Agglomerations, New Economic Geography.

Acknowledgement: This research has been partially supported by the Kajima Foundation and the Grant in Aid for Research (Nos. 21243021, 21330055, 22330076 and Global COE Program: "Raising Market Quality: Integrated Design of Market Infrastructure") of Ministry of Education, Culture, Sports, Science and Technology of Japan.

*Institute of Economic Research, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501 Japan. Email: mori@kier.kyoto-u.ac.jp. Phone: +81-75-753-7121. Fax: +81-75-753-7198.

†Research Institute of Economy, Trade and Industry (RIETI), 11th floor, Annex, Ministry of Economy, Trade and Industry (METI) 1-3-1, Kasumigaseki Chiyoda-ku, Tokyo, 100-8901 Japan.

‡Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: tesmith@seas.upenn.edu. Phone: +1-215-898-9647. Fax: +1-215-898-5020.

1 Introduction

A procedure for identifying spatial patterns of industrial agglomeration was developed in Mori and Smith [28], hereafter referred to as [MS]. The present paper applies this procedure to the case of manufacturing industries in Japan. As mentioned in [MS], most studies of agglomeration focus on the overall degree of agglomeration in industries, and typically measure the discrepancy between industry-specific regional distributions of establishments/employment and some hypothetical reference distribution representing "complete dispersion."[1] But even if industries are judged to be similar with respect to these indices, their spatial patterns of agglomeration may appear to be quite different. Thus the main feature of our present approach is to develop explicit spatial representations of agglomeration patterns that allow more detailed types of spatial comparisons.

Toward this end, a second objective of the present paper is to propose a set of pattern measures that may facilitate such comparisons. The specific measures to be developed are largely inspired by theoretical results from the "new economic geography"(NEG)[2] where industrial location is modeled in continuous space.[3] Here it has been shown that the spatial structure of agglomeration and dispersion can change at different scales of analysis, depending on a host of factors including plant-level increasing returns, product differentiation and transport costs. These effects are well illustrated by considering the spatial effects of transportation costs in simple "core-periphery" models of industrial location (e.g., Tabuchi, 1998; Murata and Thisse, 2005). At very high levels of transport costs, the dispersion of consumers between the "core" and "periphery" regions leads to a corresponding dispersion of manufacturing. But as transport costs decrease and distance to consumers becomes less critical, manufacturing tends to concentrate (in

[1] Examples of such reference distributions are (1) the regional distribution of all-industry employment, used by Ellison and Glaeser [7], (2) the regional distribution all-industry establishments, used by Duranton and Overman [5], and (3) the regional distribution of economic area used by Mori, Nishikimi and Smith [26].

[2] See, e.g., Fujita, Krugman and Venables [10] and Combes, Mayer and Thisse [4] for an overview of the literature.

[3] See Fujita and Mori [12, 13] for a survey.

the core region). Finally, at even lower levels of transport costs, the reduction in commuting costs (together with congestion effects in the core region) can induce a second phase of manufacturing dispersion (popularly referred to as "re-dispersion" or "revival"). Indeed, these two dispersion patterns often appear to be exactly the same (i.e., a symmetric distribution of manufacturing between the two regions). But, in NEG models involving more general location spaces (e.g., Krugman [22]; Fujita and Mori [25]), these two dispersion phases have qualitatively different spatial patterns. In particular, while dispersion of manufacturing at high levels of transportation costs continues to be global (as in core-periphery models), the second phase of dispersion at low levels of transport costs is much more localized in nature, and perhaps better described as an expansion of existing core areas rather than re-dispersion to peripheral areas (Mori [25]).

Such theoretical findings raise important questions as to whether this diversity of patterns can in fact be identified empirically. Hence the specific measures proposed here are designed to quantify pattern differences both in terms of their global and local properties. While the details of these measures require a more formal definition and construction of agglomeration patterns, the basic ideas can be illustrated by a preview of the types of patterns we have identified for Japanese manufacturing industries. First, there are industries which clearly exhibit strong spatial concentration, such as the "motor vehicle, parts and accessories" industry. The agglomeration pattern derived for this industry is shown in Figure 4.12(b), where the areas marked in gray denote industrial agglomerations.[4] While some establishments of this industry are attracted to port cities along the northern coast, the main industrial concentration lies along the inland Industrial Belt extending westward from Tokyo to Hiroshima. Moreover, the individual clusters of establishments within this belt are seen to be densely packed from end to end. We describe this type of agglomeration pattern as "globally confined" and "locally dense" (here with respect to the Industrial Belt). In particular, this pattern is reminiscent of the type of "second-phase" dispersion of manufacturing identified

[4]See Section 4.3 below for a more detailed discussion of these figures.

in the NEG models described above. But even globally dispersed industries often form small agglomerations at local scales. For example, the agglomeration pattern for the "soft drinks and carbonated water products" industry shown in Figure 4.5(b) is spread throughout the nation, but exhibits a large number of local agglomerations. Such patterns, which we describe "globally dispersed" and "locally sparse," are closer in spirit to the "first-phase" dispersion of manufacturing in the NEG models above. Aside from these extremes, there are a variety of other patterns that can be identified, as discussed more fully in Sections 3 and 4.3 below. Finally, it is important to emphasize that the full range of patterns identified here actually bears a close relation to those identified in the new economic geography.[5] We return to this issue briefly in the concluding remarks.

The paper is organized as follows. In Section 2 below we develop the formal framework for analysis, and briefly sketch the cluster identification procedure developed in [MS]. This is followed in Section 3 with a development of our summary measures for analyzing and classifying the agglomeration patterns obtained. Finally, our application of these methods to manufacturing industries in Japan is developed in Section 4. The paper concludes in with brief discussions of related research in Section 5.

2 Identification of Industrial Clusters

As in [MS], we begin with a set, R, of relevant *regions* (municipalities), r, within which each industry can locate. An *industrial cluster* is then taken roughly to be a spatially coherent subset of regions within which the density of industrial establishments is unusually high. Since the explicit construction of such clusters will have consequences for the summary measures to be developed, it is appropriate to outline this construction more explicitly. The present notion of "spatial coherence" is taken to include the requirement that such regions be contiguous, and as close to one another as possible – where "closeness" is defined with respect to the relevant underlying road network. By

[5]See Krugman [22], Fujita and Mori [11], Fujita et al. [9], Ikeda, Akamatsu and Kono [17], and Akamatsu, Takayama and Ikeda [1] for the case of "globally dispersed"and "locally sparse"agglomeration patterns, and Mori [25] for the case of "globally confined"and "locally dense"agglomeration patterns.

using travel distances between regional centers, we define *shortest paths* between each pair of regions, r_i and r_j, to be sequences of intermediate regions, $(r_i, r_1, .., r_k, r_j)$ reflecting minimum travel distances with respect to the road network.[6] Our key requirement for spatial coherence of a cluster is that it be *convex* in the sense that it includes all shortest paths between its member regions. But as shown in [MS, Figure 5] this weaker notion of convexity can in principle allow "holes" in regional clusters. Hence unlike the usual notion of planar convexity with respect to Euclidean distance, these convex clusters can be more like "pretzels". However, by distinguishing regions inside clusters from those "outside" with respect to the boundary of the full regional system, R, it is possible to formalize a procedure for "solidifying" such clusters in a reasonable way. As shown in [MS], this *convex-solidification* procedure yields a class of sets in R, designated as *convex solids*,[7] which then constitute the desired class of candidate *clusters* for our purposes. Examples of such clusters (for the "livestock products" industry in Japan) are illustrated and discussed in Section 2.2 below.

2.1 Cluster Schemes

Most industries consist of multiple clusters in R that together define the agglomeration *pattern* for that industry. In fact, the spacing between such clusters is a topic of considerable economic interest (as discussed further in Section 5.1 below). Hence it is essential to model such patterns as explicit spatial arrangements of multiple clusters. The model proposed in [MS] is a *cluster scheme*, $\mathbf{C} = (R_0, C_1, .., C_{k_\mathbf{C}})$, that partitions R into one or more disjoint clusters (convex solids), $C_1, .., C_{k_\mathbf{C}}$, together with the residual set, R_0, of all non-cluster regions in R. The individual clusters are implicitly taken to be areas in R where industry density is unusually high. But within each cluster, C_j, all that is assumed for modeling purposes is that location probabilities for randomly sampled industrial establishments are uniform across the feasible locations in C_j. More precisely,

[6] Technically these shortest paths may in many cases be longer than actual shortest routes on the network. For additional details see Section 3.1 of [MS].

[7] In fact, it is shown in Property 3.5 of [MS] that for any initial set, $S \subset R$, this procedure generates the *smallest convex solid* containing S.

if the *feasible area* as defined in Section 4.1.2 below for locations in each region, $r \in R$, is denoted by a_r, so that the total area of C_j is given by $a_{C_j} = \sum_{r \in C_j} a_r$, then location probabilities in C_j are take to be uniform over a_{C_j}.[8] In particular, this implies that the conditional probability of an establishment locating in $r \in C_j$ given that it is located in C_j is simply a_r/a_{C_j}. With this assumption, the only unknown probabilities are the marginal location probabilities, $p_{\mathbf{C}}(j)$, for clusters C_j in $\mathbf{C}$. Hence each cluster scheme, $\mathbf{C}$, generates a possible *cluster probability model*, $p_{\mathbf{C}} = [p_{\mathbf{C}}(j) : j = 1, .., k_{\mathbf{C}}]$, of establishment locations for the industry.[9] If there are n establishments in the given industry, then each cluster probability model, $p_{\mathbf{C}}$, amounts formally to multinomial sampling model with sample size, n, and outcomes given by the $k_{\mathbf{C}} + 1$ sets in cluster scheme, $\mathbf{C}$, with respect to samples of size n. Finally, since the observed relative frequencies, $\widehat{p}_{\mathbf{C}} = [\widehat{p}_C(j) = n_j/n : j = 1, .., k_{\mathbf{C}}]$, of establishments in each cluster are well known to be the maximum-likelihood estimates of these (multinomial) probabilities, such estimates yield a family of well-defined candidate probability models for describing the agglomeration patterns of each industry.

2.2 Cluster-Detection Procedure

The key question remaining is how to find a "best" cluster-scheme for capturing the observed distribution of industry establishments. It is argued in [MS] that the *Bayes Information Criterion* (BIC) offers an appropriate measure of model fit in the present setting. In particular, it is shown in [MS] that for any given cluster scheme, $\mathbf{C}$, the (multinomial) log-likelihood of $\widehat{p}_{\mathbf{C}}$ is given by

$$L_{\mathbf{C}}(\widehat{p}_{\mathbf{C}}) = \sum_{j=0}^{k_{\mathbf{C}}} n_j(x) \ln\left(\frac{n_j(x)}{n}\right) \; + \; \sum_{j=0}^{k_{\mathbf{C}}} \sum_{r \in C_j} n_r \ln\left(\frac{a_r}{a_{C_j}}\right) \tag{1}$$

[8] Feasible area is here taken to be *economic area*.

[9] This probability model is completed by the condition that $p_C(R_0) = 1 - \Sigma_j p_{\mathbf{C}}(j)$.

and that in terms of $L_{\mathbf{C}}(\widehat{p}_{\mathbf{C}})$, the appropriate value of BIC is given for each candidate cluster scheme, $\mathbf{C}$, by

$$BIC_{\mathbf{C}} = L_{\mathbf{C}}(\widehat{p}_{\mathbf{C}}) - \frac{k_{\mathbf{C}}}{2}\ln(n)\ . \tag{2}$$

Hence BIC is a "penalized likelihood" measure, where the second term in (2) essentially penalizes cluster schemes with large number of clusters, $k_{\mathbf{C}}$, to avoid "over fitting" the data.

Given this criterion function, the *cluster-detection procedure* developed in [MS] amounts to a systematic way of searching the space of possible cluster probabity models to find a cluster scheme, $\mathbf{C}^*$, with a maximum value of $BIC_{\mathbf{C}^*}$.[10] While the details of this search procedure will play no role in the present analysis, the results of this procedure for Japanese industries will play a crucial role. Hence it is appropriate to illustrate these results in terms of the "livestock products" industry in Japan, shown in Figure 4.6 in Section 4.3.1 below.

Here Figure 4.6(a) shows the relative density of "livestock products" establishments in each municipality of Japan, where darker patchs correspond to higher densities.[11] Figure 4.6(b) shows the cluster scheme, $\mathbf{C}^*$, that was produced for the "livestock products" industry by this cluster-detection procedure. Here it is seen that not all isolated patches of density are clusters. But the highest density areas do indeed yield significant clusters. Notice also that the convex solidification procedure above has produced easily recognizable clusters that do seem to reflect the shapes of these high density areas.

2.3 A Test of Significant Clustering

Finally it should be emphasized that even random locational patterns will tend to exhibit some degree of clustering. So there remains the statistical question of whether the "locally best" cluster scheme, $\mathbf{C}^*$, found for an industry by the above procedure

[10] However, it should be emphasized that this space of probability models is very large, and hence that one can only expect to find *local* maxima (with respect to the particular perturbations defined by the search procedure itself).

[11] These municipalities are mapped in Figure 4.1 below.

is significantly better (in terms of BIC values) than would be expected in a random location pattern. A test is developed in [MS, Section 4.3] that will be used in the present application. Basically, a "random" location pattern is taken to be one in which location probabilities in all regions, $r \in R$, are proportional to their feasible areas, a_r. Hence a Monte Carlo test can be constructed by (i) generating N random location patterns for the establishments of a given industry, (ii) determining the locally optimal values, say BIC_s^*, for each simulated pattern, $s = 1, .., N$, and (iii) comparing the value, $BIC_{\mathbf{C}^*}$, with this sampling distribution of BIC values. If $BIC_{\mathbf{C}^*}$ is sufficiently large (say in the top 1% of these values), then one may conclude that the clustering captured by $\mathbf{C}^*$ is significantly higher than what would be expected under randomness. Otherwise, $\mathbf{C}^*$ is said to involve *spurious clustering*. Results of this testing procedure for the present application will be discussed in Section 4.2 below.

3 Measures for Classifying Agglomeration Patterns

As emphasized in the Introduction, the main strength of our cluster detection approach is to identify cluster schemes in a manner that preserves their two-dimensional spatial properties. By so doing, it is possible to analyze the spatial patterns of industrial agglomerations in more detail. As we will see for the case of Japanese manufacturing industries in Section 4 below, agglomerations of given industries often tend to concentrate within specific subregions of the nation, i.e., are themselves "spatially contained." Hence our first task below is to construct an operational definition of such containments, designated as the *essential containment* (*e-containment*) for each industry. Our next task is to construct a meaure of the relative size of these e-containments, designated as the *global extent* of the industry. Industries with small global extents can be regarded as relatively "confined," and those with large global extents can be regarded as relatively "dispersed." Finally, industries can also differ with respect to their patterns of agglomeration *within* these e-containments. Some patterns may be "dense" and others "sparse." To compare such patterns, we construct a measure of the *local density* of agglomerations within

each e-containment. This will yield a useful classification of agglomeration patterns ranging from *maximally concentrated* patterns with agglomerations densely packed in small e-containments to *minimally concentrated* patterns with agglomerations sparsely distributed over large e-containments.

3.1 Essential Containment

To formalize the notion of an industry's essential containment, we start by assuming that an optimal cluster scheme, $\mathbf{C} = \mathbf{C}^*$, has been indentified for the industry.[12] The main idea is to identify an appropriate subset of "most significant" clusters in $\mathbf{C}$, and then take *essential containment* to be the convex solidification of this set of clusters in R. To identify "most significant" clusters, we proceed recursively by successively adding those clusters in $\mathbf{C}$ with maximum incremental contributions to *BIC*.[13] This recursion starts with the "empty" cluster scheme represented by $\mathbf{C}_0 \equiv \{R_{0,0}\}$ where $R_{0,0}$ denotes the full set of regions, R. If the set of (non-residual) *clusters* in $\mathbf{C}$ is denoted by $\mathbf{C}^+ \equiv \mathbf{C} - \{R_0\}$, then we next consider each possible "one-cluster" scheme created by choosing a cluster, $C \in \mathbf{C}^+$, and forming $\mathbf{C}_0(C) = \{R_{0,0}(C), C\}$, with $R_{0,0}(C) = R_{0,0} - C$. The "most significant" of these, denoted by $\mathbf{C}_1 = \{R_{1,0}(C), C_{1,1}\}$, is then taken to be the cluster scheme with the *maximum BIC value* (defined below). If this is called *stage* $t = 1$, and if the *most significant cluster scheme* found at each stage $t \geq 1$ is denoted by

$$\mathbf{C}_t \equiv \{R_{t,0}, C_{t,1}, \ldots, C_{t,t}\} \ , \tag{3}$$

then the recursive construction of these schemes can be defined more precisely as follows.

