*CSIS Discussion Paper No*. **119R**

# Decomposition approach to the measurement of spatial segregation

Yukio Sadahiro* and Seong-Yun Hong**

July 2014

Center for Spatial Information Science, University of Tokyo
5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan
* sada@csis.u-tokyo.ac.jp
** yun.hong@csis.u-tokyo.ac.jp

**Keywords:** spatial segregation, decomposition approach, measurement

**Abstract**

This paper develops a new method of evaluating segregation among point distributions. We argue that there are three main sources of segregation: difference in the spatial arrangements of points; imbalance in the number of points; and lack of diversity of points, and that segregation has three distinctive components, each of which is due to one of these sources. The proposed measures evaluate each component separately, and this approach has five desirable properties, which are not fully possessed by existing measures. To demonstrate the validity of the method, we apply it to an analysis of three data sets. The results show that the proposed measures can evaluate the degree of segregation effectively from three different perspectives, and their interpretation can be straightforward because each of the measures is linked with only one component of segregation.

## 1. Introduction

Segregation is a complex and multidimensional socio-spatial phenomenon. A complete, or near-complete, understanding of this phenomenon requires us to evaluate segregation from various perspectives, and for this reason, numerous indices have been proposed in geography, sociology, and other related fields. The index of dissimilarity (Duncan and Duncan, 1955), the Gini index, and the entropy-based indices (Theil and Finizza, 1971), for example, have been frequently used in the literature, in part because of their ease of calculation and interpretation. These earlier indices, however, are often criticized as *aspatial* since they are insensitive to the arrangement of spatial units (Morrill, 1991). To overcome this limitation, a number of alternative, *spatial* segregation indices have been developed in more recent years (e.g., White, 1983; Morrill, 1991; Wong, 1993; Reardon and O'Sullivan, 2004).

While all these indices aim to evaluate segregation from different perspectives, they are overlapping to some extent with each other. We thus need to choose an appropriate set of indices that explain the whole picture of segregation efficiently and address different dimensions of segregation separately. As Massey and Denton (1988) pointed out, however, there has been no consensus on which indices are necessary and sufficient to fully describe patterns of segregation. Though several papers have proposed criteria for segregation indices (James and Taeuber, 1985; Reardon and Firebaugh, 2002; Reardon and O'Sullivan, 2004), they do not directly discuss the relationship between the indices, and hence the criteria are not enough for choosing an appropriate set of indices.

To resolve the problem, Massey and Denton (1988) introduced five conceptually distinct dimensions (evenness, exposure, clustering, centralization, and concentration). They performed factor analysis on twenty segregation indices for the 1980 U.S. census data, from which they nominated five indices as the best indicators of the five dimensions. Reardon and O'Sullivan (2004), on the other hand, argued that the distinction between evenness and clustering is somewhat arbitrary, and proposed two primary dimensions and two indices. Johnston *et al*. (2007) conducted principal component analysis of the U.S. census data for three different years and concluded that two basic dimensions are sufficient to grasp the residential segregation in the U.S. metropolitan areas. They suggested the use of principal components' (PC) scores derived from principal component analysis as measures of segregation.

The concept of such *distinct dimensions* provides us a logical basis for choosing a set of indices. Since the dimensions of segregation are conceptually distinctive from each other, the chosen indices should also evaluate individual dimensions separately. Each index should fully represent one, and only one dimension, and overlaps between the indices are undesirable as it not only reduces the efficiency of analysis but also makes dimensions incomparable.

However, the existing indices do not completely satisfy these properties. The five indices chosen by Massey and Denton (1988) are only partially correlated with the corresponding conceptual dimensions, implying that selected indices cannot completely explain the associated aspects of

segregation. Moreover, the indices are correlated not only with their corresponding dimensions but also with others. The indices overlap with each other not only empirically but also theoretically, which prohibits us from evaluating each dimension separately. Reardon and O'Sullivan (2004), on the other hand, did not prove that proposed dimensions are fully covered by their representative indices and that the two indices do not overlap. Furthermore, the lower bound of one index called the spatial information theory segregation index is not specified; thus, it cannot be standardized into a relative form (a revision of this index is given in Sadahiro and Hong, 2013). The representativeness of PC scores proposed by Johnston *et al*. (2007) depends on the data set used in principal component analysis. One PC that is chosen to represent a particular dimension of segregation in one data set may not be able to describe satisfactorily the same dimension in another data set. It is not assured that each dimension is fully represented by a single PC. In addition, the interpretation of PC scores is not straightforward, as there is no one-to-one correspondence between the PCs and the conceptual dimensions of segregation.

As seen above, the existing sets of indices are not entirely appropriate for the measurement of distinct dimensions. To resolve the problems, this paper proposes a new set of indices for measuring segregation. We focus on point distributions that represent zero-dimensional objects and approximate higher-dimensional objects in the real world. Examples include individual persons, animal species, houses, and land parcels. The proposed indices permit us to capture different aspects of segregation with satisfying the desirable properties discussed above. In the next section, we argue that segregation can be divided into three distinctive components, and then propose three indices, each of which evaluates one component of segregation. Section 3 applies the proposed approach to an analysis of three spatial data sets. Section 4 discusses in detail the properties of proposed measures, and Section 5 summarizes the conclusions.

## 2. Method

Many of aspatial indices evaluate segregation in terms of spatial pattern such as concentration, clustering, and unevenness. Spatial indices, on the other hand, often focus on the possibility of social interaction between individuals since it permits us to consider the geographic scale explicitly (Echenique and Fryer, 2007; Kaplan and Holloway, 2001; O'Sullivan and Wong, 2007; Reardon and Firebaugh, 2002; White, 1983; Wong, 1993). Both spatial pattern and social interaction give us a measure of segregation, though they take different perspectives. The former focuses on a source of segregation while the latter puts more weight on the result of segregation. This paper follows the latter approach, i.e., a focus on the social interaction between individuals to consider the spatial aspect of segregation more explicitly.

Segregation indices should address distinctive dimensions of segregation separately without overlap. To this end, this paper distinguishes the different sources of segregation. Existing papers

almost agree on that there are at least spatial and aspatial dimensions in segregation. For instance, among the five dimensions proposed in Massey and Denton (1988), evenness and exposure are aspatial while clustering, centralization, and concentration are spatial (Reardon and O'Sullivan, 2004). Reardon and O'Sullivan (2004) proposed spatial evenness and spatial exposure that put more focus on spatial and aspatial dimensions, respectively. Spatial and aspatial dimensions imply two sources of segregation, i.e., spatial and aspatial sources. Extending these sources, this paper considers three sources of segregation: 1) different arrangement of points, 2) imbalance in the number of points, and 3) lack of diversity in points. The different arrangement of points is a spatial source of segregation while the latter two are aspatial sources.

Each source causes segregation in a different manner. We introduce a term *component* to refer to an element of segregation caused by a specific source. The three sources cause three different components of segregation, which we name the components locational, compositional, and qualitative segregations. In the following subsection, we will explain the concepts of sources and components in more detail, and propose the indices associated with the three components.

*2.1 The three sources and components of segregation*

Figure 1 shows the distribution of individuals of different ethnicities. The black and white points exhibit almost the same spatial arrangement in Figure 1a. On the other hand, points are clearly separated from each other in Figure 1b. The probability of contact between the black and white points in the latter is obviously smaller than that in the former due to the different spatial configuration of the points. The term "locational segregation" refers to segregation due to such different arrangements of points. Figure 1c displays black and white points that share the same proportional distributions, but differ in the number of points. One may think that segregation does not exist because the two types of points share the same proportion at individual locations. However, if we consider the probability of contact between points, each white point has a higher chance of meeting points of the same color; thus, they could be considered to be more segregated than those in Figure 1a. In this paper, we use the term "compositional segregation" to refer to segregation caused by such imbalance between the groups (i.e., point types) in terms of the number of points. In Figure 1d, there are four different types of points. Although they are arranged in the same manner as the points in Figure 1a, the probability that a certain type of point meets a different type of point is clearly larger than in Figure 1a. This implies that the increasing variety of points reduces the level of segregation, and the term "qualitative segregation" refers to segregation caused by a lack of diversity of points.

The difference in the spatial arrangements of points is a spatial source of segregation, although the imbalance in the number of points and the lack of diversity of points are aspatial elements. Thus, the locational segregation can be considered to address the spatial aspect of segregation, and the compositional and qualitative segregation to evaluate the aspatial aspects. Previous studies have

discussed the aspatial aspects of segregation primarily with respect to the imbalance in the proportion of different types of points, which corresponds to the compositional segregation. As seen above, however, a close relationship exists between the variety of points and the level of segregation. Thus, we advocate qualitative segregation as the third component of segregation.

*2.2 Measurement of segregation*

This subsection proposes a set of measures to evaluate the three components of segregation. Suppose a bounded region $R$ of area $T$ in which $\Psi$, a set of $K$ types of points are distributed. The location of $j$th point of type $i$ is denoted by $\mathbf{z}_{ij}$. Let $N_i$ and $N$ be the number of type $i$ points ant the total number of points, respectively.

The measurement of segregation depends on the geographic scale of analysis (Wong, 1993; Wong *et al.*, 1999; Kaplan and Holloway, 2001; Reardon *et al.*, 2008). Figure 1e shows the distribution of black and white points. Segregation is not observed from a usual point of view because both black and white points are uniformly distributed. However, if we take a local view, we find segregation between the points; no white point exists in the close neighborhood of black points, and vice versa. White points are less probable to meet black points in Figure 1e than in Figure 1a. Segregation is caused by the difference in the spatial arrangement of points at a local scale.

To deal with the geographic scale of segregation, we adopt what is called the surface-based approach often used in existing papers (Reardon and O'Sullivan 2004; O'Sullivan and Wong, 2007; Reardon *et al.*, 2008; Spielman and Logan, 2013). The surface-based approach transforms a point distribution into a continuous surface that indicates the degree of spatial clustering of points. The surface represents the probability of contact between points since points are more probable to meet where many points are clustered.

