J Geogr Syst (2014) 16:211–231
DOI 10.1007/s10109-013-0190-7

ORIGINAL PAPER

Measuring geographic segregation: a graph-based approach

Seong-Yun Hong · Yukio Sadahiro

Received: 20 February 2013 / Accepted: 26 July 2013 / Published online: 9 August 2013

Abstract Residential segregation is a multidimensional phenomenon that encompasses several conceptually distinct aspects of geographical separation between populations. While various indices have been developed as a response to different definitions of segregation, the reliance on such single-figure indices could oversimplify the complex, multidimensional phenomena. In this regard, this paper suggests an alternative graph-based approach that provides more detailed information than simple indices: The concentration profile graphically conveys information about how evenly a population group is distributed over the study region, and the spatial proximity profile depicts the degree of clustering across different threshold levels. These graphs can also be summarized into single numbers for comparative purposes, but the interpretation can be more accurate by inspecting the additional information. To demonstrate the use of these methods, the residential patterns of three major ethnic groups in Auckland, namely Māori, Pacific peoples, and Asians, are examined using the 2006 census data.

Keywords Segregation measures · Residential segregation · Segregation profiles · Concentration profile · Spatial proximity profile

JEL Classification C4 · C43 · J15

1 Introduction

Residential segregation is a multidimensional phenomenon that encompasses several conceptually distinct aspects of geographical separation between

S.-Y. Hong (✉) · Y. Sadahiro
Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan
e-mail: yun.hong@csis.u-tokyo.ac.jp

populations (Massey and Denton 1988; Reardon and O'Sullivan 2004). In order to capture the different dimensions of segregation, various indices have been proposed, and the use of reliable and appropriate measurement is crucial in studying the geographic patterns of population distributions.

The index of dissimilarity has been the most widely used measure of segregation in the literature due to its easy calculation and interpretation (White 1983; Morrill 1991; Wong et al. 2007). One limitation of this classical measure is, however, that it does not take the absolute population sizes into account (Allen et al. 2009; Johnston et al. 2010). If, for example, the index of dissimilarity is calculated for two groups, one of which has a very small population size and the other is relatively large, the overall segregation level can be over- or underestimated by random fluctuations in the small group (Winship 1977; Falk et al. 1978). Furthermore, as will be illustrated in the next section, the index of dissimilarity alone cannot distinguish various "uneven" settings in the population composition, and it is unable to capture the spatial dimension of segregation—clustering (White 1983; Morrill 1991; Wong 1993; Reardon and O'Sullivan 2004; Wong et al. 2007).

While recently developed spatial segregation indices address the last problem, these single-figure indices are not always informative enough to describe the actual state of segregation under examination (Johnston et al. 2010). In this context, this study introduces an alternative graph-based approach to measuring segregation. The concentration profile visually represents how evenly, in proportion to the population size of each census tract, an individual group is distributed over the study region (Poulsen et al. 2002), and the spatial proximity profile depicts the level of clustering across different threshold levels. These graphs can also be used to derive summary measures of segregation, but the interpretation can be more accurate by inspecting the additional information portrayed in the profile curves.

In the next sections, we review existing measures of segregation and their weaknesses in more detail and then describe the definitions of alternative graph-based methods. The subsequent section demonstrates the use of the proposed approach by examining two idealized landscapes as well as the residential patterns of three major ethnic groups in Auckland, Māori, Pacific peoples, and Asians, in the 2006 census. In the final section, we conclude with a discussion of the strengths of the graph-based methods as well as limitations that may be overcome in future research.

2 Existing measures

As mentioned in the beginning of this paper, the index of dissimilarity has been the dominant method for measuring residential segregation over the last half a century. The index of dissimilarity, or D, is a numerical summary statistic of segregation based on the proportions of two mutually exclusive population groups in census tracts (Duncan and Duncan 1955). The groups are traditionally referred to as non-whites and whites and accordingly denoted by N and W. For a study region consisting of k census tracts, the index is defined as:

$$D = \frac{1}{2}\sum_{i=1}^{k}\left|\frac{N_i}{N} - \frac{W_i}{W}\right|$$

where N_i and W_i denote the population counts of non-whites and whites, respectively, in the ith census tract, and N and W are the total population counts of the two groups. The value of D ranges from 0 to 1, indicating the proportion of the non-white population, N, that should be moved to make each census tract have the same population composition (White 1983).

Although the index of dissimilarity is useful in comparing the distribution of two population groups across spatial units, this single-figure index can be easily misinterpreted. Figure 1, for example, illustrates three hypothetical data sets in hexagonal grids, each of which has a different pattern of segregation. The first data set exhibits the most uneven distribution of populations: 70 % of the non-white population is concentrated in a single tile, and the rest is scattered over three adjacent units. In the second data set, there is a couple of areas where 40 % of the minority individuals are located, resulting in the modest level of segregation, whereas in the last data set, the population is equally distributed across four hexagonal tiles. Despite these differences, the index of dissimilarity is the same for all the synthetic landscapes, $D = 0.667$, meaning that the reliance on this simple measure can lead to a misunderstanding of the actual situation.