For each $t \geq 1$ let $\mathbf{C}^+_{t-1}$ denote the (non-residual) clusters in $\mathbf{C}_{t-1}$ (so that for $t = 1$ we have $\mathbf{C}^+_{t-1} = \mathbf{C}^+_0 = \varnothing$), and for each cluster not yet included in $\mathbf{C}_{t-1}$, i.e., each

[12] For notational simplicity we drop the asterisk in $\mathbf{C}^*$.

[13] At this point it should be emphasized that the following procedure for identifying "significant clusters" in $\mathbf{C}$ is different from the one used to indentify $\mathbf{C}$ in Section 2.2. In particular, the only candidate clusters now being considered are those in $\mathbf{C}$ itself.

$C \in \mathbf{C}^+ - \mathbf{C}^+_{t-1}$, let $\mathbf{C}_{t-1}(C)$ be defined by,

$$\mathbf{C}_{t-1}(C) = (R_{t-1,0}(C), C_{t-1,1}, \ldots, C_{t-1,t-1}, C) \ , \tag{4}$$

where

$$R_{t-1,0}(C) = R_{t-1,0} - C \ . \tag{5}$$

Then the *most significant additional cluster*, $C_t (\equiv C_{t,t})$ $(\in \mathbf{C}^+ - \mathbf{C}^+_{t-1})$, at stage $t \geq 1$ is defined by

$$C_t \equiv \arg \max_{C \in \mathbf{C}^+ - \mathbf{C}^+_{t-1}} L\left(\widehat{p}_{\mathbf{C}_{t-1}(C)} | \mathbf{C}_{t-1}\right) \ , \tag{6}$$

where $L\left(\widehat{p}_{\mathbf{C}_{t-1}(C)} | \mathbf{C}_{t-1}\right)$ is the *estimated maximum log-likelihood value* for model $p_{\mathbf{C}_{t-1}(C)}$ given [in a manner paralleling expression (1) above] by

$$L\left(\widehat{p}_{\mathbf{C}_{t-1}(C)} | \mathbf{C}_{t-1}\right) = \sum_{C' \in \mathbf{C}_{t-1}(C)} n_{C'} \ln\left(\frac{n_{C'}}{n}\right) + \sum_{C' \in \mathbf{C}_{t-1}(C)} \sum_{r \in C'} n_r \ln\left(\frac{a_r}{a_{C'}}\right) \ , \tag{7}$$

where $n_{C'} \equiv \sum_{r \in C'} n_r$ and $n \equiv \sum_{r \in R} n_r$. Thus, at each stage $t \geq 1$ the likelihood-maximizing cluster, C_t, is removed from the residual region, $R_{t-1,0}$, and added to the set of significant clusters in $\mathbf{C}_{t-1}$. The resulting BIC value at each stage t is then given by

$$BIC_{\mathbf{C}_t} = L_{\mathbf{C}_t} - \frac{t}{2} \ln(n) \tag{8}$$

with

$$L_{\mathbf{C}_t} = \sum_{C \in \mathbf{C}_t} n_C \ln\left(\frac{n_C}{n}\right) + \sum_{C \in \mathbf{C}_t} \sum_{r \in C} n_r \ln\left(\frac{a_r}{a_C}\right) \tag{9}$$

Finally, the *incremental contribution* of each new cluster, C_t, to BIC is given by the increment for its associated cluster scheme, $\mathbf{C}_t$, as follows:

$$\triangle BIC_t \equiv BIC_{\mathbf{C}_t} - BIC_{\mathbf{C}_{t-1}} \tag{10}$$

To identify the relevant set of "significant clusters" in $\mathbf{C}$, it would thus seem most

natural to simply add clusters as long as the increments are positive. But from the original construction of $\mathbf{C}$ it should be clear that these increments may often be positive for *all* $t = 1, .., k_{\mathbf{C}}$. Hence our first requirement for significance of cluster C_t is that it yield a "substantial" increment to *BIC*. One hypothetical illustration with $k_{\mathbf{C}} = 7$ is given in Figure 3.1(a) below, where each successive increment to *BIC* is seen to be positive [and where the values on the horizontal axis can be ignored for the moment]. By the nature of our recursive procedure, it can be expected that the first increment ($t = 1$) will be the largest, and that successive increments will continue to diminish in size.[14] In the example shown, it appears that the increments for $t = 2, 3$ are comparable to $t = 1$, but that there is a noticeable decrease at $t = 4$ and beyond. Hence one simple criteria for a "substantial increment," $\triangle BIC_t$, would be to require that it be at least some specified fraction, μ, of $\triangle BIC_1$.[15] In terms of this criterion, the procedure would stop at the first stage, t^e, where additional increments fail to satisfy this condition, i.e., where $\triangle BIC_{t^e+1} < \mu \triangle BIC_1$.

Figure 3.1 here

But while this *substantial-increment condition* provides a reasonable criterion for identifying the set of most significant clusters with respect to *BIC*, such clusters may in some cases represent only a small subset of all clusters in $\mathbf{C}$. More importantly, they may represent only a small portion of all *establishments* in such clusters. Hence, if the "essential containment" for the industry is to include a substantial portion of these *agglomeration establishments*, then it is desirable to impose an additional condition on the stopping rule above. In particular, if the *share* of agglomeration establishments in each cluster scheme, $\mathbf{C}_t$, of expression (3) is denoted by

$$s\left(\mathbf{C}_t\right) = \frac{\sum_{C \in \mathbf{C}_t^+} n_C}{\sum_{C \in \mathbf{C}^+} n_C}, \tag{11}$$

then it is reasonable to require that the above recursive procedure continue until this

[14] This situation is somewhat analogous to successive increments in adjusted R-square resulting from a foward stepwise regression procedure.

[15] The values $\mu = .03$ and $\mu = .05$ were selected for our application in Section 4.3 below..

share has reached some specified fraction, ζ, of all agglomeration establishments.[16] If the desired *stopping point* is again denoted by $t^e \in \{1,..,k_{\mathbf{C}}\}$, then this modified stopping rule can be formalized as follows: (i) if $k_{\mathbf{C}} = 1$, set $t^e = 1$; (ii) if $k_{\mathbf{C}} \geq 2$, and if for the given pair of threshold fractions, $\mu, \zeta \in (0,1)$, there is at least one stage, $t \in \{2,3,\ldots,k_{\mathbf{C}}-1\}$ satisfying the following two conditions,

$$\triangle BIC_{t+1} < \mu \triangle BIC_1 \qquad [\textit{substantial-increment condition}] \tag{12}$$

$$s(\mathbf{C}_t) \geq \zeta \qquad [\textit{substantial-establishments condition}] \tag{13}$$

then choose t^e to be the *smallest* of these; and otherwise, (iii) set $t^e = k_{\mathbf{C}}$. This stopping rule is again illustrated by Figure 3.1 above where hypothetical shares of agglomeration establishments, $s(\mathbf{C}_t)$, are shown at each stage, $t = 1,..,k_{\mathbf{C}}(=7)$, on the horizontal axis. Hence if $\zeta = .80$ and if $\triangle BIC_t / \triangle BIC_1$ first falls below the specified value of μ at $t = 4$ in Figure 3.1(a), then $t^e = 3$. However, if the shares if agglomeration establishments are as shown in Figure 3.1(b) [which uses the same BIC increments as in Figure 3.1(a)], then the procedure will not terminate until stage $t^e = 5$.

If the set of *essential clusters* in $\mathbf{C}$ is now defined to be $\mathbf{C}^e = \mathbf{C}^+_{t^e}$, then the desired *essential containment* (*e-containment*), $ec(\mathbf{C})$, for an industry with cluster scheme $\mathbf{C}$ is taken to be the smallest solid convex set in R containing $\mathbf{C}^e$, i.e., the convex solidification of $\mathbf{C}^e$.[17]

These concepts can be illustrated by the stylized location patterns in Figure 3.2 below. For example, if the relevant cluster scheme, $\mathbf{C}$, for a given industry corresponds to the five clusters (shown in black) in Figure 3.2(a), and if the subset of essential clusters, $\mathbf{C}^e$, consists of the three largest clusters on the left, then the essential containment, $ec(\mathbf{C})$, for this industry is given by the filled square containing these three clusters. Similar interpretations can be given to the filled rectangles of Figures 3.2(b,c,d).

Figure 3.2 here

[16] Note that this condition could also be formulated in terms of *agglomeration employment*.

[17] In terms of the d-convex solidification operator, σc_d, defined in expression (26) of [MS] (with respect to shortest-path travel distance, d), the formal definition of *e-containment* is given by $ec(\mathbf{C}) = \sigma c_d(\mathbf{C}^e)$.

3.2 Global Extent and Local Density

With these definitions we next seek to compare e-containments for different industries in terms of their relative sizes. In order to reflect the actual spatial extent of such containments, it is now more appropriate to measure "size" in terms of total *geographic area* rather than the more limited notion of *feasible area* (employed for modeling the potential locations of individual establishments, as in Sections 2.1 above). Hence if we now let A to denote *geographic area*, then the economic areas for *basic regions* (a_r), *clusters* (a_C), and the *entire nation* (a), are here replaced by A_r, A_C, and A, respectively. With these conventions, the *global extent* (GE) of an industry is now taken to be simply the total area of its e-containment, $ec(\mathbf{C})$, relative to that of the entire nation, i.e.,

$$GE(\mathbf{C}) = \frac{\sum_{r \in ec(\mathbf{C})} A_r}{A} \in (0, 1] \, . \tag{14}$$

Industries with relatively small global extents might be classified as "globally confined" industries [illustrated by the industries in Figures 3.2(a,c)]. Similarly, industries with substantially larger global extents might be classified as "globally dispersed" industries [illustrated by those in Figures 3.2(b,d)].[18]

Finally, we consider the relative denseness of essential clusters within the e-containment for each industry. As a parallel to global extent, we now define the *local density* (LD) of a given industry to be simply the total area of its essential clusters, $\mathbf{C}^e$, relative to that of its e-containment, $ec(\mathbf{C})$, i.e.,

$$LD(\mathbf{C}) = \frac{\sum_{r \in \mathbf{C}^e} A_r}{\sum_{r \in ec(\mathbf{C})} A_r} \in (0, 1] \, . \tag{15}$$

Industries with a relatively high density of agglomerations in their e-containments might be classified as "locally dense" industries [illustrated by the industries in Figures 3.2(a,b)]. Similarly, industries with a substantially lower density of agglomerations in

[18] One might consider more exact classifications, such as $GE < 1/2$ for "globally confined" and $GE \geq 1/2$ for "globally dispersed." But in our view, the appropriate ranges of GE may often be context dependent.

their e-containments might be classified as "locally sparse" industries [illustrated by those in Figures 3.2(c,d)].

More generally, Figure 3.2 is intended to summarize the main features of this classification system. First, the concept of *essential containment* is designed to capture the region of most significant agglomeration for an industry, while at the same time including most of its establishments. This is illustrated in each of the figures by filled regions containing the largest agglomerations for the cluster schemes shown. In each case, the "outlier" agglomerations excluded from this region are implicitly assumed to be less significant, both in terms of their contributions to BIC and their overall share of establishments for the industry.

In addition, Figure 3.2 illustrates the four possible extreme cases in this classification system. As already mentioned, *maximal spatial concentration* in this system corresponds to the case of globally confined and locally dense agglomeration patterns, such as Figure 3.2(a). The opposite extreme of *minimal spatial concentration* is characterized most naturally by globally dispersed and locally sparse agglomeration patterns, such as Figure 3.2(d).[19] The two "intermediate" extremes are somewhat more difficult to interpret, but do indeed occur (as will be seen in Sections 4.3.2 and 4.3.3 below). Here it should be noted that these intermediate extremes do have implications for the overall *size* of the industries involved. In particular, only industries with many establishments can exhibit dense patterns of significant agglomerations over large areas [such as Figure 3.2(b)], and only industries with small numbers of establishments can exhibit sparse patterns of agglomerations in confined areas [such as Figure 3.2(c)]. Additional features and examples of this classification system will be developed in Section 4 below.

[19] However, it should be borne in mind that "minimal spatial concentration" in our present framework is not the same as "complete spatial randomness." In particular, since *all* spatial patterns are assumed to have passed the "spurious cluster" test developed above, even globally dispersed and locally sparse patterns must contain some significant degree of local clustering.

4 Application

In this section, we apply the above set of cluster-analytic tools to study the agglomeration patterns of manufacturing industries in Japan. We begin in Section 4.1 with a description of the relevant data for analysis. This is followed in Section 4.2 with a summary of results for the spurious-cluster test in Section 2.3 above. The classification scheme developed in Section 3 above is then given an operational form for the present application. Finally, this classification scheme is illustrated by means of a number of selected examples in Section 4.3.

4.1 Data for Analysis

The data required for this application includes both quantitative descriptions of the relevant system of regions and the class of industries to be studied. We consider each of these data types in turn.

4.1.1 Basic Regions

The relevant notion of a "basic region" for this analysis is taken to be the *shi-ku-cho-son* (SKCS), which is a municipality category equivalent to a city-ward-town-village in Japan. The specific SKCS boundaries are taken to be those of October 1, 2001.[20] While there are a total of 3363 SKCS's in Japan, we only consider 3207 of these (as shown in Figure 4.1), namely those that are *geographically connected to the major islands of Japan* (*Honshu, Hokkaido, Kyushu and Shikoku*) *via a road network*. This avoids the need for ad-hoc assumptions regarding the effective distance between non-connected regions.

Figure 4.1 here

The only exception here is Hokkaido, which is one of the four major islands (refer to Figure 4.1), but is disconnected from the road network covering the other three. Given

[20]The data source for these SKCS boundaries is the Statistical Information Institute for Consulting and Analysis [31, 32].

its size (217 SKCS's), as seen in Figure 4.1, we still include Hokkaido as a potential location for establishments. Aside from this exceptional case, we adopt the following conventions. First, while we allow establishments to locate freely within the 3207 municipalities, we do not allow the formation of any clusters including basic regions in both Hokkaido and other major islands.[21] Second, *e-containments* for each industry are obtained as the union of the two convex solidified subsets of essential clusters within and without Hokkaido [see, for example, the cases of "soft drinks, and carbonated water," "livestock products," and "sliding doors and screens," shown in Figures 4.5(c), 4.6(c) and 4.7(c), respectively, in Section 4.3 below].

4.1.2 Economic Area of Regions

To represent the areal extent of each basic region we adopt the notion of "economic area," obtained by subtracting forests, lakes, marshes and undeveloped area from the total area of the region (available from the Statistical Information Institute for Consulting and Analysis[31, 32]).[22] The economic area of Japan as a whole (120,205km^2) amounts to only 31.8% of total area in Japan. Among individual SKCS's this percentage ranges from 2.1% to 100%, with a mean of 48.5%. Not surprisingly, those SKCS's with highest proportions of economic area are concentrated in urban regions. In this respect, our present approach is relatively more sensitive to clustering in rural areas.[23]

[21] In terms of our δ-neighborhood definition in Section 4.2.2 of [MS], the distances between Hokkaido regions and those of the major islands are implicitly assumed to exceed δ.

[22] There is of course a certain degree of interdependence between the size of economic areas and the presence of industries in those areas. In particular, industrial growth in a region may well lead to a gradual increase in the economic area of that region (say by land fills or deforestation). But to capture agglomeration patterns at a given point in time, we believe that it is more reasonable to adopt economic area than total area as the potential location space for establishments. In Japan, for example, it is doubtful that mountainous forested regions (which account for 98% of non-economic areas) can be easily be made available for industrial location in the short run.

[23] In other words, for any given number of firms, n_r, in a basic region r, our clustering algorithm implicitly regards n_r as a more significant concentration in regions with smaller economic areas (other things being equal).