We adopt the Gaussian kernel in converting point distributions into surfaces (Silverman, 1986):

$$k\left(\mathbf{x}, \mathbf{z}_{ij}, h\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\left\|\mathbf{x}-\mathbf{z}_{ij}\right\|}{h}\right)^2}. \tag{1}$$

The parameter $h$ is called the bandwidth that determines the geographic scale parameter of analysis. A large $h$ gives us a global view in the measurement of segregation while a small $h$ permits us to evaluate segregation at a local scale. The degree of spatial clustering of type $i$ points at $\mathbf{x}$ is represented by the summation of Equation (1) in a standardized form:

$$D_i(\mathbf{x}, h) = \frac{N_i \sum_j k(\mathbf{x}, \mathbf{z}_{ij}, h)}{\int_{\mathbf{y} \in R} \sum_j k(\mathbf{y}, \mathbf{z}_{ij}, h) d\mathbf{y}} . \quad (2)$$

We call $D_i(\mathbf{x}, h)$ the *density function* of type *i* points. It shows a large value where points are clustered. It becomes smooth with an increase in the bandwidth *h*.

So far we have assumed that the location of individual points is known. However, spatial data are often provided only in an aggregated form to keep the confidentiality of individuals. This is a primary reason why most of segregation measures proposed in the literature assume spatial data aggregated by spatial units such as census tracts and zip code units. Aggregated data are undesirable since they inherently bring the Modifiable Areal Unit Problem, i.e., calculated indices depend on the spatial units used for data aggregation (Wong, 1993; Reardon and O'Sullivan, 2004). Though this problem is not fully avoidable, the method proposed in this paper can be applied to aggregated spatial data. Suppose region *R* is divided into subregions $R_1, R_2, ..., R_V$. Let us denote $T_v$ and $p_{iv}$ as the area of *v*th region and the number of type *i* points in the *v*th region, respectively. A simple approach is to substitute the point density $p_{iv}/T_v$ for $D_i(\mathbf{x})$ everywhere in the *v*th region. In this method, however, we cannot evaluate the geographic scale of segregation. A more sophisticated alternative is to assume that points are uniformly distributed in every region and calculate the density functions of points based on Equation (2):

$$D_i(\mathbf{x}, h) = \frac{N_i \sum_v \int_{\mathbf{z} \in R_v} \frac{p_{iv}}{T_v} k(\mathbf{x}, \mathbf{z}, h) d\mathbf{z}}{\int_{\mathbf{y} \in R} \sum_v \int_{\mathbf{z} \in R_v} \frac{p_{iv}}{T_v} k(\mathbf{y}, \mathbf{z}, h) d\mathbf{z} d\mathbf{y}} . \quad (3)$$

Unlike aspatial measures, we can consider the geographic scale of analysis explicitly by surface conversion.

Using the density functions defined by Equation (2), we define a *local segregation measure* as

$$s(\mathbf{x}, \Psi, h) = \sum_i \left\{ \frac{D_i(\mathbf{x}, h)}{D(\mathbf{x}, h)} \right\}^2 , \quad (4)$$

where:

$$D(\mathbf{x}, h) = \sum_i D_i(\mathbf{x}, h) .$$

(5)

Though the entropy index has been often used as a measure of segregation in the literature, this paper adopts the above definition because the entropy index cannot evaluate qualitative segregation (Sadahiro and Hong, 2013). The local segregation measure estimates the level of the imbalance between different types of points at location **x**. It can be interpreted as the sum of the proportion of each type of point that needs to change for balanced proportions. Figure 2a shows the density functions of five types of points in a 1D space. Figure 2b shows the distribution of $s(\mathbf{x}, \Psi, h)$ calculated from the density functions in Figure 2a. It has a large value if the proportion of different types of points at location **x** varies widely. It becomes small when each type of point occupies a similar proportion of the total.

We define a global measure of segregation by integrating $s(\mathbf{x}, \Psi, h)$ weighted by the total density of points $D(\mathbf{x}, h)$ over the entire region:

$$S(\Psi, h) = \frac{1}{N} \int_{\mathbf{x} \in R} D(\mathbf{x}, h)\, s(\mathbf{x}, \Psi, h)\, d\mathbf{x}.$$

(6)

This indicates the average degree of segregation over the entire region. We call it the *overall segregation measure*, whose range is $1/K < S(\Psi, h) \leq 1$ (a proof is given in Sadahiro and Hong, 2013). As seen in Equation (6), it is a function of point set $\Psi$ and bandwidth $h$, which implies that the measure depends on the geographic scale of analysis represented by $h$.

We then define the measures for the three components of segregation. To evaluate a particular component, we need to consider two situations: where the source of that component is present and where it is absent. We first calculate the overall segregation measure $S(\Psi, h)$ by assuming the latter and comparing the result with the observed value; the difference represents the effect of the source of that component.

Next we begin with the evaluation of locational segregation. The source of locational segregation is absent when all types of points have exactly the same spatial arrangement. In this case, the fragmented pieces of $K$ types of points $\{q_1, q_2, \ldots, q_K\}$ are distributed at every location of the original points, where $q_i$ is a piece of $N_i/N$ fraction of type $i$ point (Figure 3). In Figure 3a, there are four black points and eight white points. If we replace every point in this figure with a point comprising 1/3(=4/12) black and 2/3(=8/12) white points (Figure 3b), the black and white points have exactly the same spatial arrangement. In this situation, the density functions of points are given by,

$$D_i(\mathbf{x}, h) = \frac{N_i}{N} D(\mathbf{x}, h), \forall i.$$

(7)

This equation indicates that the relative share of each type of point is constant everywhere in $R$.

We can confirm this in Figure 2c, which shows the density functions above applied to the

data in Figure 2a. The difference in the spatial arrangement of points disappears in Figure 2c. This makes the distribution of the local segregation measure $s(\mathbf{x}, \Psi, h)$ uniform, as seen in Figure 2d; consequently, it reduces the overall level of segregation. Substituting Equation (7) into Equations (4) and (6), we obtain

$$S'(\Psi) = \sum_i \left( \frac{N_i}{N} \right)^2 . \tag{8}$$

The reduction of overall segregation measure indicates the degree of locational segregation:

$$\begin{aligned} S_L(\Psi, h) &= S(\Psi, h) - S'(\Psi) \\ &= S(\Psi, h) - \sum_i \left( \frac{N_i}{N} \right)^2 . \end{aligned} \tag{9}$$

$S_L(\Psi, h)$ is the *locational segregation measure*, and its range is $0 \le S_L(\Psi, h) < S(\Psi, h)$. This index indicates the variation in the proportional share of points across locations, and has a large value when the proportional share of points varies greatly across locations (i.e., when the different types of points are separated from each other). The value of $S_L(\Psi, h)$ becomes zero when the proportional share is constant in $R$.

We then consider compositional segregation. To evaluate this component, we again compare two hypothetical cases. The first case is the same as the one that was considered during the evaluation of locational segregation, i.e., the absence of a difference in the spatial arrangements of points. For the second, we assume that both a difference in the spatial arrangement of points and an imbalance in the actual number of points do not exist. To represent this situation mathematically, suppose that the fragmented pieces of $K$ types of points $\{q_1, q_2, \ldots, q_K\}$ are distributed at every location of the original points, where each $q_i$ has the $1/K$ proportion of the total of that point. In this situation, the density functions of the points are given by,

$$D_i(\mathbf{x}, h) = \frac{D(\mathbf{x}, h)}{K}, \forall i . \tag{10}$$

This equation indicates that all types of points' are present in the same proportion across every location in $R$ (Figure 2e), and is applicable where both a difference in the spatial arrangements of points and an imbalance in the actual number of points are absent.

Figure 2f shows the distribution of the local segregation measure for this situation. Compared with Figure 2d, it decreases everywhere in $R$, so does the overall segregation. This reduction occurs because the imbalance in the number of points has vanished. Substituting Equation (10) into

Equations (4) and (6), we obtain

$$S''(\Psi) = \frac{1}{K}.$$

(11)

The degree of compositional segregation can be estimated by the reduction from $S'(\Psi, h)$ to $S''(\Psi, h)$:

$$\begin{aligned} S_C(\Psi) &= S'(\Psi) - S''(\Psi) \\ &= \sum_i \left( \frac{N_i}{N} \right)^2 - \frac{1}{K} \end{aligned} .$$

(12)

$S_C(\Psi)$ is the *compositional segregation measure*, and its range is $0 \le S_C(\Psi) < S(\Psi, h)$. Unlike $S_L(\Psi, h)$, we omit $h$ in the notation since Equation (12) is independent of the bandwidth $h$. This measure has a large value when the proportional share of the point types is unbalanced across locations in $R$, and becomes zero when all the types of points are present in the same proportion at every location.

Now we move on to the evaluation of qualitative segregation. To explore this component of segregation, we compare the case wherein a difference in the spatial arrangement of points and an imbalance in the number of points do not exist (i.e., Equation (10)) with another hypothetical case wherein all three sources of segregation do not exist. To represent the latter case, we first assume that the fragmented pieces of $K$ types of points $\{q_1, q_2, \ldots, q_K\}$ are distributed at every location of the original points, where each $q_i$ has a $1/K$ fraction of the type $i$ points. Subsequently, we increase $K$ infinitely to resolve the lack of diversity of points. Under this circumstance, the density function of the points approaches zero:

$$D_i(\mathbf{x}, h) \to 0, \forall i$$

(13)

Figure 2g shows the density functions of the points when all three sources of segregation do not exist. The local segregation measure infinitely approaches zero everywhere in $R$ as illustrated in Figure 2h, as does the overall segregation. Thus, the qualitative segregation can be evaluated using the reduction from $S''(\Psi)$ to zero:

$$\begin{aligned} S_Q(\Psi) &= S''(\Psi) - 0 \\ &= \frac{1}{K} \end{aligned} .$$

(14)

$S_Q(\Psi)$ is the *qualitative segregation measure*. This measure represents the homogeneity of the points, ranging from zero to one. A large value of $S_Q(\Psi)$ implies that there are only a few different types of points in $R$, whereas a small value indicates that a wide variety of points exist.

The locational segregation evaluates the spatial aspect of segregation, and the compositional and qualitative segregation deal with the aspatial aspects of segregation as mentioned previously. This is, in fact, reflected in the definitions of these measures. The definition of $S_L(\Psi, h)$ in Equation (9) contains the bandwidth $h$ that determines the geographic scale of analysis, while $S_C(\Psi)$ and $S_Q(\Psi)$ are defined only by aspatial variables as seen in Equations (12) and (14).