4.1.3 Interregional Distances

The travel distance between each pair of neighboring SKCS's is computed as the length of the shortest route between their municipality offices along the road network.[24] From the computed pairwise distances between neighboring (contiguous) SKCS's, the *shortest-path distances* (and associated sequences of neighboring SKCS's) are computed in terms of expression (15) in [MS].[25] While there is of course some degree of interdependency between industrial locations and the road network, the spatial structure of this network is mainly determined by topographical factors. With respect to topography, it should also be noted that since Japan is quite mountainous with very irregular coastlines (along which the majority of industrial sites are found), shortest-path distances are generally much longer than straight-line distances. Hence the use of shortest-path distances is particularly important for countries like Japan.

4.1.4 Industry and Establishments Data

Finally, the industry and establishments data used for this analysis is based on the Japanese Standard Industry Classification (JSIC) in 2001. Here we focus on three-digit manufacturing industries, of which 163 industrial types are present in the set of basic regions chosen for this analysis.[26] The establishment counts (n) across these 163 industries is taken from the Establishment and Enterprise Census of Japan [20] in 2001, and the frequency distribution of these counts is shown in Figure 4.2. The mean and median establishment counts per industry are respectively 3958 and 1825. In addition, 147 (90%) of these industries have more than 100 establishments, and 125 (77%) have more than 500 establishments.

[24] This road network data is taken from Hokkaido-chizu Co. Lit. [15], and includes both prefectural and municipal roads. However, if a given municipality office is not on one of these roads, then minor roads are also included.

[25] As noted in Section 3.1 of [MS], shortest-path distances are always at least as large as shortest-route distances. But in the present case, shortest-path distance appears to approximate shortest-route distance quite well. For the distribution of ratios of short-path over shortest-route distances across all 4,491,991 relevant pairs of municipalities, the median and mean are both equal to 1.14. In fact, the 99.5 percentile point of this distribution is only 1.28.

[26] More precisely, out of the 164 industrial types in Japan, all but one have establishments in at least one of our basic regions.

Figure 4.2 here

4.2 Tests of Spuriousness of Cluster Schemes

Using the cluster-detection procedure developed in Section 2.2 above, optimal cluster schemes, $\mathbf{C}_i^*$, were identified for each industry, $i = 1, .., 163$.[27] Each cluster scheme, $\mathbf{C}_i^*$, was then tested for spuriousness using the testing procedure developed in Section2.2.[28] Among the 163 industries studied, the null hypothesis of complete spatial randomness (Section 2.3 above) was strongly rejected for 154 (95%) of these industries, with p-values virtually zero. For the remaining nine industries, this null hypothesis could not be rejected at the .01 level. The main reason for non-rejection in these cases [which include seven arms-related industries (JSIC331-337), together with tobacco manufacturing (JSIC135) and coke (JSIC213)], appears to be the small size of these industries, with $n < 40$ in all cases.[29] In view of these findings, we chose to drop the nine industries in question and focus our subsequent analyses on the 154 industries exhibiting significant clustering.

For these 154 industries, Figure 4.3 shows the frequency distribution of the share of establishments for each industry i that are included in the clusters of its cluster scheme, $\mathbf{C}_i^*$. These shares range from 39.1% to 100% with a median [resp., mean] share of 95.2% [resp., 93.6%]. The industries with the smallest shares of establishments in clusters are typically those which exhibit the weakest tendency for clustering. For

[27] The computation time required to identify the best cluster scheme for a given industry varies depends on the number and the spatial distribution of establishments of this industry, and of course, computational environment. Other things being equal, an industry with a smaller number of establishments requires a smaller amount of time. Computation takes more time for an industry with spatially larger clusters, e.g., in the case of industrial belt (refer to Section 4.3.4). In our computational environment [Intel C++ version 9.1 on a computer with Xeon Westmere-EP) 2.8GHz with 4GB random access memory per computational core), the computational time for detecting the best cluster scheme ranges from less than a minute to about two hours. However, it should be noted that computational time depends strongly on the implementation of the algorithms. Since the computational efficiency is not the main theme of the present paper, there should be a large room for improvement on the actual implementation of the algorithms.

[28] These tests of spuriousness were based on the BIC values for a sample of 1000 completely random location patterns for each industry.

[29] These industries are also rather special in other ways. Tobacco manufacturing and arms-related industries are highly regulated industries, so that their location patterns are not determined by market forces. Finally, Coke is a typical declining industry in Japan (steel industries have gradually replaced coke production by less expensive powder coal after the 1970s).

instance, "paving materials" industry (JSIC215) and "sawing, planning mills and wood products" industry (JSIC161) have 39.1% and 54.0% of their establishments in the clusters, respectively. Since both of these industries are typically sensitive to transport costs, their establishment locations tend to reflect population density.

Figure 4.3 here

4.3 Classification of Cluster Patterns

Figure 4.4 plots LD versus GE for each of 154 industries (with non-spurious clusters) under four different sets of threshold levels, μ and ζ [see expressions (12) and (13) above]. The patterns are essentially the same for a reasonable range of μ and ζ values, although the range of (GE, LD) pairs tends to become more diverse for smaller values of ζ. In particular, there is seen to be wide variation in both measures, i.e., in both the global extent and local density of cluster schemes across industries. Note also that their is no clear relationship between them, indicating that all four extremes in Figure 3.2 do in fact occur.[30] However, the overall dispersion of (GE, LD) pairs appears to be relatively more sensitive to values of ζ than μ. For example, under $\zeta = 0.8$ [Diagrams (a) and (c)], there are a few industries in the northwest section of the diagram, but not under the larger value, $\zeta = 0.9$ [Diagrams (b) and (d)]. Since these industries exhibit a high degree of spatial concentration, i.e., they have only a few significant clusters, the inclusion of only a single additional cluster can dramatically affect the size of their e-containment, and hence their global extent. For example, in Section 4.3.4 below, Figures 4.9(c) and 4.10 show the essential containment of "leather gloves and mittens" (JSIC245) under $\zeta = 0.8$ and $\zeta = 0.9$ (with $\mu = 0.03$), respectively. In the latter case, the essential containment contains a large vacant area since it includes a remote cluster in Tokyo, while the former captures a more compact and highly specialized region around Hikita-Ohuchi-Shiratori. Note also that a visual comparison of JSIC245 in Figure 4.9(c)

[30] The relative positions of Diagrams (a) through (d) in Figure 3.2 are arranged to match the relative positions in each diagram of Figure 4.4. In particular, globally confined patterns in Figures 3.2(a,c) [resp., locally dense patterns in Figures 3.2(a,b)] are found in western [resp., northern] part of each diagram in Figure 4.4.

with "motor vehicle, parts and accessories" (JSIC311) in Figure 4.12(c) suggests that the former is more "spatially concentrated," even though the latter appears to be "closer" to the maximally-concentrated northwest corner of Figure 4.4(a). Hence it should also be clear that even these two measures, GE and LD, taken together can be expected to provide only a rough classification of spatial-concentration types.

Figure 4.4 here

With these general observations in mind, it is of interest to consider more detailed examples of industries with cluster schemes exhibiting a variety of (GE, LD) combinations. Here we focus mainly on the case of Figure 4.4(a) [$\mu = 0.03$ and $\zeta = 0.8$] which is seen to exhibit the widest variation of GE and LD values. Figures 4.5 through 4.13 each focus on a different industry. For each industry i, the associated figure displays its density of establishments in each basic region (Diagram, a), the spatial pattern of clusters in its derived cluster scheme, $\mathbf{C}_i^*$ (Diagram, b), and the essential containment, $ec(\mathbf{C}_i^*)$, of this cluster scheme (Diagram, c). In Diagram (a), basic regions with higher densities of establishments are shown as darker. In Diagram (b), the individual clusters in scheme $\mathbf{C}_i^*$ are represented by enclosed gray areas. The portion of each cluster in lighter gray shows those basic regions which contain no establishments (but are included in $\mathbf{C}_i^*$ by the process of convex solidification). Finally, the hatched area in Diagram (c) depicts the e-containment, $ec(\mathbf{C}_i^*)$, of this cluster scheme.

4.3.1 Globally Dispersed and Locally Sparse Patterns

Industries with relatively high values of GE and low values of LD [near the southeast corner of Figure 4.4(a)] can be described as exhibiting patterns of agglomeration that are both "globally dispersed and locally sparse." A clear example is provided by the "soft drinks and carbonated water" industry (JSIC131) shown in Figure 4.5 [with $GE = 0.589$ and $LD = 0.133$]. Bottled/packed soft drinks are weight/bulk-gaining products. Thus to minimize transport costs, establishments in this industry are naturally attracted to individual market locations, resulting in a pattern of global dispersion. In addition,

the individual clusters shown in Figure 4.5(b) appear to be locally concentrated, due to scale economies of production combined with relatively modest needs for land. Thus in terms of total area occupied, this pattern of clusters is relatively sparse.

Figure 4.5 here

A second example [mentioned in the Introduction] is provided by the "livestock products" industry (JSIC121) depicted in Figure 4.6 [with $GE = 0.771$ and $LD = 0.281$]. Here the perishable nature of livestock products again leads to market-oriented location behavior, and hence to global dispersion. But in this case, the extensive land requirements for livestock production produce higher local densities in terms of area occupied, and thus result in larger clusters than JSIC131 [as seen in Figure 4.6(b)].

Figure 4.6 here

4.3.2 Globally Dispersed and Locally Dense Patterns

Industries with both high values of GE and LD [near the northeast corner of Figure 4.4(a)] can be described as exhibiting patterns of agglomeration that are "globally dispersed and locally dense." Such industries are by definition present almost everywhere, and can equivalently be described as *ubiquitous industries*. While there are no extreme examples in Figure 4.4(a), one relatively ubiquitous example is provided by the "sliding doors and screens" (JSIC173) [with $GE = 0.777$, $LD = 0.473$]. As seen in Figure 4.7(a), the establishments of this industry are indeed found almost everywhere, with clusters densely distributed throughout the nation [Figure 4.7(b)]. Such products are often custom made and require face-to-face contact with customers. Hence there are strong market-attraction forces that contribute to the ubiquity of this industry. In such cases, the local density of clusters tends to correspond roughly to that of population.

Figure 4.7 here

It is also of interest to note (as mentioned at the end of Section 3.2) that such ubiquitous industries are by their very nature quite large in terms of establishment numbers.

In the present case, industry JSIC173 has 15,363 establishments, which is well above the mean of 4189 for all industries (with no spurious clusters, i.e., exhibiting significant agglomeration). In terms of establishments in clusters, JSIC173 has 13,565 establishments relative to a mean of only 3896 for all industries.

4.3.3 Globally Confined and Locally Sparse Patterns

The opposite extreme of "globally confined and locally sparse" agglomeration patterns [in the southwest corner of Figure 4.4(a)] is well illustrated by the "ophthalmic goods" (JSIC326) [with $GE = 0.166$ and $LD = 0.139$]. As seen in Figure 4.8(a) this industry has only a small number of establishments (located mainly between Tokyo and Osaka), with a disproportionate concentration in the small town, Sabae with population of 65,000. In fact, this single remote town accounts for more than 90% of the national market share in ophthalmic goods. As with many specialized industries, the location pattern of this industry is governed more by historical circumstances than economic factors. In terms of establishment numbers, such industries are necessarily small in size. In the present case, JSIC326 has only 1139 establishments, which is well below the mean of 4188 for all industries (as above). Even given the fact that all 1139 establishments are in clusters, this number is still well below the mean of 3896 for all industries (as above).

Figure 4.8 here

A similar example of this pattern is the "leather gloves and mittens" industry (JSIC245) depicted in Figure 4.9 [with $GE = 0.019$ and $LD = 0.418$]. Like JSIC326, this industry is not concentrated in large cities. Rather, its major concentration (accounting for 90% of the leather glove market in Japan) is confined to a cluster of three small towns, Hikita-Ohuchi-Shiratori with population of 38,000, shown in Figure 4.9(b).

Figure 4.9 here

Here it is of interest to note that while the value of LD for JSIC245 seems relatively large compared to JSIC326 above, this is mostly due to its small e-containment, as

reflected by its low level of GE relative to JSIC326 [compare Figures 4.8(c) and 4.9(c)]. When GE is very small for an industry, its value of LD is necessarily sensitive to the number of clusters in its e-containment.

In addition, it is also important to note that for globally confined industries with few clusters (such as JSIC245 and JSIC326), the values of GE and LD are both quite sensitive to the cut-off criteria, μ and ζ, in (12) and (13), respectively. As one illustration, Figure 4.10 shows the essential containment of JSIC245 obtained with $\zeta = 0.9$ rather than $\zeta = 0.8$ as in Figure 4.9(c). While this higher value of ζ allows the inclusion of only one additional cluster, the location of this cluster in Tokyo leads to the inclusion of a large vacant area between Osaka and Tokyo in the resulting convex solidification of these clusters.

Figure 4.10 here

A final example is provided by the larger "publishing industry" (JSIC192) depicted in Figure 4.11 [with $GE = 0.342$ and $LD = 0.232$]. Unlike JSIC326 and JSIC245, publishing is a typical "urban-oriented" industry with a location pattern generally reflecting urban density. As seen in Figure 4.11(b) this pattern is more concentrated toward the Pacific coast area between Tokyo and Osaka, with a narrow band stretching beyond Osaka to include the major metro areas further west (Kobe, Okayama, Hiroshima, and Fukuoka).

Figure 4.11 here

4.3.4 Globally Confined and Locally Dense Patterns

Finally, as mentioned in Section 3 above, those industries with agglomeration patterns that are both "globally confined and locally dense" [i.e., in the northwest corner of Figure 4.4(a)] constitute the single most spatially concentrated class of industries. Such industries are well illustrated by the example used in the Introduction, namely the "motor vehicles, parts and accessories" (JSIC311) in Figure 4.12 [with $GE = 0.221$ and $LD = 0.751$]. A comparison of the e-containment for this industry in Figure 4.12(c)

with that of the urban-oriented publishing industry in Figure 4.11(c) shows that JSIC311 again follows the chain of large metro areas extending westward from Tokyo through Osaka to Hiroshima. But here the containment is even more concentrated along this chain, and coincides with the Industrial Belt that constitutes the manufacturing core of Japan. This manufacturing core is in fact dominated by the major auto assembly plants in this industry, which by definition produce weight/bulk-gaining products requiring proximity to consumers in the metro centers. Moroever, the chain of contiguous clusters seen in Figure 4.12(b) essentially fills in the gaps between these metro centers, creating the effect of a single "megalopolis." The outputs of JSIC311 provide an important clue to the nature of this "filling-in" process. In particular, "parts and accessories" are basically factor inputs to the auto assembly process ("motor vehicles"). Moreover, since parts suppliers tend to sell to more than one car assembler,[31] the intermediate locations between these assemblers provide natural market economies for such suppliers.[32]

Figure 4.12 here

A second example is provided by the "plastic compounds and reclaimed plastics" industry (JSIC225) [with $GE = 0.298$ and $LD = 0.465$]. From Figure 4.13(b) it is clear that most clusters for this industry also follow the Industrial Belt.[33] Moreover, the outputs of this industry are again primarily intermediate inputs for a variety of manufactured goods, and in particular for motor vehicles (such as the molded plastic parts for seats, fenders, and instrument panels). Thus the intermediate locations between these manufacturers again constitute natural market-oriented locations for this industry. Hence the filling-in process that created this industrial belt is largely a consequence of the fact that typical automobiles consist of as many as 20,000 to 30,000 separate parts.

Figure 4.13 here

[31] In 1999, parts suppliers on average sold to 3.05 of the 9 auto assemblers in Japan, while auto assemblers on average bought the same parts from 2.46 different suppliers (Kinnou [21]).

[32] For a theoretical development of this "filling-in" process in the context of the new economic geography model see Mori [25].

[33] The lower density for this industry is due mainly to the fact that the e-containment in Figure 4.13(c) also includes the clusters on the Sea of Japan coast around Himi and Takaoka which have a large historical agglomeration of molding and casting industries (refer to Figure 4.13(b)).

5 Concluding Remarks

In this paper we have applied the cluster-detection procedure developed in [MS] to study the agglomeration patterns of manufaturing industries in Japan. In addition, we have proposed a simple classification of pattern types based on measures of *global extent* and *local density* derived from cluster schemes. But the ultimate utility of this approach will of course depend on how it can be applied in practical situations.

As alluded to in the Introduction, these measures can already help to sharpen certain concepts in the literature. For example, the differences between spatial dispersion of manufacturing at high versus low levels of transport costs, as derived in general NEG models, can be characterized in terms of these measures. In particular, the type of dispersion associated with high levels of transport costs ("first-phase" dispersion) can in principle be quantified empirically in terms of large global extent (GE) values and small local density (LD) values.[34] In contrast, dispersion patterns associated with low levels of transport costs ("second-phase" dispersion) might be quantified in terms of small GE values and large LD values. Hence, within a given *GE-LD space* of such values (as illustrated in Figure 4.4 for various levels of μ and ζ), such differences between dispersion patterns might quantified in terms of directed distances in this space. In fact, given appropriate historical data on industrial location patterns at various stages of transportation technology, one might even be able to test the significance of such differences.

As another illustration, it was pointed out in Section 4.3.4 above that the Japanese Industrial Belt is an instance of the more general notion of a "megalopolis," first proposed by Gottman [14] to describe the continuum of cities along the US Atlantic seaboard (stretching from Boston to Washington, DC, via New York). But to date, no formal methods have been developed for identifying such agglomeration structures statistically. In this light, the analysis of Section 4.3.4 shows that such structures can also be regarded

[34] Here it should be noted that since firms have no "area" in such continuous models, our present notion of *local density* is somewhat ambiguous. But given fixed employment levels for industries, the essence of this type of dispersion is that individual agglomerations become smaller and more scattered throughout the spatial continuum. So local "employment" density decreases under this type of dispersion.

as natural instances of "globally confined and locally dense" agglomeration patterns. Hence, the emergence of such large scale structures might in principle be characterized in term of urban agglomeration pattern shifts within an appropriate *GE-LD* space.

But it should also be emphasized that these two measures are by no means the only relevant properties of agglomeration patterns that can quantified. Indeed, our present construction of such patterns in terms of cluster schemes provides a potentially rich spatial data set for studying a wide range of problems. Along these lines, it is appropriate to mention two possible research directions involving, respectively, the spacing of agglomertions within industries and the coordination of agglomerations between industries.

5.1 Agglomeration Spacing within Industries

Within the new economic geography, a number of models have been developed to explain the spacing between individual agglomerations for a given industry (e.g., Krugman [22], Fujita and Krugman [8], Fujita and Mori [11], Fujita et al. [10, Ch.6], Ikeda et al. [17], Akamatsu et al. [1]). From the view point of general equilibrium theory, these models predict whether an agglomeration of industrial firms will be viable at a given location, depending on how other agglomerations of the same industry (as well as population) are distributed over the location space. In these models, industrial agglomeration is typically induced by demand externalities arising from the interactions between product differentiation, plant-level scale economies and transport costs. In particular, Fujita and Krugman [8] have shown that each agglomeration casts a so-called *agglomeration shadow* in which firms have no incentive to relocate from the existing agglomerations. For within this "shadow" firms are too close to existing agglomerations (i.e., competitors) to realize sufficient local monopoly advantages. Hence the presence of such shadows serves to limit the number of viable agglomerations within each industry. Note also that since the level of internal competition differs between industries (depending on their degree of product differentiation and transport costs), the size of

agglomeration shadows should also be industry specific. Hence the presence of such shadows has a number of observable spatial consequences.

But while there has been empirical work to study the spacing between urban centers (as for example in Chapter 7 of Marshall [24] and in Ioannides and Overman [18]), there have to our knowledge been no systematic efforts to study the spacing between industrial agglomerations – and in particular, no efforts to identify the presence of actual agglomeration shadows. However, it should be clear that our present approach to cluster identification offers a promising method for doing so. In particular, since our cluster-detection procedure enables one to identify individual agglomerations for each industry, it is a simple matter to construct explicit measures of the spacing between them. In the present setting, the most natural measure of spacing between any pair of clusters, C_j and C_h in a given cluster scheme, $\mathbf{C}$, is the *shortest-path distance*, $d(C_j,C_h)$, between their closest basic regions on the given road network. Hence the *size* of the agglomeration shadow cast by any cluster, $C_j \in \mathbf{C}$, can be modeled by the distance to it closest neighbor in $\mathbf{C}$. In these terms, a simple summary measure of agglomeration spacing for $\mathbf{C}$ is given by the mean nearest-neighbor distance between its constituent clusters.

To test whether this spacing is larger (or more uniform) than would be expected by chance alone, one could in principle generate random versions of $\mathbf{C}$ involving clusters of roughly the same size with the actual ones but with possibly very different spacing. While such random collections of disjoint sets are of course more difficult to construct than random point patterns, initial investigations with a variety of rejection-sampling techniques suggest that this is certainly possible. Hence by constructing mean nearest-neighbor distances for each random version sampled, one can use this sampling distribution to test a variety of agglomeration spacing properties in terms of cluster schemes, $\mathbf{C}$. Such spacing analyses will be reported in subsequent work.[35]

[35]Here it is of interest to note that initial investigations of such spacing properties suggest that further restrictions need to be imposed. In particular, for those industries with small e-containments, it is clear that random versions located throughout all of Japan will necessarily tend to exhibit larger mean spacing for rather spurious reasons. One possibility here is to preserve the e-containment of each industry, and

5.2 Agglomeration Coordination between Industries

Within the context of Christaller's [3] celebrated theory of *Central Places*, a topic of major interest has long been the spatial coordination of locations across industries. In particular, the "Hierarchy Principle" underlying this theory asserts that the set of industries found in smaller metro areas is always a subset of those found in larger metro areas.[36] Theoretical efforts to explain this phenomenon have focused mainly on the role of demand externalities in determining industrial locations (see Quinzii and Thisse [29], Fujita et al. [9], Tabuchi and Thisse [33] and Hsu [16]).[37] In particular, the types of demand externalities which induce industrial agglomerations are often shared by many different industries, so that their spatial markets overlap. In such cases, it is natural for these industries to co-locate. Moreover, in terms of market sizes, it is also natural for agglomerations in more concentrated industries (with larger markets) to coincide with those of less concentrated industries (with smaller markets), thus leading to the type of synchronization predicted by the Hierarchy Principle.

But while these theoretical arguments are quite plausible, there has been surprisingly little work done to actually test the empirical validity of the Hierarchy Principle itself. One approach proposed by Mori and Smith [27] focuses on the hierarchical industrial structure of cities implied by this principle. In particular, the present cluster-detection procedure was used to identify those cities containing establishments that are actually part of clusters for the industry. By restricting the classical Hierarchy Principle to these "cluster-based choice cities" for each industry, it was shown that this *cluster-based Hierarchy Principle* holds even more strongly than the classical version for our Japanese data.

However, the detailed spatial structure of cluster schemes also permits more direct comparisons of spatial coordination between individual industries. In particular, if we

to restrict random versions to these e-containments. This should provide more meaingful tests of the presence of agglomeration shadows in which the overall spatial scale of each industry is preserved.

[36] Obviously, this principle implicitly assumes a certain degree of industry aggregation, since it could not hold if industries are fully disaggregated, i.e., where each industry consists of one establishment.

[37] There were earlier attempts by, e.g., Christaller [3], Lösch[23], Beckmann [2] and Eaton and Lipsey [6]. But, all lacked formal microeconomic foundations leading to the Hierarchy Principle.

associate larger market sizes with smaller numbers of clusters (agglomerations) for an industry,[38] then one may ask whether industries with larger market sizes do in fact tend to coordinate their spatial locations with industries having smaller market sizes. More formally, for any pair of industries, $i = 1, 2$, with cluster schemes, $\mathbf{C}_i = (R_{0i}, C_{1i}, .., C_{k_{\mathbf{C}_i}})$, satisfying $k_{\mathbf{C}_1} < k_{\mathbf{C}_2}$, we may ask whether the clusters in $\mathbf{C}_1$ are "closer" to those in $\mathbf{C}_2$ than would be expected by chance alone. As one possible measure of "closeness", we can proceed as in Section 5.1 above by identifying the (*shortest path*) *nearest-neighbor distance* from each cluster, $C_{1h} \in \mathbf{C}_1$, to those in $\mathbf{C}_2$:

$$d(C_{1h}, \mathbf{C}_2) = \min\{d(C_{1h}, C_{2j}) : C_{2j} \in \mathbf{C}_2\} \tag{16}$$

and then to define the *mean distance* between $\mathbf{C}_1$ and $\mathbf{C}_2$ to be the average of these:

$$d(\mathbf{C}_1, \mathbf{C}_2) = \frac{1}{k_{\mathbf{C}_1}} \sum_{h=1}^{k_{\mathbf{C}_1}} d(C_{1h}, \mathbf{C}_2) \tag{17}$$

To employ this summary measure as a test statistic, one could again use the procedure in Section 5.1 to generate many random versions, $\mathbf{C}_1'$ of $\mathbf{C}_1$, and test whether $d(\mathbf{C}_1, \mathbf{C}_2)$ is significantly smaller than would be expected from the sampling distribution of mean-distance values, $d(\mathbf{C}_1', \mathbf{C}_2)$. Applications of this testing procedure will also be reported in subsequent work.

[38] In fact this relationship underlies the results in the theoretical papers above.

References

[1] Akamatsu, Takahshi, Takayama, Yuki, Ikeda, Kiyohiro (2010) "Spatial discounting, fourier, and racetrack economy: A recipe for the analysis of spatial agglomeration models," MPRA Paper 21738, University Library of Munich, Germany.

[2] Beckmann, Martin J. (1958), "City Hierarchies and the Distribution of City Size," *Economic Development and Cultural Change*, **6**, 243-248.

[3] Christaller, William (1933), *Die Zentralen Orte in Suddeutschland* (Jena, Germany: Gustav Fischer). English translation by Baskin, Carlisle W. (1966), *Central Places in Southern Germany* (London: Prentice Hall).

[4] Combes, Pierre-Phillippe, Mayer, Thierry, Thisse, Jaques-Francǫis (2008), *Economic Geography: The Integration of Regions and Nations* (Princeton, NJ: Princeton University Press).

[5] Duranton, Gilles, Overman, Henry G. (2005), "Testing for Localization Using Micro-Geographic Data,"*Review of Economic Studies*, **72**, 1077-1106.

[6] Eaton, B. Curtis, Lipsey, Richard G. (1982), "An Economic Theory of Central Places," *Economic Journal*, **92(365)**, 56-72.

[7] Ellison, Glenn, Glaeser, Edward L. (1997), "Geographic Concentration in US Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, **105(5)**, 889-927.

[8] Fujita, Masahisa, Krugman, Paul R. (1995), "When Is the Economy Monocentric?: von Thüen and Chamberlin Unified," *Regional Science and Urban Economics* **25**, 505-528.

[9] Fujita, Masahisa, Krugman, Paul R., Mori Tomoya (1999) "On the Evolution of Hierarchical Urban Systems," *European Economic Review*, **43**, 209-251.

[10] Fujita, Masahisa, Krugman, Paul R., Venables, Anthony J. (1999), *The Spatial Economy: Cities, Regions and International Trade* (Cambridge, MA: The MIT Press).

[11] Fujita, Masahisa, Mori, Tomoya (1997), "Structural Stability and Evolution of Urban systems," *Regional Science and Urban Economics*, **27**, 399-442.

[12] ———— (2005), "Frontiers of the New Economic Geography," *Papers in Regional Science*, **84(3)**, 307-405.

[13] ———— (2005), "Transport Development and the Evolution of Economic Geography," *Portuguese Economic Journal*, **4(2)**, 129-156.

[14] Gottman, Jean (1961), *Megalopolis: The Urbanized Northern Seaboad of the United States* (New York: The Twentieth Century Fund).

[15] Hokkaido-chizu, Co. Ltd. (2002), *GIS Map for Road*.

[16] Hsu, Wen-Tai (2009), "Central Place Theory and the City Size Distribution" (Manuscript, Department of Economics, Chinese University of Hong Kong).

[17] Ikeda, Kiyohiro, Akamatsu, Takashi, Kono, Tatsuhito (2012) "Spatial Period Doubling Agglomeration of a Core-Periphery Model with a System of Cities," forthcoming in *Journal of Economic Dynamics and Control*.

[18] Ioannides, Yannis M., Overman, Henry G. (2004), "Spatial Evolution of the US Urban System," *Journal of Economic Geography*, **4(2)**, 131-156.

[19] Japan Statistics Bureau (2000), *Population Census of Japan* (in Japanese).

[20] Japan Statistics Bureau (2001), *Establishments and Enterprise Census of Japan* (in Japanese).

[21] Konnou, Yoshinori (2003), "Jidosha Buhin Torihiki No Open-ka No Kensho," *Keizaigaku Ronshu* **68(4)**, 58-86 (in Japanese).

[22] Krugman, Paul R. (1993), "On the Number and Location of Cities," *European Economic Review*, **37**, 293-298.

[23] Lösch, August (1940), *The Economics of Location* (Jena, Germany: Fischer). English translation by Stolper, Wolfgang F. and Woglom, William H. (1954) (New Haven, CT: Yale University Press).

[24] Marshall, John U. (1989), *The Structure of Urban Systems* (Toronto: University of Toronto Press).

[25] Mori, Tomoya (1997), "A Modeling of Megalopolis Formatin: the Maturing of City Systems," *Journal of Urban Economics* **42**, 133-157.

[26] Mori, Tomoya, Nishikimi, Koji, Smith, Tony E. (2005), "A Divergence Statistic for Industrial Localization," *Review of Economics and Statistics*, **87(4)**, 635-651.

[27] Mori, Tomoya, Smith, Tony E. (2011), "An Industrial Agglomeration Approach to Central Place and City Size Regularities," *Journal of Regional Science* 51(4), 694-731.

[28] ———— (2011) "A probabilistic modeling approach to the detection of industrial agglomerations: Metholodogical framework", Discussion paper No.777, Institute of Economic Research, Kyoto University.

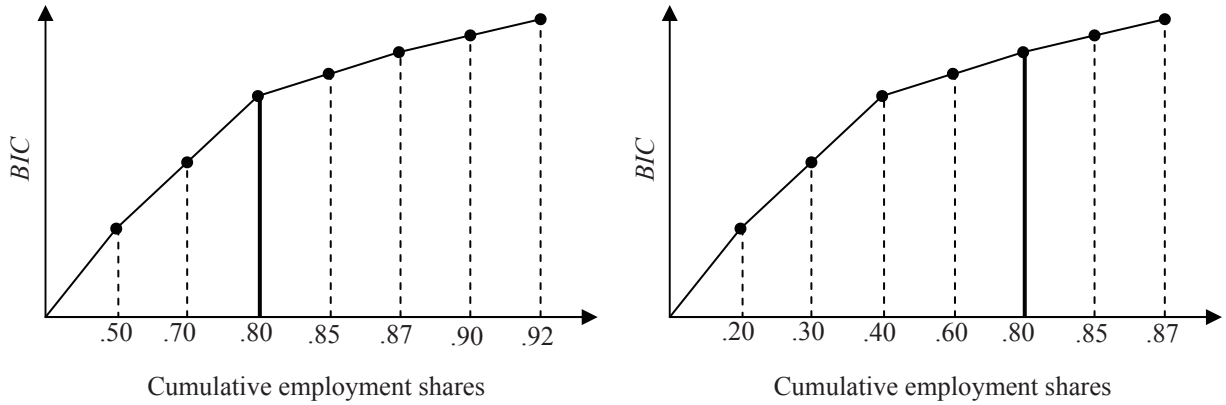
[29] Quinzii, Martine, Thisse, Jacques-François (1990) "On the optimality of central places", *Econometrica* 58, 1101-1119.

[30] Rosenthal, Stuart S., Strange, William C. (2004), "Evidence on the Nature and Sources of Agglomeration Economies," in Henderson, J. Vernon and Thisse, Jacques-Francçois (eds.), *Handbook of Regional and Urban Economics*, Vol.4 (Amsterdam: North-Holland), Ch.49.

[31] Statistical Information Institute for Consulting and Analysis (2002), *Toukei de Miru Shi-Ku-Cho-Son no Sugata* (in Japanese).