The relationship between these three measures can be expressed as:

$$S(\Psi, h) = S_L(\Psi, h) + S_C(\Psi) + S_Q(\Psi). \tag{15}$$

This indicates that segregation can be decomposed into locational, compositional, and qualitative segregations. The measures cover the whole range of segregation without overlap. This relationship makes the three measures comparable with each other so that we can tell what sources are more influential on segregation. If, for instance, $S_L(\Psi, h)$ is larger than $S_C(\Psi)$, the difference in the spatial arrangement of points is a more influential source than the imbalance in the number of points. If $S_C(\Psi)$ is predominant among the three measures, the imbalance in the number of points is a primary source of segregation among the three sources.

*2.3 Geographic scale of segregation*

The measurement of segregation depends on the geographic scale of analysis as mentioned earlier. This paper argues that the scale dependence occurs because segregation measured at a certain scale consists of a collection of smaller elements observed at larger scales. This subsection discusses the geographic scale of segregation in detail, and proposes another segregation measure.

Of the three components of segregation, we focus on locational segregation here because the others represent aspatial aspects of segregation. Figure 4a illustrates $S_L(\Psi, h)$ as a function of the bandwidth $h$. This figure is equivalent to the dissimilarity index $D$ represented as a function of grid size (Wong *et al*., 1999), the *GD* index proposed by Wong (2005), and the segregation profile proposed by Reardon *et al*. (2008). The measure $S_L(\Psi, h)$ is usually a monotonically decreasing function of scale parameter $h$, since the level of observed segregation tends to decrease with an increase in the scale of analysis.

The surface conversion is a kind of low-pass filter (Silverman, 1986; Sonka *et al.*, 2014), i.e., it conceals the spatial patterns in a point distribution with a scale smaller than the scale parameter $h$. The surface obtained by the surface conversion of scale $h$ represents a collection of spatial patterns of scales larger than $h$ in the point distribution, and consequently, the measure $S_L(\Psi, h)$ is the summation of segregation of scales larger than $h$. This implies that the locational segregation can be decomposed further into segregations of different scales, which we call *differential locational segregations*. The locational segregation observed at scale $h$ is a collection of differential locational

segregations measured at scales larger than $h$. This relationship is mathematically represented as

$$S_L(\Psi, h) = \int_h^\infty S_{DL}(\Psi, u)\,\mathrm{d}u\,, \tag{16}$$

where $S_{DL}(\Psi, h)$ is the *differential locational segregation measure* of scale $h$. Differentiating both sides of this equation with respect to $h$, we obtain

$$S_{DL}(\Psi, h) = -\frac{\mathrm{d}}{\mathrm{d}h} S_L(\Psi, h)\,, \tag{17}$$

Figure 4b presents $S_{DL}(\Psi, h)$ calculated from $S_L(\Psi, h)$ in Figure 4a. The measure $S_L(h_1)$ in Figure 4a is equal to the gray-shaded area in Figure 4b.

Calculation of $S_{DL}(\Psi, h)$ is, in a sense, a frequency resolution of $S_L(\Psi, h)$. Substituting Equation (16) into (15), we obtain

$$S(\Psi, h) = S_C(\Psi) + S_Q(\Psi) + \int_h^\infty S_{DL}(\Psi, u)\,\mathrm{d}u\,. \tag{18}$$

This equation indicates that segregation measured at scale $h$ can be decomposed into compositional segregation, qualitative segregation, and differential locational segregations of scales larger than $h$. Differential locational segregations of scales smaller than $h$ are concealed in $S(\Psi, h)$, which yields the scale dependence of $S(\Psi, h)$.

*2.4 Evaluation of spatial segregation by the entropy index*

The entropy index has been often used as a measure of segregation in the literature (Theil and Finizza, 1971; White, 1986; Allen and Turner, 1989; Hårsman and Quigley, 1995; Reardon and O'Sullivan, 2004). This paper adopts different measures because the entropy index cannot evaluate the three components of spatial segregation separately. However, if such a separation is not necessary, we can use the entropy index within our framework. This subsection briefly describes the evaluation of spatial segregation by the entropy index.

We replace the local segregation measure $s(\mathbf{x}, \Psi, h)$ defined by Equation (4) with that based on the entropy index:

$$s_e(\mathbf{x}, \Psi, h) = 1 + \sum_i \left\{ \frac{D_i(\mathbf{x}, h)}{\sum_k D_k(\mathbf{x}, h)} \log \frac{D_i(\mathbf{x}, h)}{\sum_k D_k(\mathbf{x}, h)} \right\}. \tag{19}$$

The overall segregation becomes

$$S_e\left(\boldsymbol{D},\Psi,h\right)=\frac{\int_{\mathbf{x}\in R}D\left(\mathbf{x},h\right)s_e\left(\mathbf{x},\Psi,h\right)d\mathbf{x}}{\int_{\mathbf{x}\in R}D\left(\mathbf{x},h\right)d\mathbf{x}}.$$

(20)

The locational, compositional, and qualitative segregation measures become

$$\begin{aligned}S_{Le}\left(\boldsymbol{D},\Psi,h\right)&=S_e\left(\boldsymbol{D},\Psi,h\right)-S_e\left(\boldsymbol{D_L},\Psi,h\right)\\&=\frac{\int_{\mathbf{x}\in R}D\left(\mathbf{x},h\right)\sum_i\left\{\frac{D_i\left(\mathbf{x},h\right)}{\sum_k D_k\left(\mathbf{x},h\right)}\log\frac{D_i\left(\mathbf{x},h\right)}{\sum_k D_k\left(\mathbf{x},h\right)}\right\}d\mathbf{x}}{\int_{\mathbf{x}\in R}D\left(\mathbf{x},h\right)d\mathbf{x}}-\sum_i\frac{N_i}{N}\log\frac{N_i}{N}\end{aligned},$$

(21)

$$\begin{aligned}S_{Ce}\left(\boldsymbol{D},\Psi,h\right)&=S_e\left(\boldsymbol{D_L},\Psi,h\right)-S_e\left(\boldsymbol{D_C},\Psi,h\right)\\&=1+\sum_i\frac{N_i}{N}\log\frac{N_i}{N}\end{aligned},$$

(22)

and

$$\begin{aligned}S_{Qe}\left(\boldsymbol{D},\Psi,h\right)&=S_e\left(\boldsymbol{D_C},\Psi,h\right)\\&=0\end{aligned},$$

(23)

respectively. Equation (23) indicates that the entropy index conceals the qualitative segregation. The ranges of the segregation measures are $0\le S_{Le}(\boldsymbol{D},\Psi,h)<S_e(\boldsymbol{D},\Psi,h)$ and $0\le S_{Ce}(\boldsymbol{D},\Psi,h)\le S_e(\boldsymbol{D},\Psi,h)$ (a proof of the former is given in Appendix A4). The equalization measures can be calculated in a similar way as that in the previous subsection.

Segregation measures defined from Equation (4) and those based on Equation (19) have both advantages and disadvantages. One strength of the former is that it permits us to decompose the aspatial components of spatial segregation into the compositional and qualitative segregations. The latter is advantageous in that the entropy index has already been widely used in the measurement of spatial segregation. Its implementation can be done with a slight modification of existing programs.

*2.5 Comparison with Reardon and O'Sullivan's paper*

This paper shares an intention and a principle with Reardon and O'Sullivan (2004). Both aim to present distinct dimensions of spatial segregation and define segregation measures based on the location of individual points. This subsection briefly compares the two papers in terms of segregation dimensions and measures.

Reardon and O'Sullivan (2004) advocates the spatial exposure and the spatial evenness as

the two dimensions of spatial segregation. The latter corresponds to the locational segregation proposed in this paper. Spatial exposure contains at least the compositional segregation, but it is not clear whether it also contains the qualitative segregation since they do not explicitly discuss the spatial exposure in terms of the multigroup segregation.

As shown in Section 2.2, the locational, compositional and qualitative segregations are independent with each other. This implies that the spatial exposure and evenness are also independent. We can change the former with keeping the latter, and vice versa. We should note, however, that this does not assure the independency between their segregation measures. It is unknown whether their spatial exposure index and spatial evenness indices are independent with each other.

To evaluate the spatial evenness, they advocate the spatial information theory segregation index (for details, see Reardon and O'Sullivan (2004)). Let $\tau_p$ and $\tilde{\pi}_{pm}$ be the density of points at location $p$ and the proportion of type $m$ points in the local environment of $p$, respectively. The total number of points is given by

$$T = \int_{p \in R} \tau_p dp \ . \tag{24}$$

The entropy index at location $p$ is

$$\tilde{E}_p = -\sum_m \tilde{\pi}_{pm} \log \tilde{\pi}_{pm} \ . \tag{25}$$

The spatial information theory segregation index is defined by

$$\widetilde{H} = 1 - \frac{1}{TE} \int_{p \in R} \tau_p \tilde{E}_p dp \ , \tag{26}$$

where $E$ is the overall entropy. The index corresponds to $S_L(\boldsymbol{D}, \Psi, h)$ in this paper in that both compare the spatial segregation with the overall segregation. While $S_L(\boldsymbol{D}, \Psi, h)$ considers the difference between the segregations, $\widetilde{H}$ evaluates their ratio.

The range of $S_L(\boldsymbol{D}, \Psi, h)$ is known as $0 \leq S_L(\boldsymbol{D}, \Psi, h) < S(\boldsymbol{D}, \Psi, h)$, while that of $\widetilde{H}$ is not known. The latter can take a negative value, and we cannot compare the spatial segregation between different patterns as mentioned in Section 1. This is presumably because the local entropy is calculated from the local density of points while the overall entropy is based on the total number of points. The summation of the local density is not always equal to the total number of points.

To avoid this problem, we suggest using the variables defined in the local environment consistently to calculate segregation measures. Let $\tilde{\tau}_p$ be the density of points in the local environment of $p$. We define the total volume of points $T'$ as

$$T' = \int_{p \in R} \tilde{\tau}_p dp \,.$$

(27)

Using these variables, we modify the spatial information theory segregation index as

$$\widetilde{H}' = 1 - \frac{1}{T'E'} \int_{p \in R} \tilde{\tau}_p \tilde{E}_p dp \,,$$

(28)

where

$$E' = -\sum_m \frac{\int_{p \in R} \tilde{\tau}_{pm} dp}{T'} \log_M \frac{\int_{p \in R} \tilde{\tau}_{pm} dp}{T'} \,.$$

(29)

The modified index would range from 0 to 1.