[32] Statistical Information Institute for Consulting and Analysis (2003), *Toukei de Miru Shi-Ku-Cho-Son no Sugata* (in Japanese).

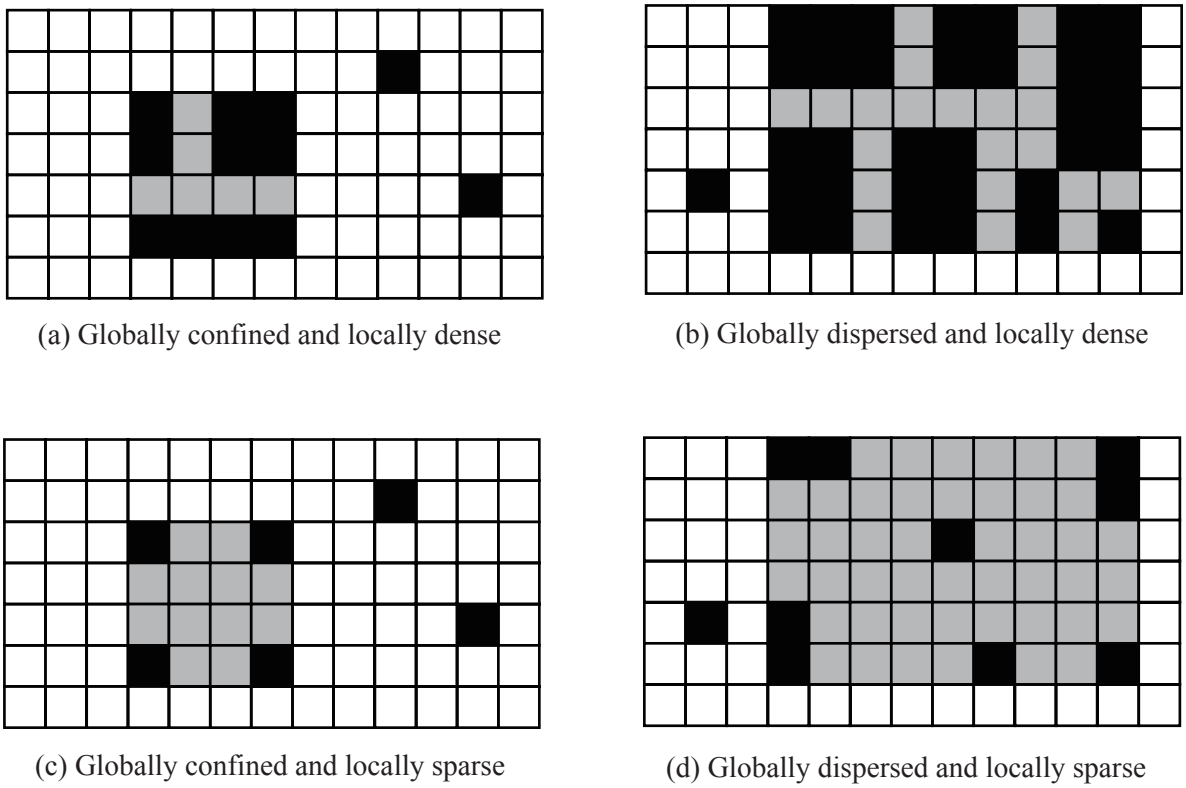
[33] Tabuchi, Takatoshi, Thisse, Jacques-François (2006), "Regional Specialization, Urban Hierarchy, and Commuting Costs," *International Economic Review*, **47**, 1295-1317.



(a) BIC-increment threshold

(b) Employment-share threshold

Figure 3.1. Thresholds for essential containment



(a) Globally confined and locally dense

(b) Globally dispersed and locally dense

(c) Globally confined and locally sparse

(d) Globally dispersed and locally sparse

Figure 3.2. Classifications of agglomeration patterns

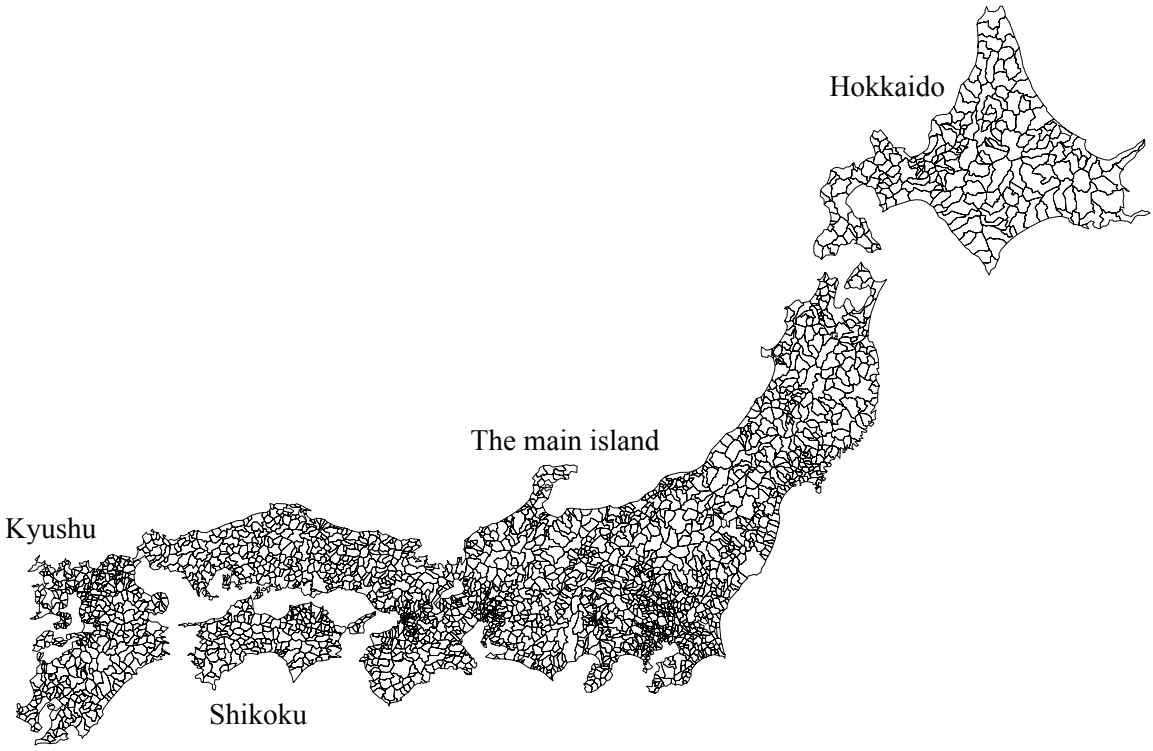


Figure 4.1. Basic regions (shi-ku-cho-son) of Japan

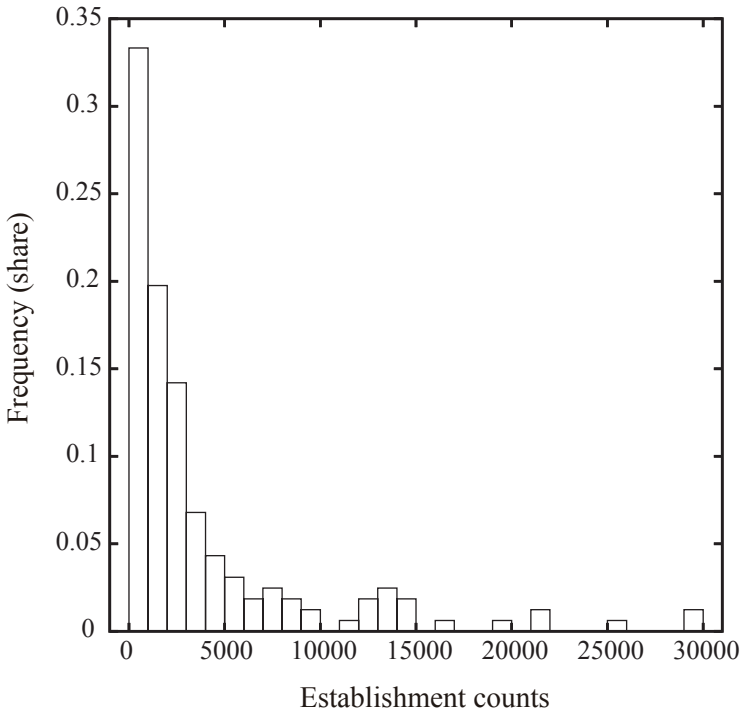


Figure 4.2. Frequency distribution of establishment counts in Japanese three-digit manufacturing industries

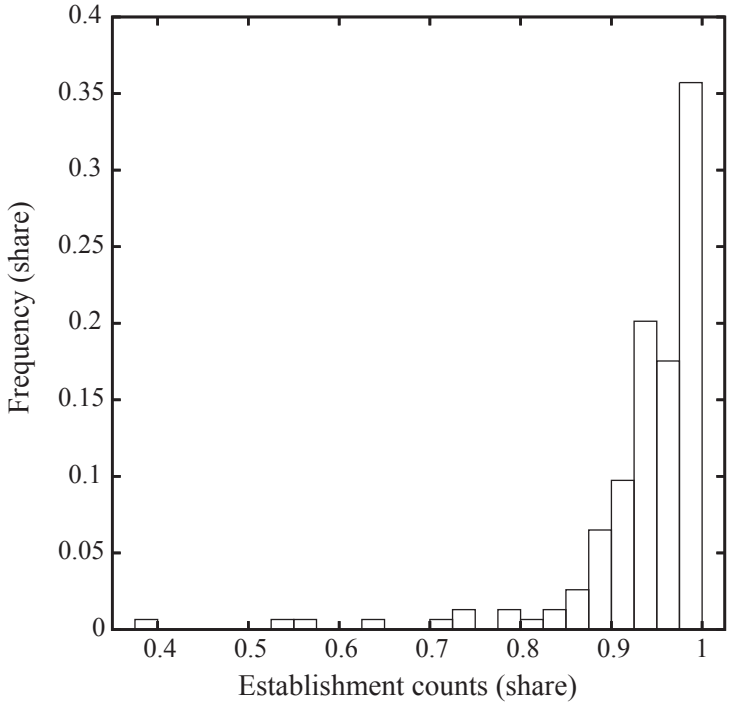


Figure 4.3. Share of establishment counts included in clusters

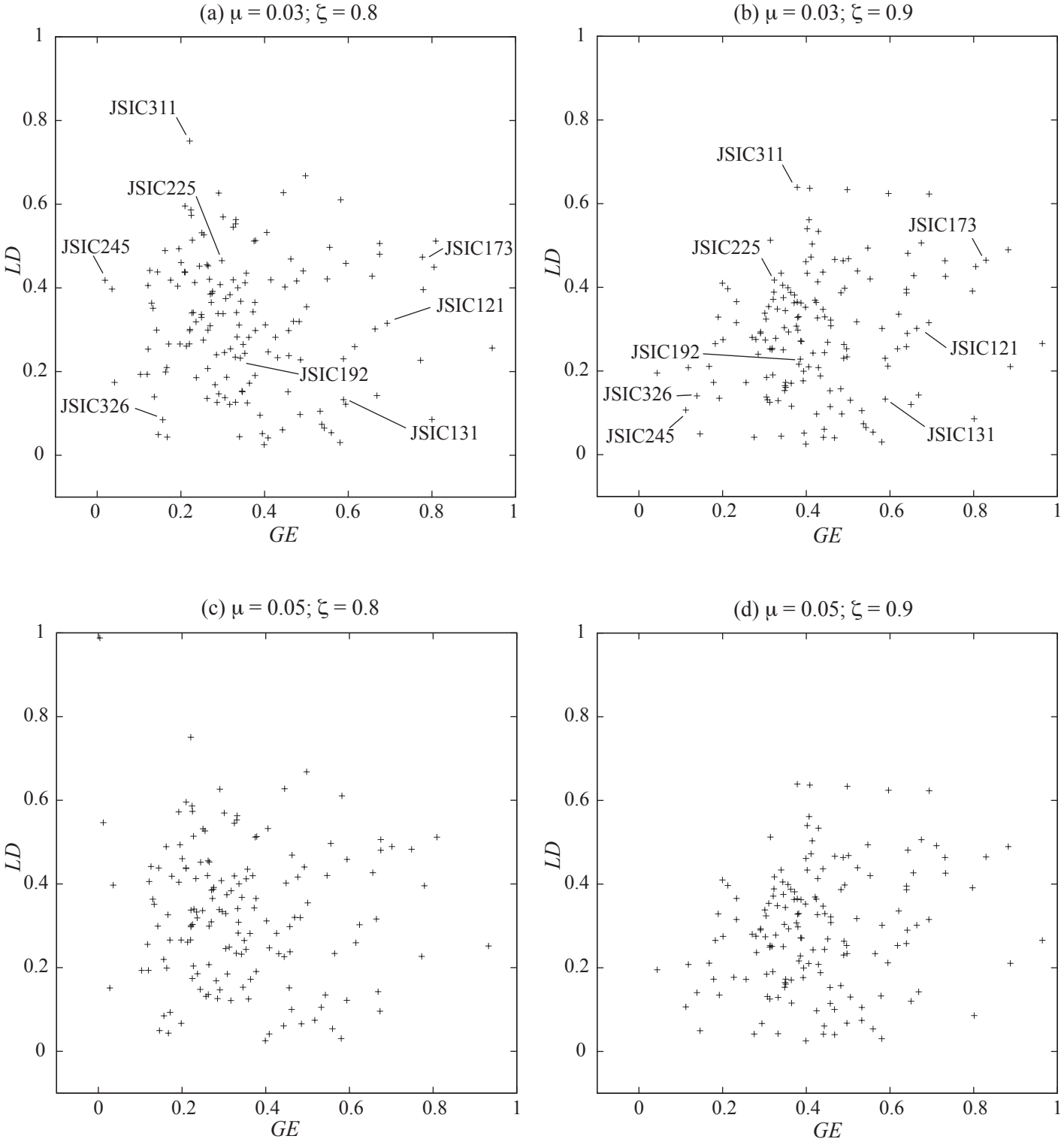
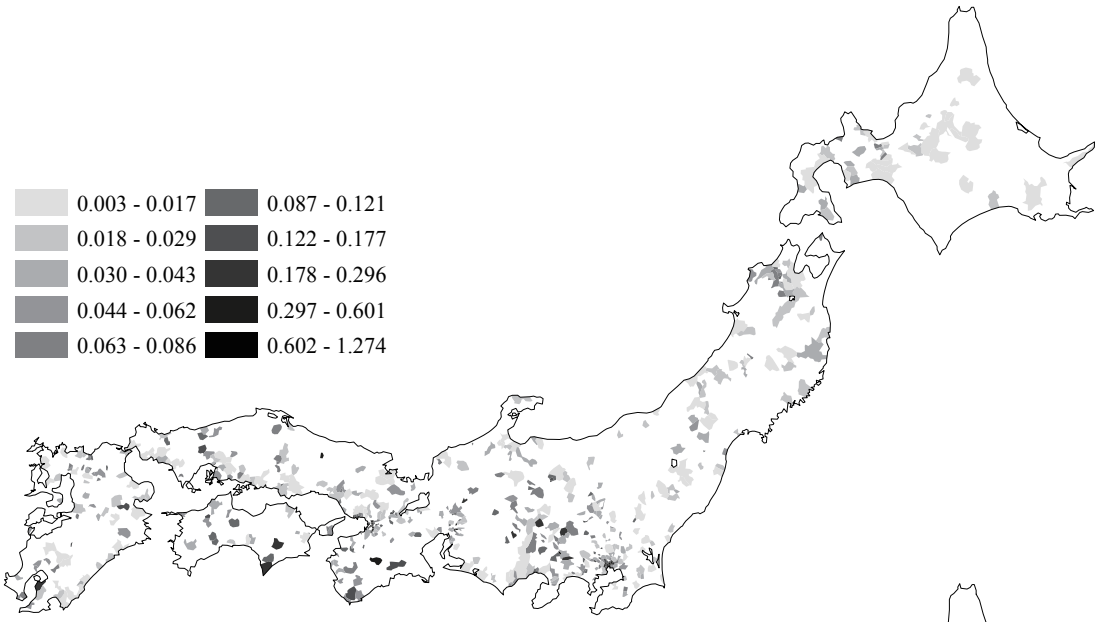
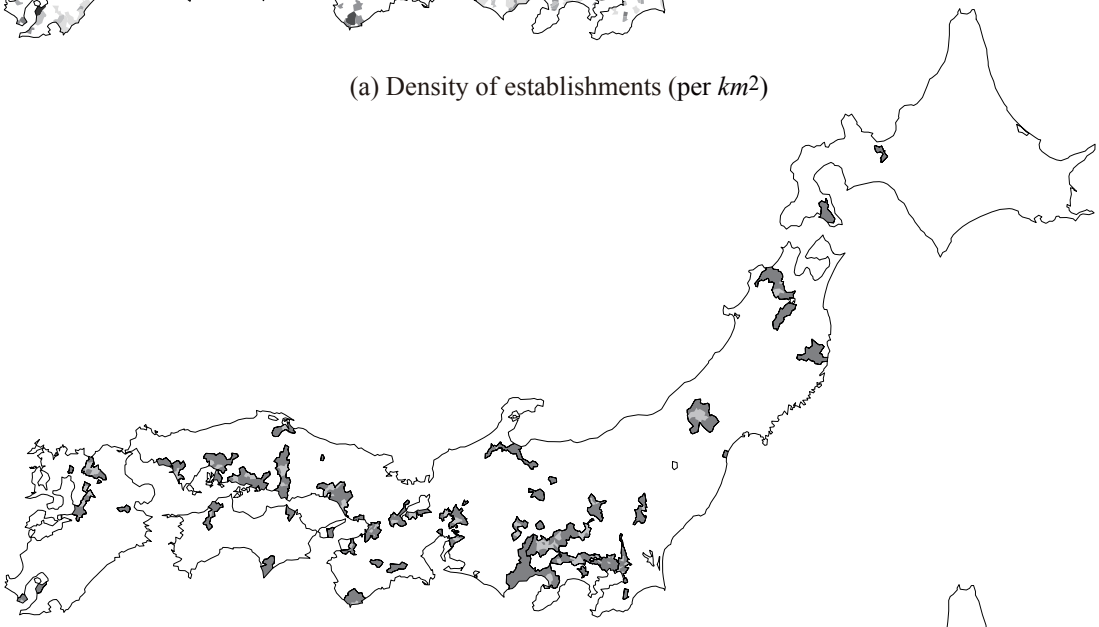


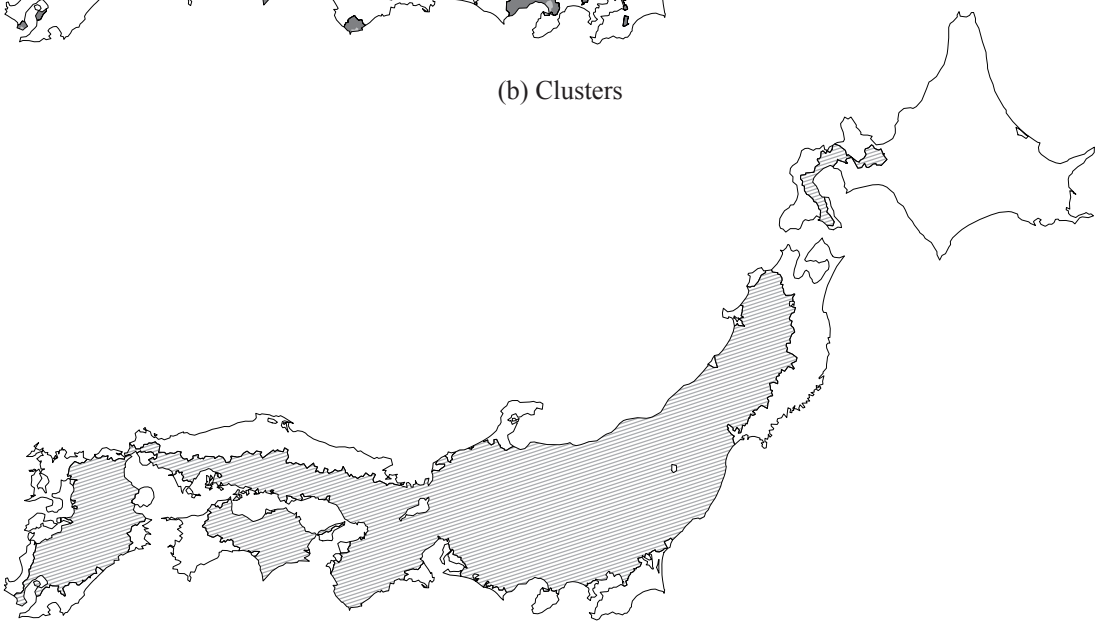
Figure 4.4. Global extent and local dispersion of clusters



(a) Density of establishments (per km^2)

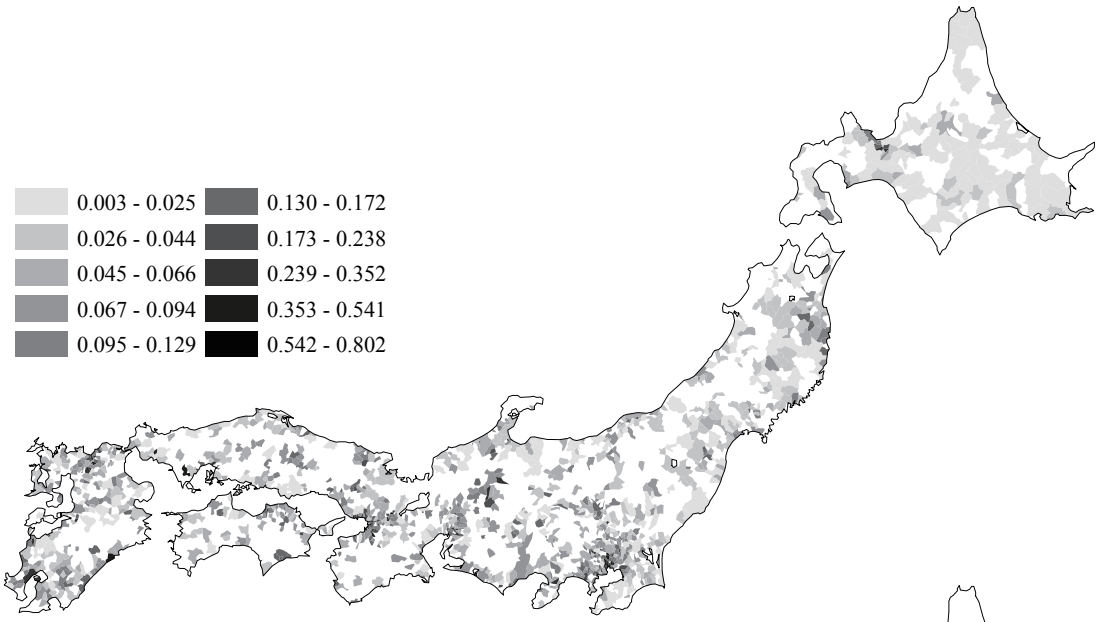


(b) Clusters

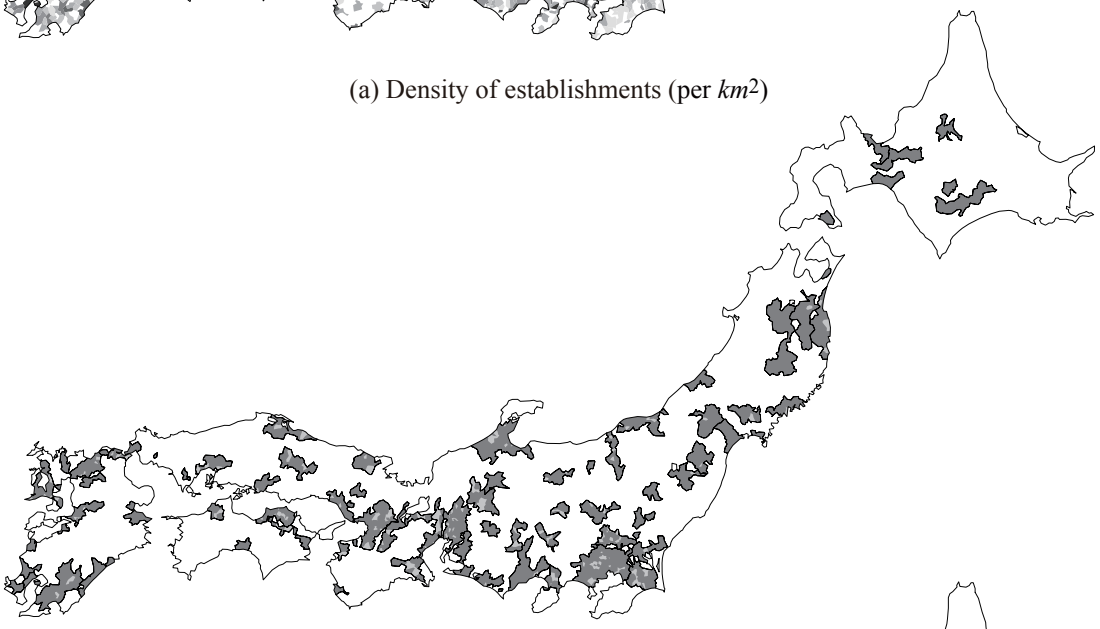


(c) Essential containment

Figure 4.5. Global dispersed and locally sparse pattern: soft drinks and carbonated water (JSIC131)



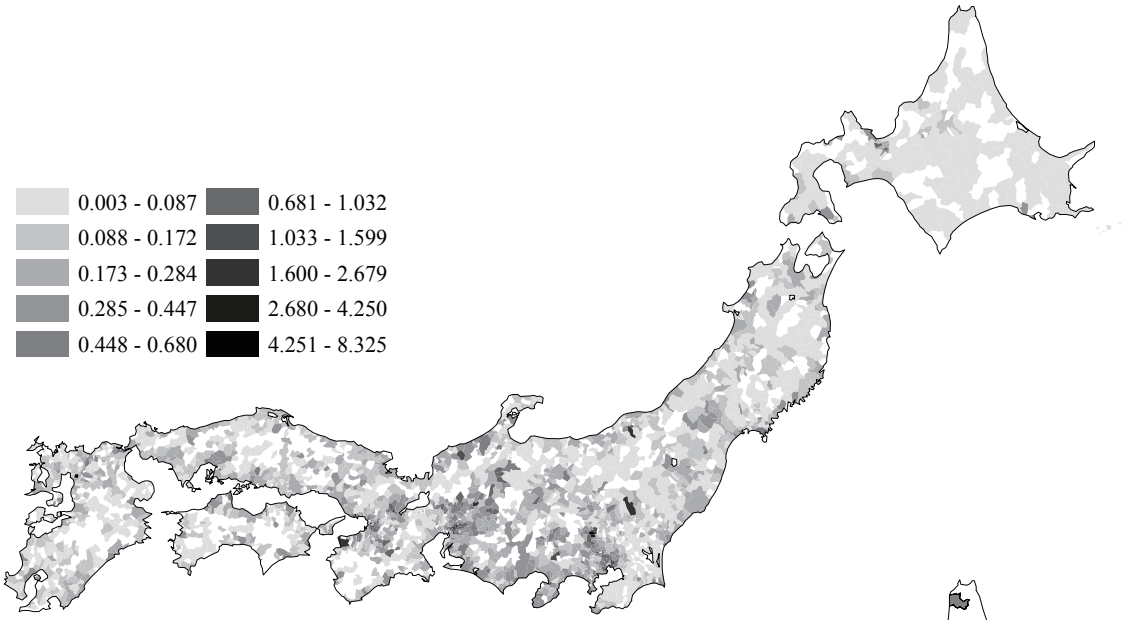
(a) Density of establishments (per km^2)



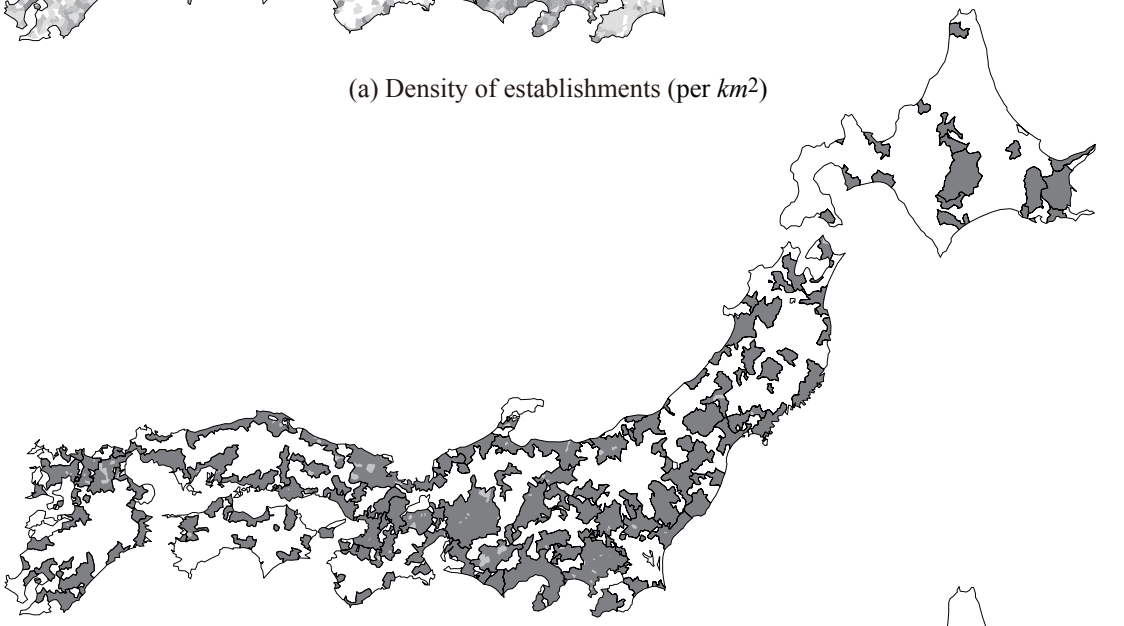
(b) Clusters

(c) Essential containment

Figure 4.6. Global dispersed and local sparse pattern: livestock products (JSIC121)



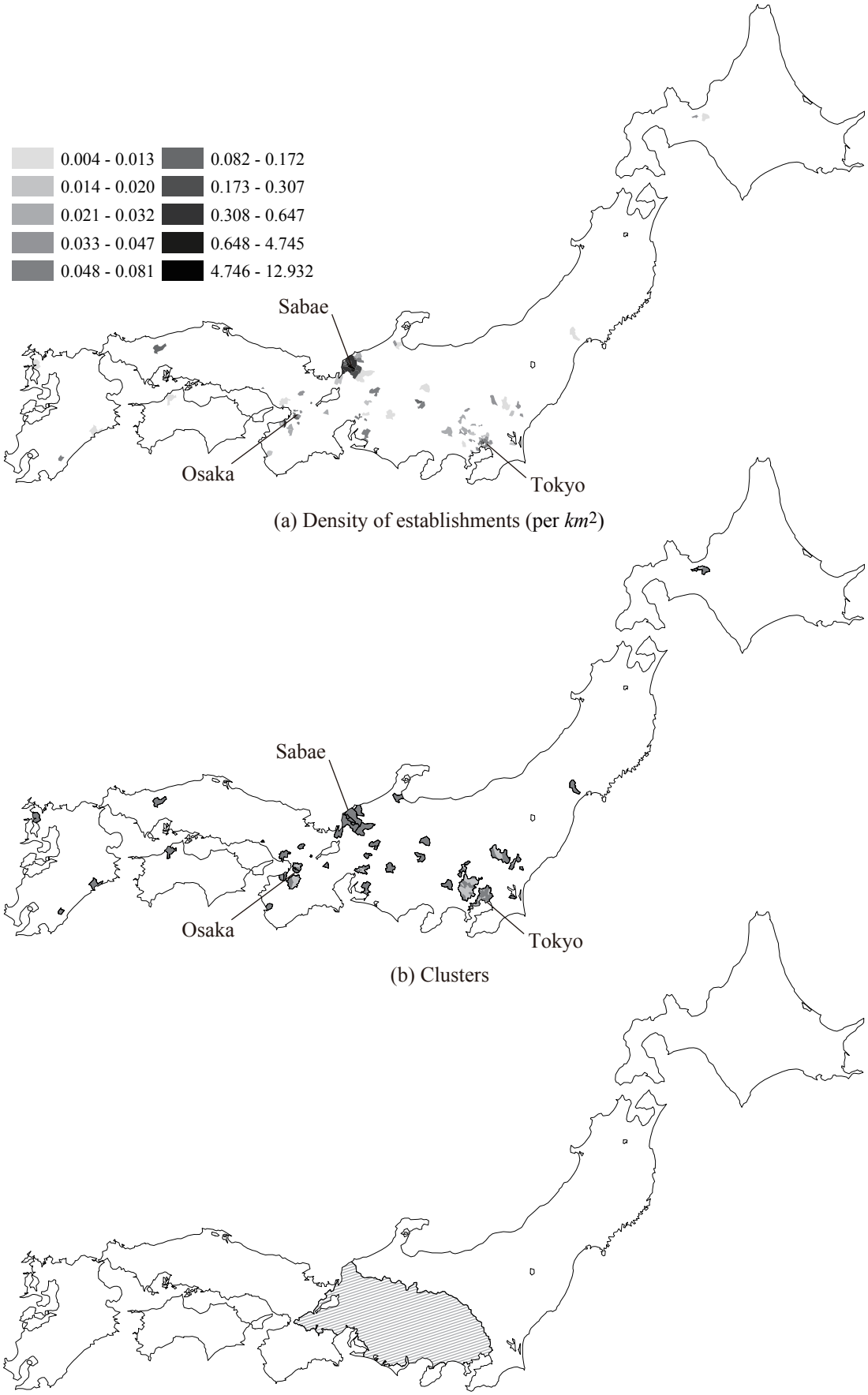
(a) Density of establishments (per km^2)