## 3. Applications

This section tests the validity of the proposed approach, by applying it to an analysis of three data sets. For simplicity, we omit the indicators ($\Psi$, $h$) and ($\Psi$) used in the notations hereafter.

### *3.1 Properties of the measures of segregation*

This subsection explores the properties of the three segregation measures using a synthetic data set. To focus on the relationship between the measures, we assume that point distributions are already converted into density functions.

We first discuss the relationship between the locational and compositional segregation measures. Figures 5a and 5b show the density distributions of two types of points in a 1D space. In both figures, the proportion of each type of point varies in the vertical direction, while it is relatively constant in the horizontal direction. The spatial arrangement of the points becomes more similar in every row from columns A to E. Figures 5c and 5d present the segregation measures calculated for the density distributions in Figures 5a and 5b, respectively. The locational segregation measure $S_L$ decreases from columns A to E as the difference in the spatial arrangement of the points decreases. The imbalance in the proportion of the two types of point increases from columns 5 to 1 in Figures 5a and 5b; consequently, it raises the compositional segregation measure $S_C$ from columns 5 to 1 in Figures 5c and 5d. The qualitative segregation measure $S_Q$ is constant at 0.5 in all the patterns. The compositional and qualitative segregation measures occupy the majority share of the overall segregation, implying that aspatial sources are the primary source of segregation in Figures 5a and 6b. From columns A to E in Figure 5c, $S_L$ changes when holding $S_C$ and $S_Q$ constant. In Figure 5d, on the other hand, $S_C$ changes when holding $S_L$ and $S_Q$ constant. This occurs because $S_L$ and $S_C$ do not overlap

with each other, i.e., they evaluate different aspects of segregation.

Figures 5e and 5f display the modified spatial information theory segregation index $\widetilde{H}'$ defined by Sadahiro and Hong (2013) calculated for the density distributions shown in Figures 5a and 5b. Comparing Figures 5c-5f, we notice that the distribution of $\widetilde{H}'$ is more similar to that of $S_L$ than to that of $S$. This is reasonable, considering that $\widetilde{H}'$ represents the spatial aspect of segregation. Figures 5g and 5h display the index of dissimilarity developed by Duncan and Duncan (1955). The two figures look similar to Figures 5e and 5f and thus the distribution of $S_L$ in Figures 5c and 5d. Though the index of dissimilarity does not consider the spatial aspect of segregation explicitly, this index looks like being sensitive to the spatial arrangement of points.

Let us move on to the relationship between the compositional and qualitative segregation measures. Figures 6a and 6c show the density distributions of points and their segregation measures, respectively. $S_L$ is constantly zero because all the density distributions are uniform. The number of types of point increases from rows 1 to 5, while the imbalance in the proportion of points decreases from columns A to E. Figure 6c demonstrates that the degree of qualitative segregation decreases with an increase in the number of types of point, as does the overall segregation. This supports our earlier discussion in Figure 1d (i.e., that an increasing variety of points decreases segregation). We can also confirm our observation in Figure 1c (i.e., an imbalance in the number of points increases segregation), by looking at the increase in $S_C$ and $S$ from the columns E to A in Figure 6c.

We then discuss the relationship between the locational and qualitative segregation measures. Figures 6b and 6d show the density distributions of points and their segregation measures, respectively. The measure $S_C$ is constantly zero because all the types of points are present in the same proportion. Similar to Figure 6c, Figure 6d shows that an increase in the number of types of point reduces $S_Q$ and $S$ from rows 1 to 5. The locational and compositional segregation measures change while keeping $S_Q$ unchanged in Figures 6c and 6d, implying that $S_L$ and $S_C$ does not overlap with $S_Q$.

Figures 6e and 6f show the modified spatial information theory segregation index $\widetilde{H}'$. The index is always equal to zero in Figure 6e, as it does not consider the lack of diversity of points as a separate source of segregation. The distribution of $\widetilde{H}'$ in Figure 6f is more similar to the distribution of $S_L$ in Figure 6d, which is compatible with our observation in Figure 5.

*3.2 Geographic scale of segregation*

This subsection investigates the geographic scale of segregation. Figure 7a displays seven different patterns of two types of points distributed on a 1D space of length 1. We convert the point distributions into density functions by Gaussian kernel smoothing of bandwidth $w$, ranging from 0.001 to 1.

Figure 7b shows the relationship between the bandwidth $h$ and the proposed segregation measures. The scale of analysis changes from local to global with an increase of $h$. The vertical axis

shows $S_L$+$S_C$, i.e., the sum of the locational and compositional segregation measures. The compositional and qualitative segregation measures are constant in all the patterns, as they are independent of the scale of analysis. The latter is always 0.5(=1/2). $S_C$ is zero in patterns A to D, and it increases from pattern E to G with an increasing imbalance in the number of points, as indicated by the arrows in Figure 7b. The locational segregation measure $S_L$ decreases monotonically with an increase in the geographic scale, which reflects a shift in our viewpoint from local to global. $S_L$ infinitely approaches zero, even when the points are clearly separated as shown in pattern D. It has a large value, however, at a very local scale, even when the points are globally well mixed, as shown in pattern A. The locational segregation increases from patterns A to D, regardless of scale $h$, with the clustering of the white points (Figure 7b).

Figure 7c demonstrates the relationship between the bandwidth $h$ and the differential locational segregation measure $S_{DL}$. The peak of $S_{DL}$ shifts from a local to a global scale from patterns A to D. This indicates that the locational segregation in pattern A is primarily due to a small-scale difference in the spatial arrangement of the points, while large-scale differences are more predominant in pattern D. A similar shift of the peak is observed from patterns E to G, reflecting an increase in the distance between the white points in Figure 7a.

*3.3 Application of the analysis to a real data set*

In this subsection, we use the proposed approach to analyze a real data set as a test of its practical feasibility. As mentioned in Subsection 2.2, many segregation measures are defined based on spatially-aggregated data. We thus use land use data in a lattice form to evaluate the performance of segregation measures calculated based on aggregated data.

We examine the changes in land use mixture observed in the urbanization process from 1974 to 1994 in Chiba, Japan. Chiba is adjacent to the east side of Tokyo Metropolis, the population of Chiba had drastically increased with the expansion of the Tokyo metropolitan area during this period. Urban sprawl spread from west to east all over Chiba, often with an undesirable land use mixture that caused traffic congestion and deteriorated the residential environment. The causes and processes of the land use mixture vary between different regions, and they have not yet been analyzed in detail. This subsection reveals the detailed process of land use mixture in Chiba using the proposed measures to evaluate the degree of mixture between different land uses.

We use the land use data of a 10-m resolution lattice provided by the Geospatial Information Authority of Japan. The data were generated by recording land use categories at sample locations that are distributed regularly at 10-m intervals on aerial photographs. To evaluate the land use mixture at a local scale, we aggregated the data into the lattice format with 500 m resolution and calculated the overall segregation measure $S$ in each cell. We did not perform the kernel smoothing to focus on the change of land use mixture rather than its spatial pattern.

Figure 8 shows the distribution of the local segregation measure $s(\mathbf{x})$. This measure indicates the imbalance in the proportion of different land use categories at location $\mathbf{x}$, which usually decreases with a change in the land use mixture. As seen in Figure 8, $s(\mathbf{x})$ had decreased all over Chiba during this period (dark shades indicate small $s(\mathbf{x})$). Land use mixture had spread in this area, especially along railway lines. This is confirmed in Figure 9a, wherein the compositional segregation measure $S_C$ had decreased from 0.41 to 0.34, as $S_C$ reflects the mean of $s(\mathbf{x})$ of the entire region. Figure 9a shows that the locational segregation measure $S_C$ had slightly increased from 0.17 to 0.18. This indicates that the variation in the pattern of land use mixture had gradually increased during this period. The spatial information theory index shown in Figure 9a is similar to $S_L$, which is again consistent with the result obtained in Subsection 3.1.

We then examine the changes in land use mixture in more detail in six subregions, as shown in Figure 8. Subregions R1 and R2 experienced a rapid expansion of urban areas during the 1980s, primarily as a result of new town construction. Subregions R3 and R4 both contain urban areas with a long history, where the land use patterns were stable over the same period. Subregions R5 and R6 are a mixture of urban, suburban, and rural areas. The former contains an old historical town, while the latter is a new town located at the border between urban and rural areas.

Figure 9b-e shows the segregation measures calculated in the six subregions from 1974 to 1994. The measures showed similar changes, except those of subregion R3. These subregions represent the overall tendency of the progress of land use mixture in Chiba. In subregion R3, $S_L$ decreased and $S_C$ increased during this period. Having examined the land use pattern in R3 in detail, we found that the urbanization was already complete in 1974. Most of this subregion was already covered by residential areas, and no significant changes occurred from 1974 to 1994. However, in general, we can conclude that the land use mixture had changed all over Chiba, with an increase in the variation in the land use mixture pattern from 1974 to 1994. Comparing Figures 9c and 9e, we do not find a clear similarity between $S_L$ and the spatial information theory index. However, the spatial information theory index is still closer to $S_L$ than $S_L$ and $S_L$, which is not incompatible with the results obtained in Subsection 3.1.

## 4. Properties of the proposed measures

The discussion in Section 1 suggests five desirable properties of segregation measures: 1) each component should be fully represented by a single measure, 2) there should be a set of measures that addresses all aspects of segregation without overlap, 3) the measures should be consistently defined independently of the data used in empirical studies, 4) the range of the measures should be analytically known, and 5) interpretation of the measures should be intuitive and straightforward.

The proposed measures are defined based on different sources of segregation independently, and they evaluate the different components of segregation. The measures have a one-to-one

correspondence to the components of segregation. Equation (15) indicates that they cover the whole range of segregation without overlap. This confirms that the measures meet the first two criteria. We should note, however, that the separability of components is not equivalent to empirical independence. A typical example is Figures 5a and 5c, where $S_L$ increases and $S_C$ decreases simultaneously from rows 1 to 5. A correlation between measures can be observed even if the measures do not overlap with each other. Empirical dependence is not incompatible with the separability of components.