(b) Clusters

(c) Essential containment

Figure 4.7. Globally dispersed and locally sparse pattern: sliding doors and screens (JSIC173)

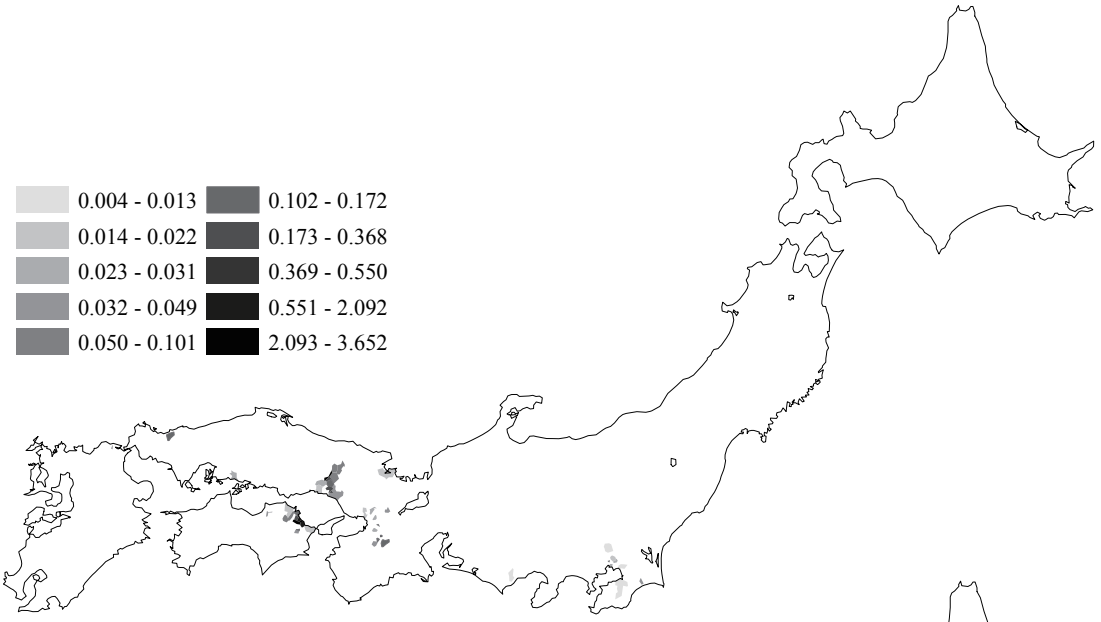


(a) Density of establishments (per km^2)

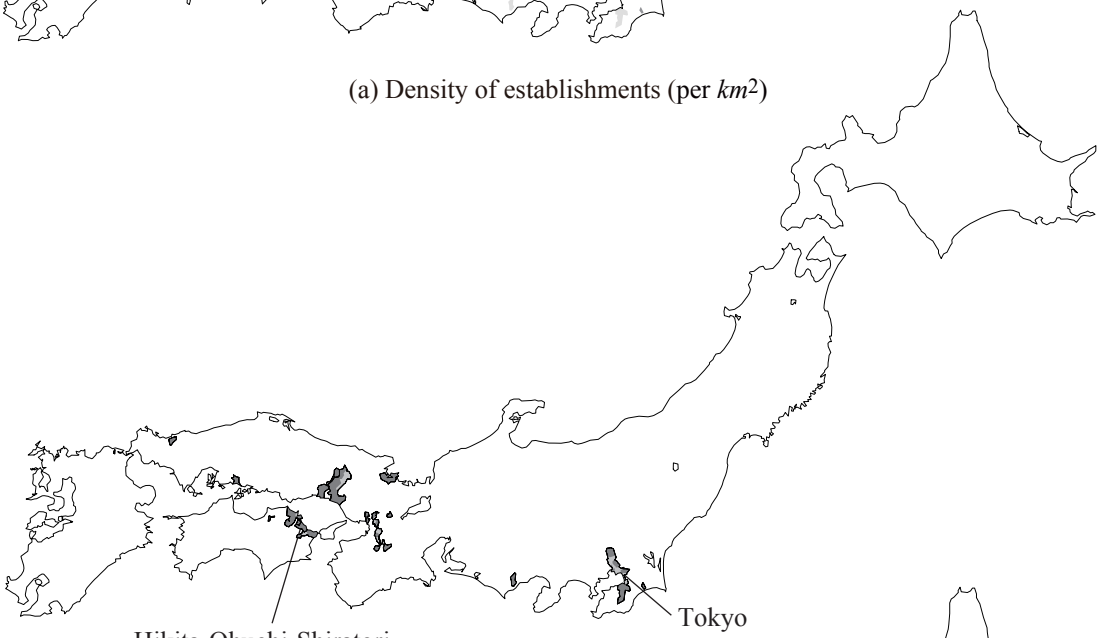
(b) Clusters

(c) Essential containment

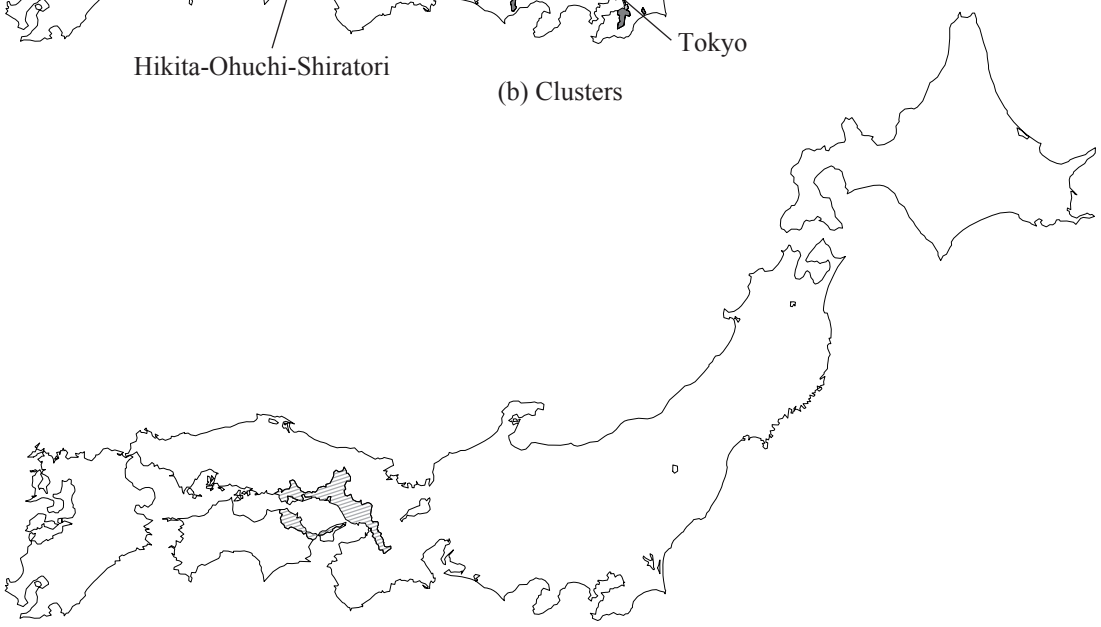
Figure 4.8. Globally confined and locally sparse pattern: ophthalmic goods, including frames (JSIC326)



(a) Density of establishments (per km^2)



(b) Clusters



(c) Essential containment

Figure 4.9. Globally confined and locally sparse pattern: leather gloves and mittens (JSIC245)

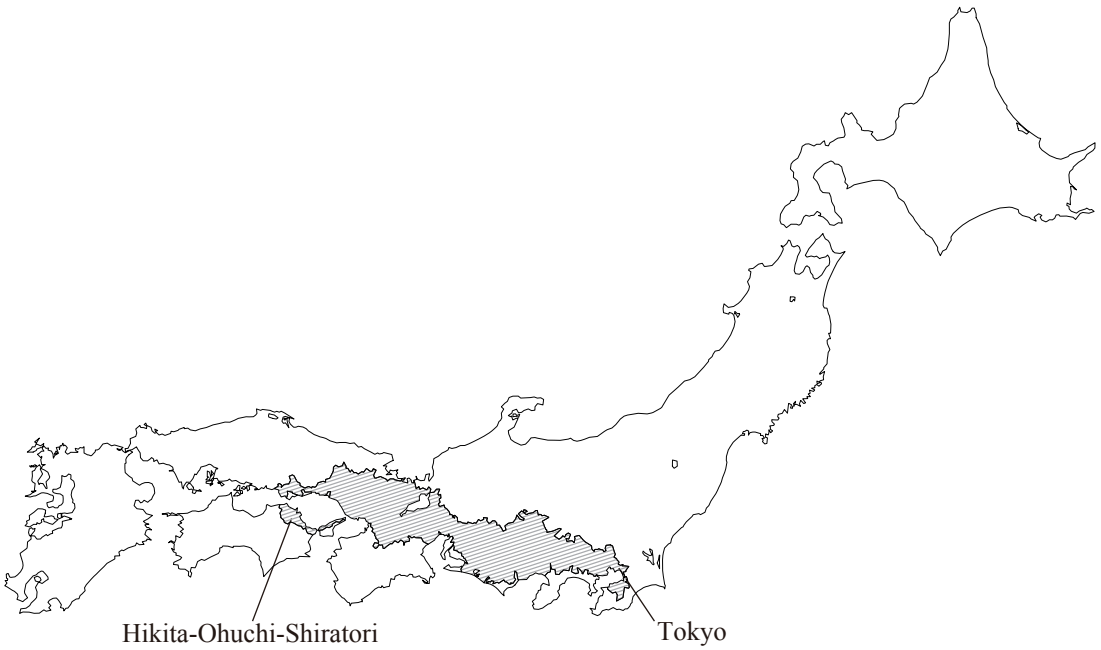
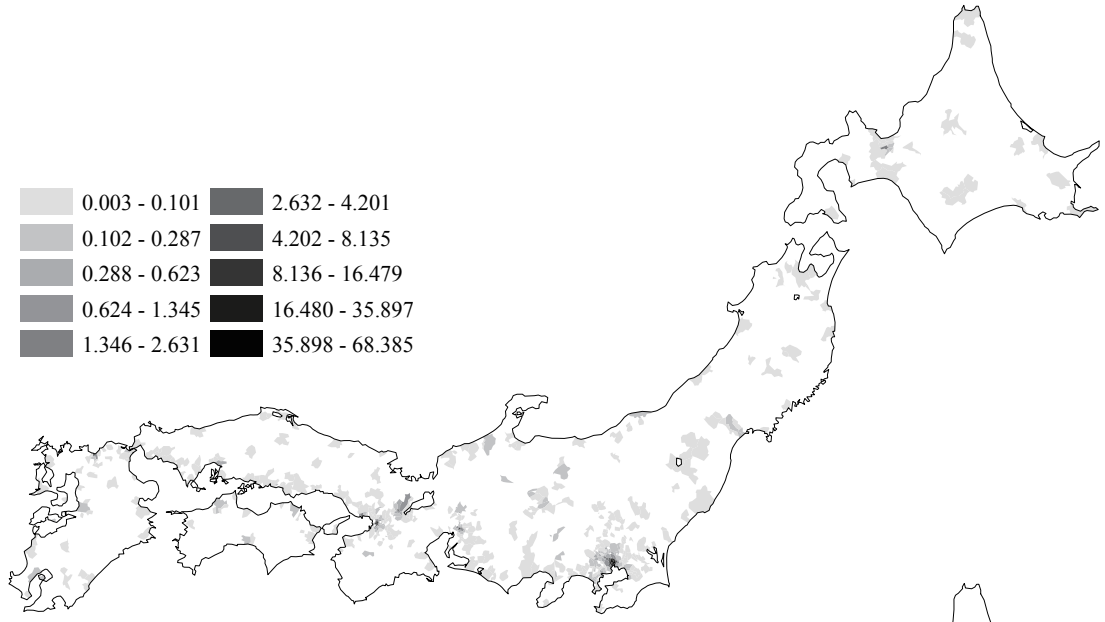
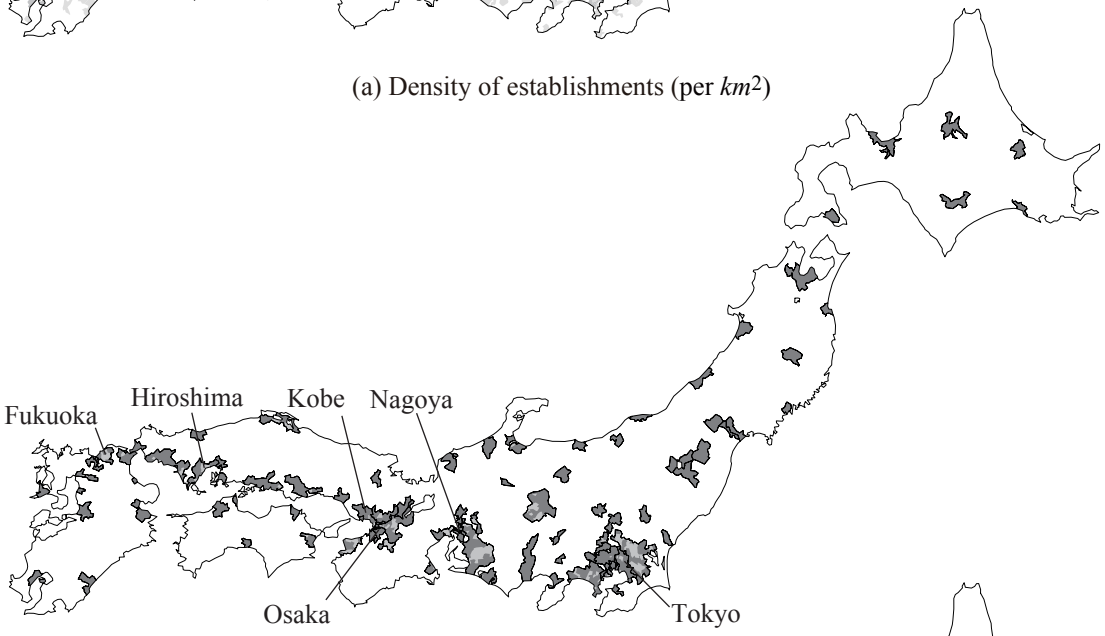


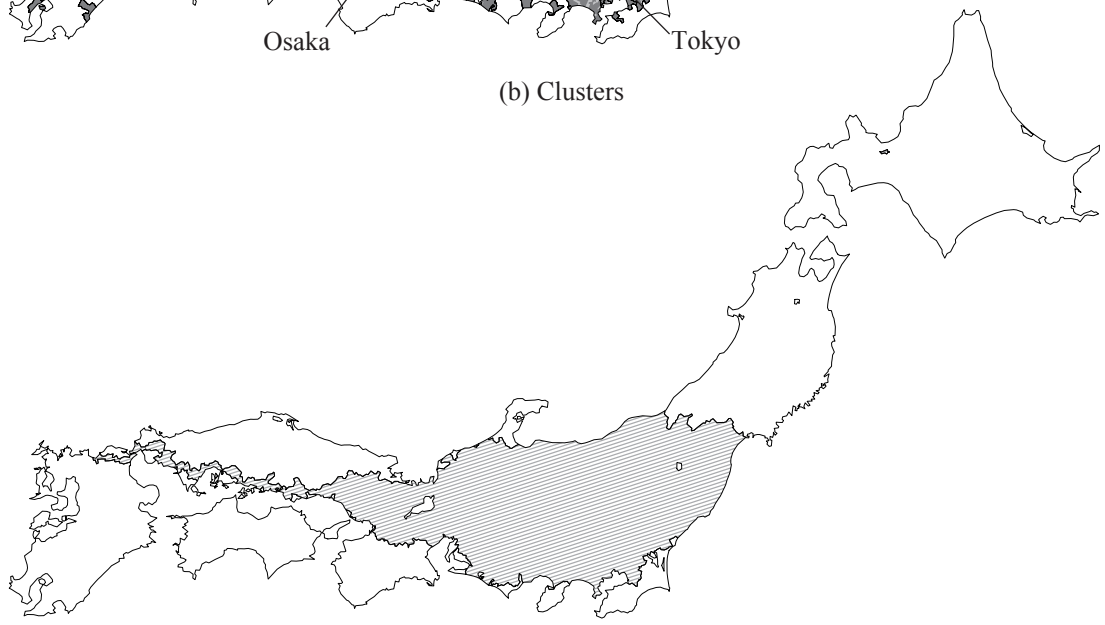
Figure 4.10. Essential containment of leather gloves and mittens (JSIC245) with $\delta = 0.03$ and $\zeta = 0.9$



(a) Density of establishments (per km^2)

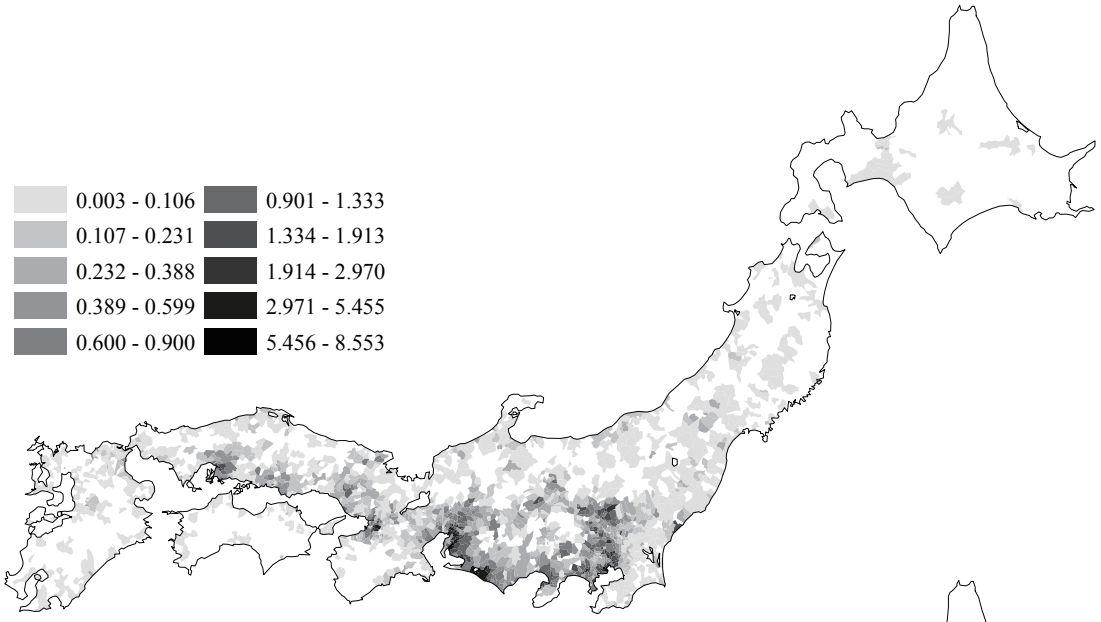


(b) Clusters

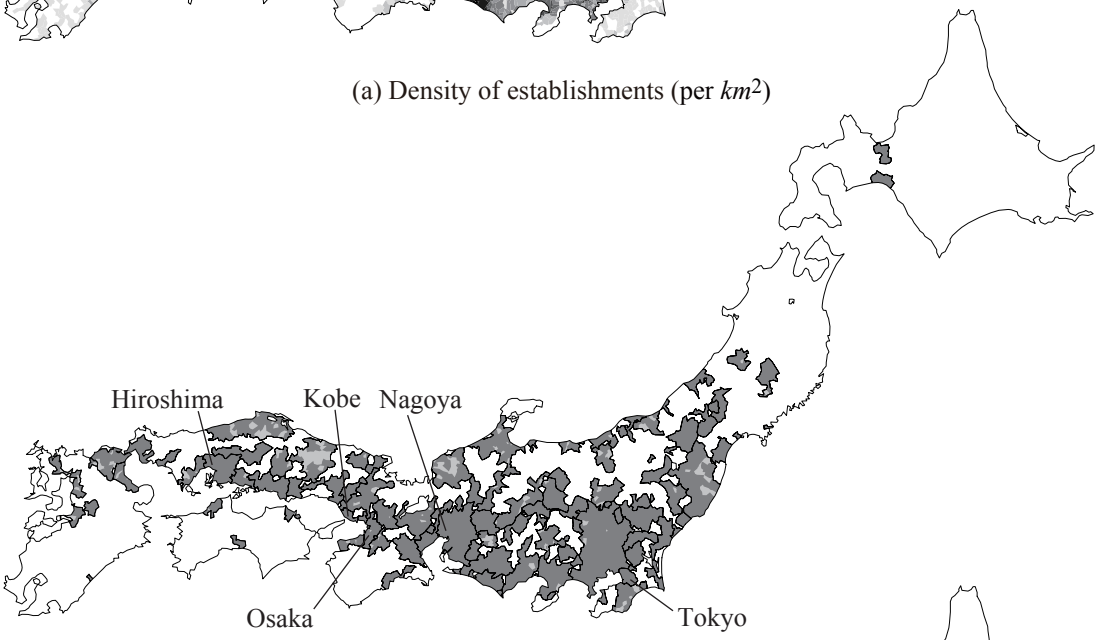


(c) Essential containment

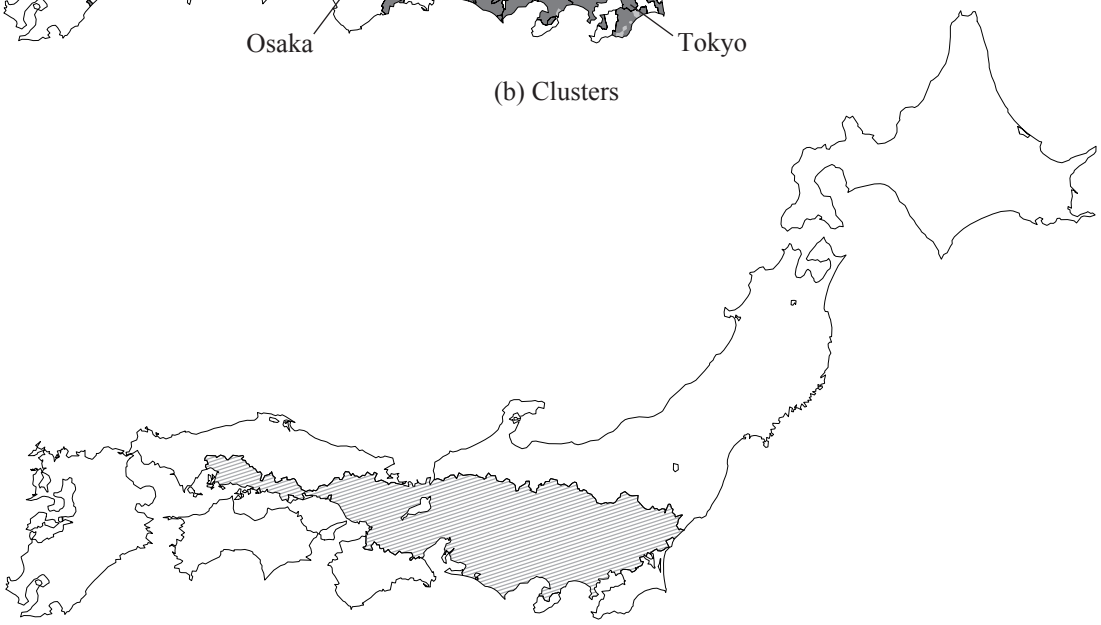
Figure 4.11. Globally confined and locally sparse pattern: publishing industries (JSIC192)



(a) Density of establishments (per km^2)

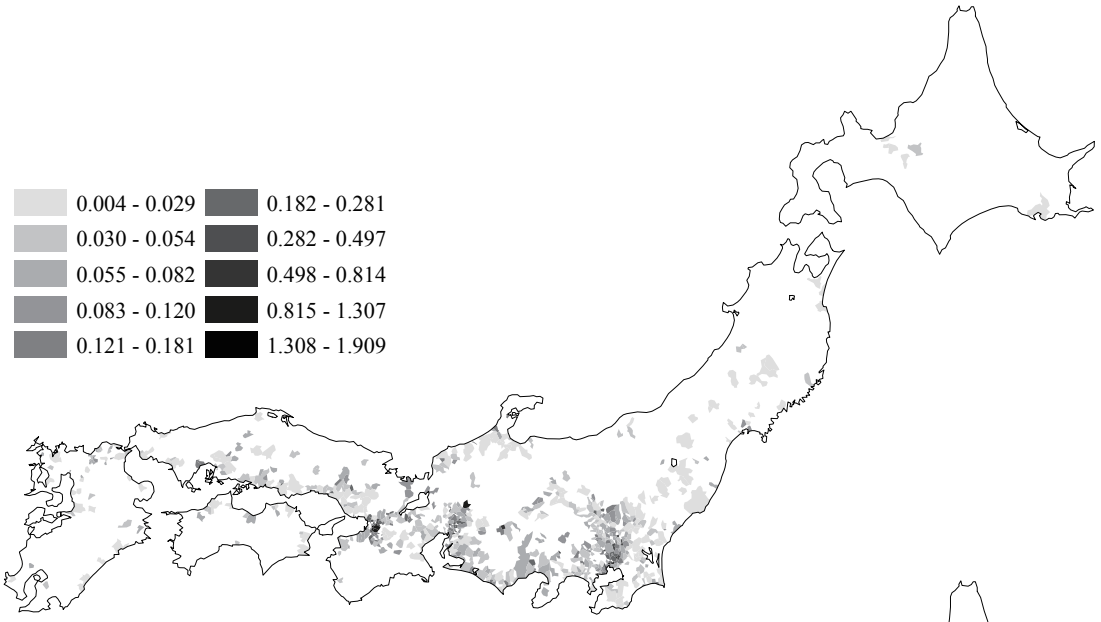


(b) Clusters

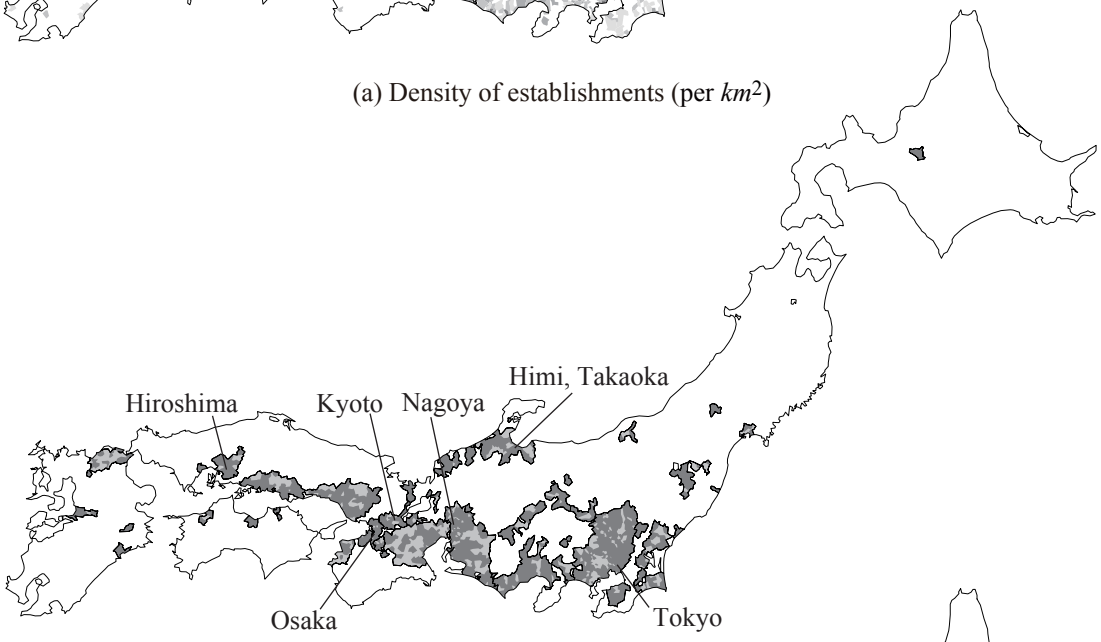


(c) Essential containment

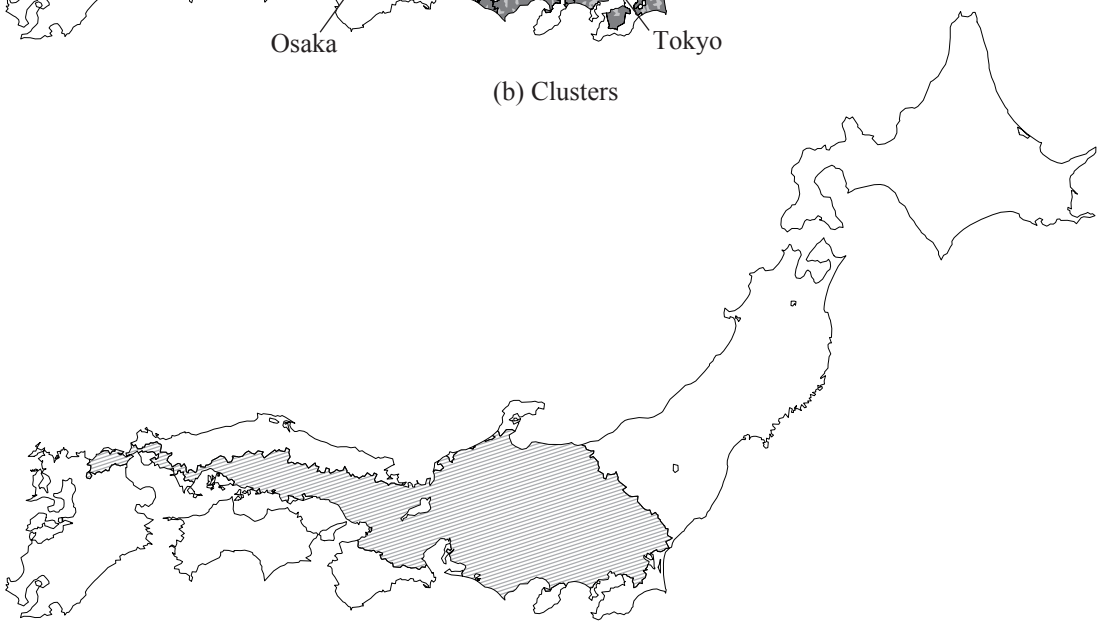
Figure 4.12. Globally confined and locally dense pattern: motor vehicle, parts and accessories (JSIC311)



(a) Density of establishments (per km^2)



(b) Clusters



(c) Essential containment

Figure 4.13. Globally confined and local dispersed pattern: compounding plastic materials, including reclaimed plastics (JSIC225)