The measures also satisfy the third and fourth criteria: the components of segregation we propose in this paper and their associated measures are defined deductively prior to empirical applications, and their range is analytically known. As seen in the applications in Section 3, the interpretation of the measures is straightforward. It is primarily because they clearly correspondent to each component of segregation.

Reardon and O'Sullivan (2004) suggested eight criteria for segregation measures: 1) scale interpretability, 2) arbitrary boundary independence, 3) location equivalence, 4) population density invariance, 5) composition invariance, 6) transfers and exchange, 7) additive spatial decomposability, and 8) additive grouping decomposability. These criteria are satisfied by the definition of $D_i(\mathbf{x}, h)$ and $s(\mathbf{x}, \Psi, h)$, and Equation (15), except scale interpretability and composition invariance. Scale interpretability requires a measure to become zero when all the groups share the same proportion in a local environment. The measure $S(\Psi, h)$ does not satisfy this criterion, as it becomes $1/K$ under this condition. However, if we interpret this criterion as being a requirement for a clear and interpretable relationship between a measure and the degree of segregation, $S(\Psi, h)$ satisfies this criterion since it increases monotonically with the degree of segregation. Composition invariance, on the other hand, was originally advocated by James and Taeuber, 1985, to which some papers, including Reardon and O'Sullivan (2004), raise objections. However, we can modify the proposed measures to satisfy it by replacing $D_i(\mathbf{x}, h)$ with the standardized density distribution of points $\eta_i(\mathbf{x}, h)$:

$$\eta_i\left(\mathbf{x}, h\right) = \frac{D_i\left(\mathbf{x}, h\right)}{\int_{\mathbf{y}\in R} D_i\left(\mathbf{y}, h\right) \mathrm{d}\mathbf{y}}. \tag{30}$$

Similar to existing measures, our measures aim to evaluate the different aspects of segregation. The locational segregation measure $S_L(\Psi, h)$ captures the spatial aspect of segregation, which is also evaluated by evenness and concentration indices such as the index of dissimilarity, the Gini index, the entropy-based indices, and the spatial information theory segregation index (Massey and Denton, 1988; Reardon and O'Sullivan, 2004). Among those, the spatial information theory segregation index often yields similar result with the locational segregation measure as seen in Section 3. The compositional and qualitative segregation measures $S_C(\Psi)$ and $S_Q(\Psi)$ focus on the aspatial aspect of segregation, which are similar to exposure indices including the isolation index, interaction

index, and the spatial exposure index. Compared with those existing indices, one strength of our measures is that they satisfy the five desirable properties of segregation measures as discussed above. The three measures evaluate the three different aspects of segregation without overlap, and they cover the whole range of segregation. This permits us to evaluate the three components separately, and compare the effect of three sources with each other.

The three measures are, in a sense, the minimum set of measures for evaluating the different aspects of segregation. This paper, however, neither prohibits us from using other existing measures nor forces us to use all the three proposed measures. Existing measures are still useful to evaluate segregation from various perspectives, especially when the distinctiveness of measures is not essential. One may argue that no segregation is observed in Figure 1c where black and white points share the same spatial pattern. If social interaction is not considered to evaluate segregation, it is enough to use only $S(\Psi, h)$ or $S_L(\Psi, h)$.

The differential locational segregation is not identical to the geographic scale discussed in Reardon *et al*. (2008), although both address the geographical scale of segregation. Differential locational segregation is a subcomponent of locational segregation, while the geographic scale is an independent dimension distinct from the spatial evenness.

**5. Conclusion**

This paper develops a new approach to the measurement of segregation among point distributions. The major contribution of the paper is the development of new measures that evaluate segregation from different perspectives. We argued that segregation can be decomposed into three components that are supported by three measures respectively. The measures satisfy the five desirable properties of segregation measures discussed in Subsection 2.2 that are not fully possessed by existing indices. Since the three components of segregation are separately measurable, the proposed measures can serve as a minimum set of measures that address distinct dimensions of segregation. The second contribution of the paper is to advocate qualitative segregation. Existing studies consider the aspatial aspects of segregation primarily with respect to the imbalance in the proportion of different types of points. As seen in Subsection 2.1, however, there exists a close relationship between the variety of points and the level of segregation. Consideration of qualitative segregation provides us a more detailed view of segregation. The third contribution of the paper is to introduce differential locational segregation. It is a representation of multi-scale aspect of segregation. Differential locational segregation permits us to consider the geographic scale of segregation explicitly, i.e., to decompose locational segregation into smaller elements of different scales.

Finally, we discuss some limitations of the paper and potential directions for future research.

First, this paper does not evaluate the statistical significance of segregation measures. One method to test the significance is to calculate the measures of spatial autocorrelation for individual

types of points. Johnston *et al*. (2010) and Poulsen *et al*. (2011) used Moran's *I* and Getis's *G* statistics to test the significance of the spatial clustering of individual ethnic groups. This method, however, does not directly evaluate the segregation of more than two types of points. A statistical test that considers the segregation of multiple types of points simultaneously needs to be developed.

Second, this paper evaluates segregation of points without considering their attributes in detail. This treatment, however, may not be appropriate in the real world. For instance, segregation between two population groups with very different socio-economic levels may have a larger impact on society in comparison with that of two groups with similar profiles. In such a case, we should put more weight on the former in the evaluation of segregation. To accomplish this, we should explicitly take into account the attributes of points in segregation measurement.

Third, as Reardon (2009) stated, most of the existing studies have discussed segregation among groups classified by nominal variables. However, groups can be categorized by other scales such as ordinal, interval, and ratio variables. Reardon (2009) developed ordinal segregation measures, and Reardon and Bischoff (2011) used them in the analysis of income segregation in the U.S. from 1970 to 2000. Extension of the proposed measures in this direction might be an important topic for future research.

**References**


Duncan OD, Duncan B, 1955, "A methodological analysis of segregation indexes" *American Sociological Review* **20** 210-217

Echenique F, Fryer RG, 2007, "A measure of segregation based on social interactions" *Quarterly Journal of Economics* **127** 1287-1338

James D R, Taeuber KE, 1985, "Measures of segregation" *Sociological Methodology* **14** 1-32

Johnston R, Poulsen M, Forrest J, 2007, "Ethnic and racial segregation in U.S. metropolitan areas, 1980-2000: the dimensions of segregation revisited" *Urban Affairs Review* **42** 479-504

Johnston R, Poulsen M, Forrest J, 2010, "Evaluating changing residential segregation in Auckland, New Zealand, using spatial statistics" *Tijdschrift voor Economische en Sociale Geografie* **102** 1-23

Kaplan DH, Holloway SR, 2001, "Scaling ethnic segregation: causal processes and contingent outcomes in Chinese residential patterns" *GeoJournal* **53** 59-70

Massey DS, Denton SA, 1988, "The dimensions of residential segregation" *Social Forces* **67** 281-315

Morrill RL, 1991, "On the measure of geographical segregation" *Geography Research Forum* **20** 210-217

O'Sullivan D, Wong DWS, 2007, "A surface-based approach to measuring spatial segregation" *Geographical Analysis* **39** 147-168

Poulsen M, Johnston R, Forrest J, 2011, "Using local statistics and neighbourhood classifications to

portray ethnic residential segregation: a London example" *Environment and Planning B* **38** 636-658

Reardon SF, 2009, "Measures of ordinal segregation" *Research on Economic Inequality* **17** 129-155

Reardon SF, Bischoff K, 2011, "Income inequality and income segregation" *American Journal of Sociology* **116** 1092-1153

Reardon SF, Firebaugh G, 2002, "Measures of multigroup segregation" *Sociological Methodology* **32** 33-67

Reardon SF, O'Sullivan D, 2004, "Measures of spatial segregation" *Sociological Methodology* **34** 121-162

Reardon SF, Matthews SA, O'Sullivan D, Lee BA, Firebaugh G, Farrell CR, Bischoff K, 2008, "The geographic scale of metropolitan racial segregation" *Demography* **45** 489-514

Sadahiro Y, Hong S-Y, 2013, "Decomposition approach to the measurement of spatial segregation" *Discussion Paper Series No. 119, Center for Spatial Information Science, The University of Tokyo* (available from http://www.csis.u-tokyo.ac.jp/dp/119.pdf)

Silverman, BW, 1986, *Density Estimation for Statistics and Data Analysis* (CRC Press, Boca Raton)

Sonka M, Hlavac V, Boyle R, 2014 *Image Processing, Analysis and Machine Vision (Fourth Edition)* (Cengage Learning, Stanford, CT)

Spielman SE, Logan JR, 2013, "Using high-resolution population data to identify neighborhoods and establish their boundaries" *Annals of the Association of American Geographers* **103** 67-84

Theil H, Finizza AJ, 1971, "A note on the measurement of racial integration of schools" *Journal of Mathematical Sociology* **1** 187-193

White MJ, 1983, "The measurement of spatial segregation" *American Journal of Sociology* **88** 1008-1018

Wong DWS, 1993, "Spatial indices of segregation" *Urban Studies* **30** 559-572

Wong DWS, 2005, "Formulating a general spatial segregation measure" *Professional Geographer* **57** 285-294

Wong DWS, Lasus H, Falk RF, 1999, "Exploring the variability of segregation index D with scale and zonal systems: an analysis of 30 US cities" *Environment and Planning A* **31** 507-522

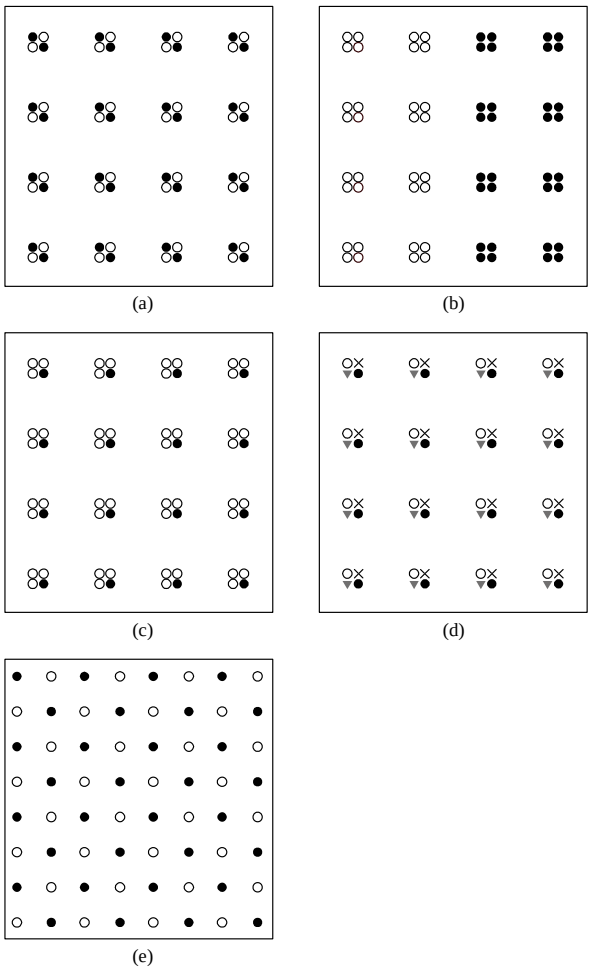



**Figure 1** Point distributions representing individuals of different ethnicities.

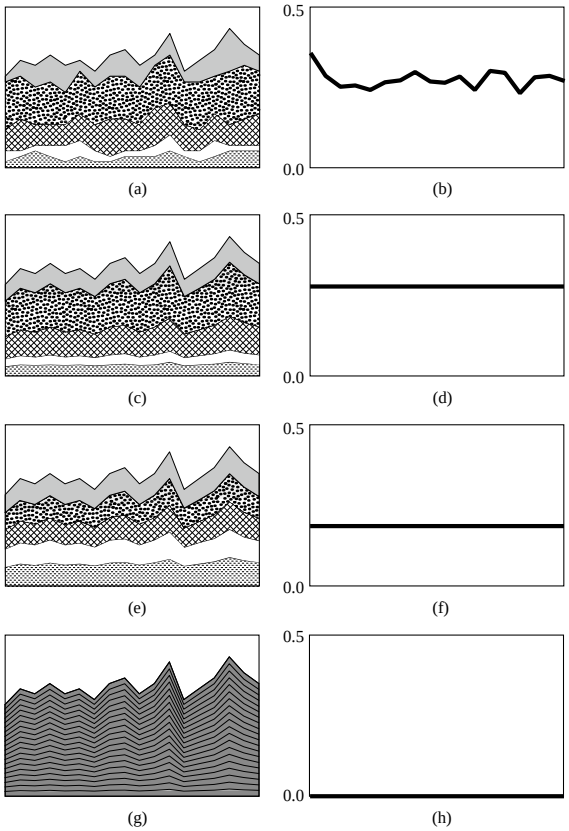
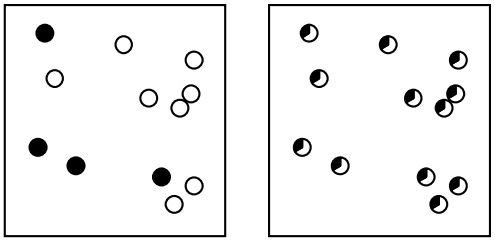



**Figure 2** Density functions of points assumed in the evaluation of segregation on a one-dimensional space and the distribution of local segregation measure. Density functions where (a) all the three sources of segregation are present, (c) the difference in the spatial arrangement of points is absent, (e) the difference in the spatial arrangement of points and the imbalance in the number of points are absent, (g) all the three sources of segregation are absent. Figures 2b, 2d, 2f, and 2h indicates the distribution of local segregation measure of density functions shown in Figure 2a, 2c, 2e, and 2g, respectively. Different textures indicate different types of points.



(a) (b)

**Figure 3** Evaluation of locational segregation. (a) Distribution of four black and eight white points. (b) A hypothetical situation where both black and white points have the same spatial arrangement.

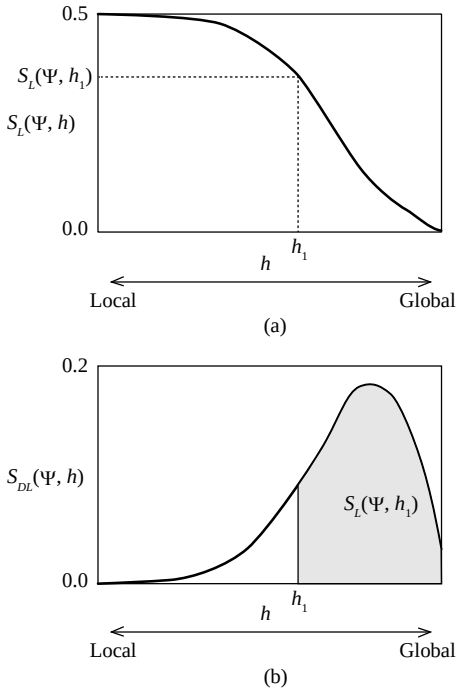



**Figure 4** Relationship between the locational segregation measure and the differential locational segregation measure. The horizontal axis indicates the window width *h* used in kernel smoothing.

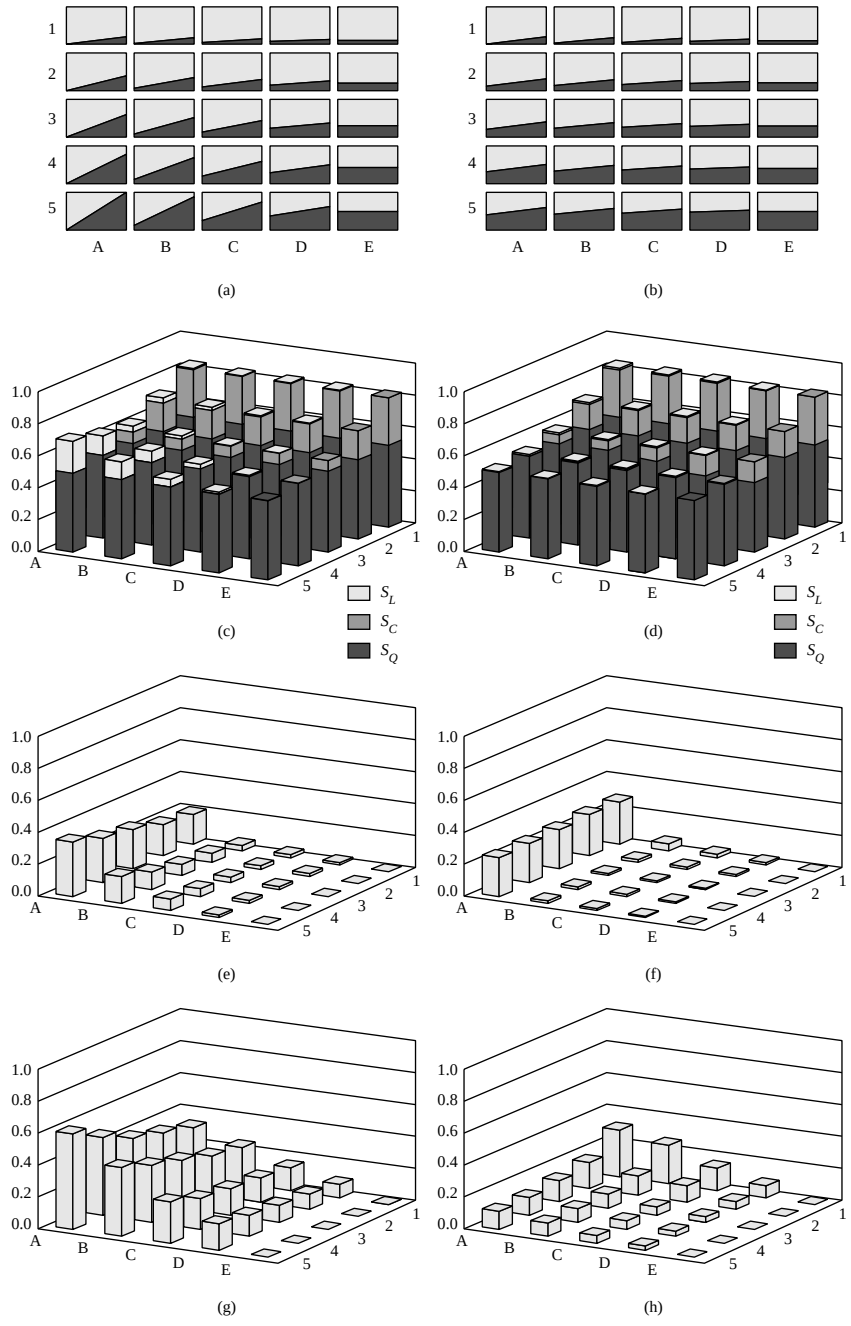



**Figure 5** Density distributions of points (a, b), their segregation measures (c, d), the spatial information theory segregation index (e, f), and the index of dissimilarity (g, h). Different colors indicate different types of points in Figures 5a and 5b.

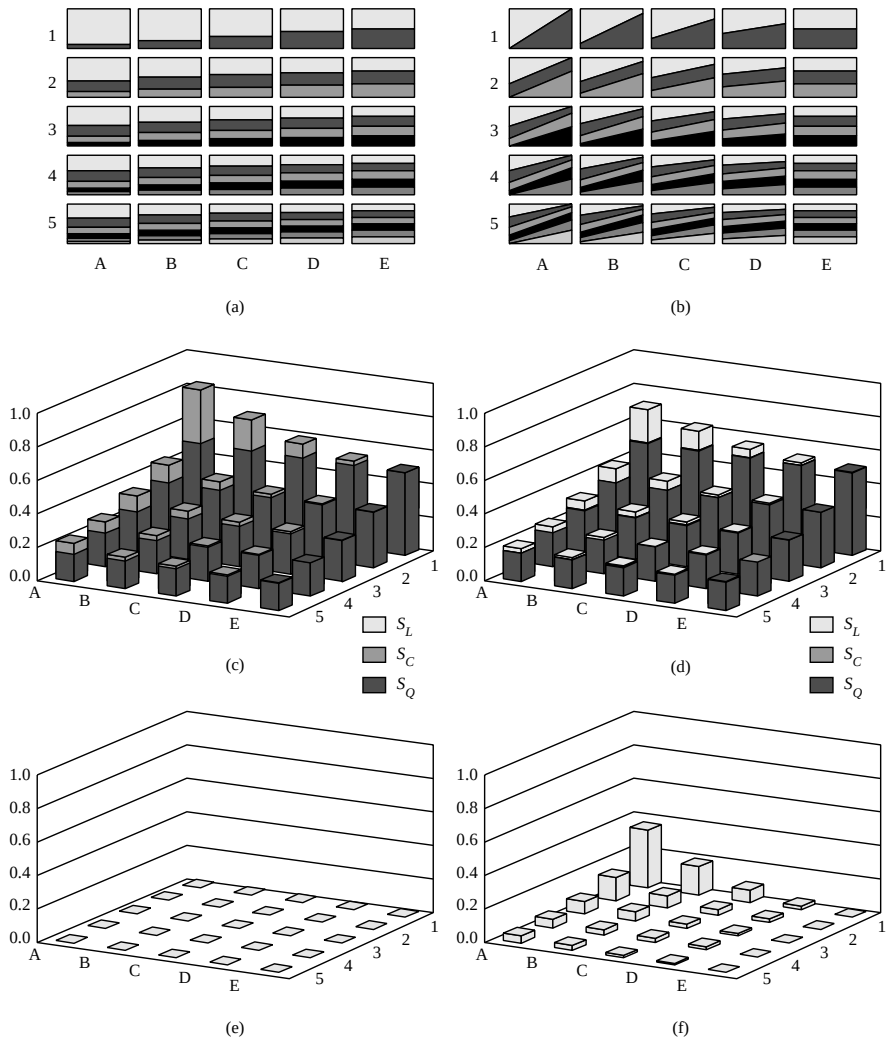



**Figure 6** Density distributions of points (a, b), their segregation measures (c, d), and the spatial information theory segregation index (e, f). Different colors indicate different types of points in Figures 6a and 6b.

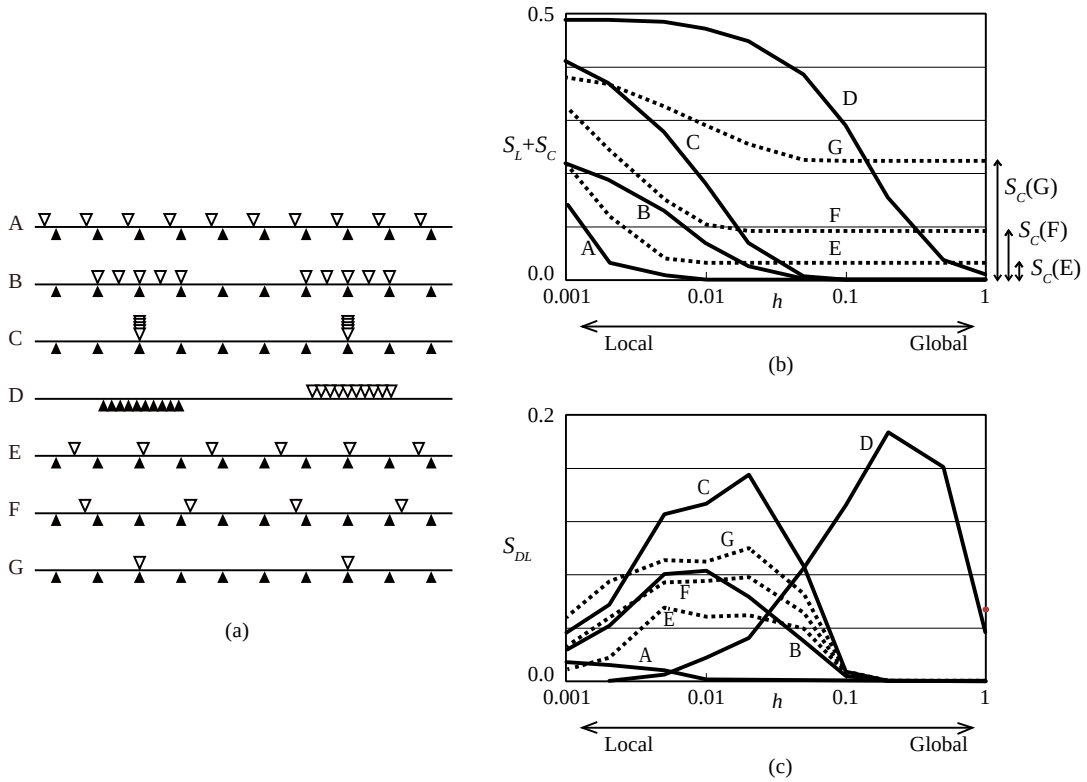
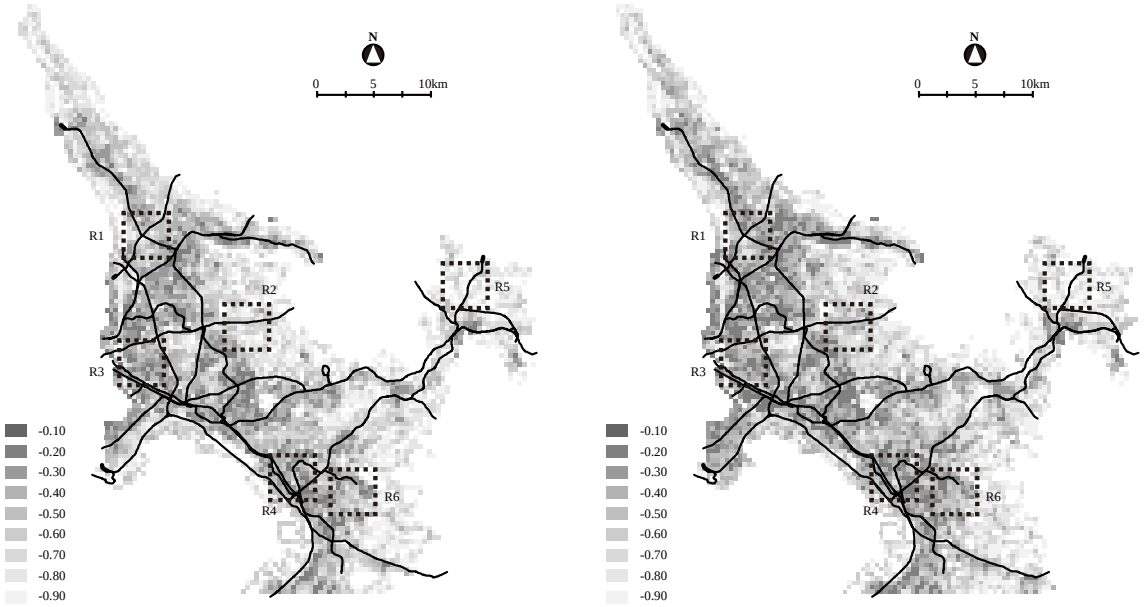



**Figure 7** Segregation measures of the distribution of two types of points on a one-dimensional space. (a) Point distributions, (b) the summation of the locational and compositional segregation measures, (c) the differential locational segregation measure. Bold solid lines indicate patterns from A to D while bold broken lines indicate patterns from E to G.




(a) (b)

**Figure 8** The local segregation measure $s(\mathbf{x})$ of land use pattern in (a) 1974 and (b) 1994.

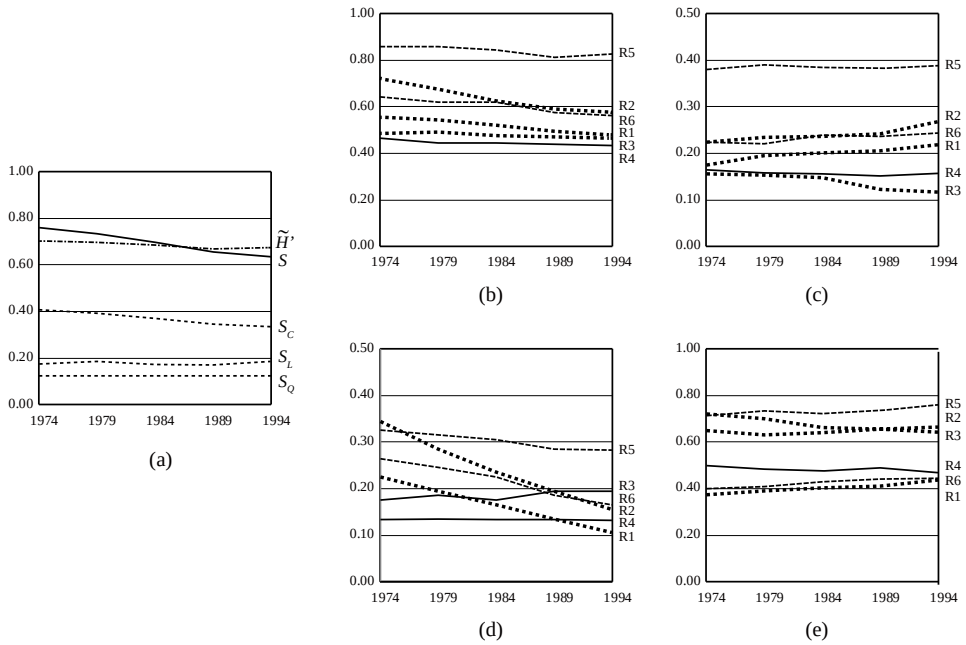



**Figure 9** The segregation measures of land use pattern from 1974 to 1994. (a) Segregation measures of the entire region, (b)-(e) Segregation measures of the subregions shown in Figure 8. (b) The overall segregation measure $S$, (c) the locational segregation measure $S_L$, (d) the compositional segregation measure $S_C$, (e) the spatial information theory segregation index.

**Appendix A1**

This appendix derives the minimum value of $s(\mathbf{x}, \Psi, h)$ defined by Equation(4) using the Langrangian multiplier method. Let us define $d_i(\mathbf{x}, h)$ as

$$d_i\left(\mathbf{x},h\right)=\frac{D_i\left(\mathbf{x},h\right)}{\sum_k D_k\left(\mathbf{x},h\right)}.$$

(31)

We minimize

$$f=\sum_i d_i^2\left(\mathbf{x},h\right)+\lambda\left(1-\sum_i d_i\left(\mathbf{x},h\right)\right)$$

(32)

with constrain

$$\sum_i d_i\left(\mathbf{x},h\right)=1,$$

(33)

where λ is the Langrangian multiplier. Partial differentiation of $f$ by $d_i(\mathbf{x}, h)$ and that by λ are

$$\frac{\partial f}{\partial d_i\left(\mathbf{x},h\right)}=2d_i\left(\mathbf{x},h\right)-\lambda$$

(34)

and

$$\frac{\partial f}{\partial \lambda}=1-\sum_i d_i\left(\mathbf{x},h\right),$$

(35)

respectively. Since

$$\frac{\partial f}{\partial d_i\left(\mathbf{x},h\right)}=0,$$

(36)

we obtain

$$d_i\left(\mathbf{x},h\right)=\frac{\lambda}{2}.$$

(37)

Substitution of Equation (37) into Equation (35) yields

$$1-\sum_i d_i\left(\mathbf{x},h\right)=1-\frac{\lambda K}{2}=0.$$

(38)

Solving this equation, we have

$$\lambda = \frac{2}{K}.$$

(39)

Substituting the above equation into Equation (38), we obtain

$$d_i(\mathbf{x}, h) = \frac{1}{K}.$$

(40)

This gives the minimum of $s(\mathbf{x}, \Psi, h)$, that is,

$$\begin{aligned} s(\mathbf{x}, \Psi, h) &= \sum_i \left\{ d_i(\mathbf{x}, h) \right\}^2 \\ &= \frac{1}{K} \end{aligned}.$$

(41)

**Appendix A2**

This appendix derives the minimum value of $S_L(\Psi, h)$. To this end, we divide region $R$ into $M$ subregions denoted by $\{R_1, R_2, ..., R_M\}$. Let $D_{ij}$ be the density of type $i$ points in $R_j$, which corresponds to $D_i(\mathbf{x}, h)$ in Equation (2). The proportion of type $i$ points in $R_j$ is given by

$$p_{ij} = \frac{D_{ij}}{\sum_i D_{ij}}.$$

(42)

The total density of points in $R_j$ is denoted by $\rho_i$. The total density of all the points is

$$T = \sum_j A(R_j) \rho_j,$$

(43)

where $A(R_j)$ is an operator that gives the area of $R_j$.

The locational segregation $S_L(\Psi, h)$ is defined as

$$\begin{aligned} S_L(\Psi, h) &= \frac{1}{T} \sum_j A(R_j) \rho_j \sum_i p_{ij}^2 - \sum_i \left( \frac{N_i}{N} \right)^2 \\ &= \frac{1}{T} \sum_i \left\{ \sum_j A(R_j) \rho_j p_{ij}^2 - \left( \frac{N_i}{N} \right)^2 T \right\} \end{aligned}.$$

(44)

We use the Langrangian multiplier method, where we minimize

$$f_i = \sum_j A\left(R_j\right)\rho_j p_{ij}^2 - \left(\frac{N_i}{N}\right)^2 T + \lambda\left(\frac{N_i}{N}T - \sum_j p_{ij}A\left(R_j\right)\rho_j\right).$$

(45)

The constraints are

$$\forall i, \sum_j p_{ij}A\left(R_j\right)\rho_j = \frac{N_i}{N}T$$

(46)

Partial differentiation of $f_i$ by $p_{ij}$ and that by $\lambda$ are

$$\frac{\partial f_i}{\partial p_{ij}} = 2A\left(R_j\right)\rho_j p_{ij} - \lambda A\left(R_j\right)\rho_j .$$

(47)

and

$$\frac{\partial f_i}{\partial \lambda} = \frac{N_i}{N}T - \sum_j p_{ij}A\left(R_j\right)\rho_j ,$$

(48)

respectively. Solving

$$\frac{\partial f_i}{\partial p_{ij}} = 0,$$

(49)

we obtain

$$p_{ij} = \frac{\lambda}{2}.$$

(50)

This equation indicates that $p_{ij}$ does not depend on $i$. We thus obtain

$$p_{ij} = \frac{N_i}{N}.$$

(51)

and the minimum of $S_L(\Psi, h)$:

$$S_L\left(\Psi, h\right) = \frac{1}{T}\sum_i\left\{\sum_j A\left(R_j\right)\rho_j\left(\frac{N_i}{N}\right)^2 - \left(\frac{N_i}{N}\right)^2 T\right\}.$$
$$= 0$$

(52)

**Appendix A3**

This appendix proves that the qualitative segregation is independent of the locational and compositional segregations. To this end, we add *m* types of points to the point set considered in Section 2. We keep the total number of points but can modify the number of each type of points. Let $N_i$' be the number of type *i* points after the addition of new points.

We first consider the independency of the qualitative segregation from the locational segregation. The locational segregation remains unchanged when the following equation holds for any **x**:

$$\sum_i \left\{ \frac{D_i(\mathbf{x},h)}{\sum_k D_k(\mathbf{x},h)} \right\}^2 = \sum_i \left\{ \frac{D_i'(\mathbf{x},h)}{\sum_k D_k'(\mathbf{x},h)} \right\}^2 . \tag{53}$$

This condition is equivalent to

$$\sum_i p_i^2 = \sum_i p_i'^2 , \tag{54}$$

under the constraint:

$$\sum_i p_i = \sum_i p_i' = 1. \tag{55}$$

We assume $m$=1 and $p_i$=$p_i$' for $i$=1, ..., $K$-2. Equations (54) and (55) become

$$p_{K-1}^2 + p_K^2 = p'^2_{K-1} + p'^2_K + p'^2_{K+1}, \tag{56}$$

and

$$p_{K-1} + p_K - p'_K - p'_{K+1} = p'_{K-1} . \tag{57}$$

We can rewrite $p_{K-1}$ as

$$p_{K-1} = p_K + a \tag{58}$$

without losing generality. Substituting Equations (57) and (58) into Equation (56), we solve it in terms of $p_K$:

$$p_K = \frac{-(p'_K + p'_{K+1}) + 2\sqrt{p'_K\, p'_{K+1} + ap'_K + ap'_{K+1}}}{2} \tag{59}$$

The condition of the existence of a positive $p_K$ is

$$4a\left(p'_K + p'_{K+1}\right) > \left(p'_K - p'_{K+1}\right)^2 .$$

(60)

It is always possible to satisfy the above inequality by choosing close values of $p'_K$ and $p'_{K+1}$. Consequently, we can change the locational segregation with keeping the qualitative segregation.

We then consider the independency of the qualitative segregation from the compositional segregation. The compositional segregation is unchanged when

$$\sum_i \left(\frac{N_i}{N}\right)^2 - \frac{1}{K^2} = \sum_i \left(\frac{N_i'}{N}\right)^2 - \frac{1}{\left(K+m\right)^2} .$$

(61)

Solving Equation (61) in terms of $m$, we obtain

$$m = \sqrt{\frac{K^2}{K^2 \sum_i \left(\frac{N_i'}{N}\right)^2 - K^2 \sum_i \left(\frac{N_i}{N}\right)^2 + 1}} - K .$$

(62)

Since $m$ is positive, the following inequality needs to hold:

$$\sum_i \left(\frac{N_i}{N}\right)^2 - \frac{1}{K^2} < \sum_i \left(\frac{N_i'}{N}\right)^2 < \sum_i \left(\frac{N_i}{N}\right)^2 .$$

(63)

It is always possible to satisfy the above inequality by slightly reducing every $N_i$ and add a new type of points. Consequently, we can change the qualitative segregation with keeping the compositional segregation.

**Appendix A4**

This appendix derives the minimum value of the locational segregation $S_{Le}(\boldsymbol{D}, \Psi, h)$. We use the same setting as that of Appendix A2. The locational segregation $S_{Le}(\boldsymbol{D}, \Psi, h)$ is written as

$$\begin{aligned} S_{Le}\left(\boldsymbol{D}, \Psi, h\right) &= \frac{1}{T}\sum_j A\left(R_j\right)\rho_j \sum_i p_{ij} \log p_{ij} - \sum_i p_{ij} \log p_{ij} \\ &= \frac{1}{T}\sum_i \left\{ \sum_j A\left(R_j\right)\rho_j p_{ij} \log p_{ij} - \frac{N_i}{N} T \log \frac{N_i}{N} \right\} . \end{aligned}$$

(64)

We use the Langrangian multiplier method to calculate the minimum value of $S_{Le}(\boldsymbol{D}, \Psi, h)$. We minimize

$$f_i = \sum_j A\left(R_j\right)\rho_j p_{ij} \log p_{ij} - \frac{N_i}{N} T \log \frac{N_i}{N} + \lambda\left(\frac{N_i}{N} T - \sum_j p_{ij} A\left(R_j\right)\rho_j\right).$$

(65)

with constraints

$$\forall i, \sum_j p_{ij} A\left(R_j\right)\rho_j = \frac{N_i}{N} T\,.$$

(66)

Partial differentiation of $f_i$ by $p_{ij}$ and that by $\lambda$ are

$$\frac{\partial f_i}{\partial p_{ij}} = \rho_j A\left(R_j\right)\left(\log p_{ij} - 1 - \lambda\right).$$

(67)

and

$$\frac{\partial f_i}{\partial \lambda} = \frac{N_i}{N} A\left(R\right) - \sum_j p_{ij} A\left(R_j\right),$$

respectively. Solving

$$\frac{\partial f_i}{\partial p_{ij}} = 0\,,$$

(68)

we obtain

$$p_{ij} = K^{1+\lambda}\,.$$

(69)

This implies that $p_{ij}$ is independent of $i$. From this we derive

$$p_{ij} = \frac{N_i}{N}\,.$$

(70)

Substitution of Equation (70) into (64) yields

$$\begin{aligned} S_{Le}\left(\boldsymbol{D}, \Psi, h\right) &= \frac{1}{T}\sum_i \left\{\sum_j A\left(R_j\right)\rho_j \frac{N_i}{N} \log \frac{N_i}{N} - \frac{N_i}{N} T \log \frac{N_i}{N}\right\}. \\ &= 0 \end{aligned}$$

(71)