Furthermore, it has been repeatedly pointed out that this simple measure does not take the spatial structure of segregation into account—the checkerboard problem (White 1983; Morrill 1991; Wong 1993; Reardon and O'Sullivan 2004; O'Sullivan and Wong 2007; Wong et al. 2007). In recent years, efforts have been directed toward developing new measures that incorporate the spatial structure of segregation into the calculation: White (1983) and Morgan (1983), for example, suggested the use of distances between the centroid of census tracts when computing their indices. In a similar vein, Morrill (1991) proposed the boundary-modified index of dissimilarity, DM, which incorporates a binary contiguity matrix, **C**, into the traditional dissimilarity index to adjust for the spatial structure:

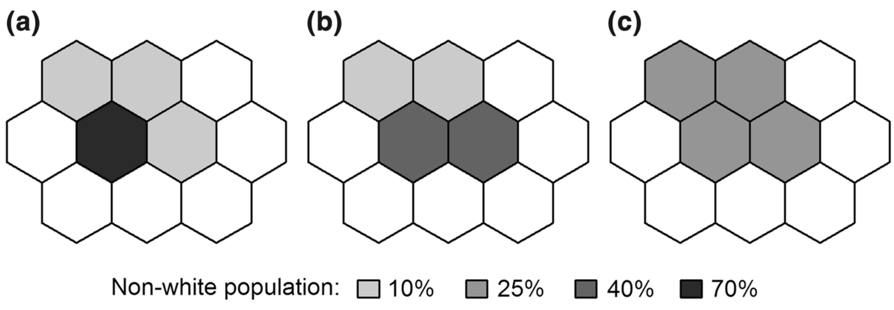

Fig. 1 Three hypothetical spatial configurations in 10 hexagonal grids

$$\mathrm{DM} = D - \frac{\sum_i \sum_j \left| c_{ij} \left(p_i - p_j \right) \right|}{\sum_i \sum_j c_{ij}}$$

where p_i and p_j represent the proportions of a minority group in census tracts i and j, and c_{ij} equals 1 if i and j are adjacent, and 0 otherwise. If the population proportions in two neighboring tracts are substantially different, it increases the second term in the equation, and so the overall level of segregation decreases.

Although the Morrill's DM and other similar adjustments, such as Wong (1993), diminish the aspatial characteristic of the original D index to some extent, they cannot fully describe the geographic distribution of populations, especially when the population size is small. Because the additional components in these methods do depend not only on the spatial arrangement of census tracts as intended but also on the proportions of the minority population, p_i, small variations in p_i can eliminate them to a large degree.

There is another important issue related to the use of these indices (and the traditional, aspatial one as well) when the minority population is small in size. That is, since this measurement works on the data in which the population counts are agglomerated into arbitrarily defined geographic areas, such as census tracts, electorates, and school zones, the resulting degree of segregation depends not only on the actual distribution of the population but also on the choice of geographic units (Openshaw 1984). The impact of this so-called modifiable areal unit problem (MAUP) is particularly large when the population of interest constitutes only a small fraction of the total population compared to the number of geographic units, as local proportions are likely to be unstable in that case (Wong 1997; Allen et al. 2009).

The spatial information theory index, $\tilde{H}$, developed by Reardon and O'Sullivan (2004) avoids the MAUP issue, as it is theoretically independent of geographic units. Unlike the index of dissimilarity and its spatial associates outlined above, this multigroup measure does not require the use of aggregate geographic units and is defined as follows:

$$\tilde{H} = 1 - \frac{1}{TE} \int\limits_{i \in R} \tau_i \left(- \sum_{m=1}^{M} p_{mi} \log_M p_{mi} \right) \mathrm{d}i$$

where T is the total population in the study region R, M is the number of population groups, τ_i is the population density in the local environment of point i, and p_{mi} is the proportion of the subgroup m in the same local environment. E denotes the overall regional entropy, which is given by:

$$\left(E = - \sum_{m=1}^{M} p_m \log_M p_m \right)$$

where p_m is the proportion of the subgroup m in the regional population composition.

While the spatial information theory index is a useful addition that complements the aspatial characteristic of D, it is important to note that the resulting segregation

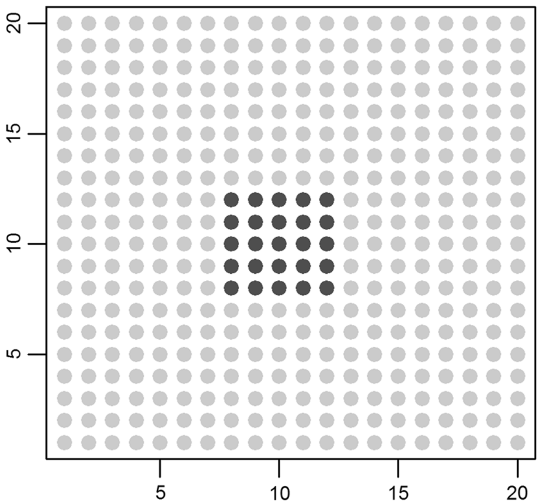

Fig. 2 Idealized distribution of individuals: *dark-gray points* are clustered in the core of the 20 m-by-20 m plot

from $\tilde{H}$ is sensitive to the definition of the local environment. In practice, a circle with a specific radius is often used to delineate the local environment for each measuring point, and depending on the chosen circle size, $\tilde{H}$ can be completely changed for the same segregation phenomenon. Figure 2, for instance, shows that dark-gray points are clustered in the central part of the plot: $\tilde{H}$ is 0.800 when the circle of radius 1 m is used as the local environment, indicating the complete segregation, but it becomes −0.184 if the radius is increased to 9 m. Although this is an important feature that enables inspecting the scale of segregation (Reardon et al. 2008, 2009), it also implies that the reliance on a single index value can result in a misinterpretation of the actual situation under examination.

To reduce the risk of oversimplifying (and even distorting sometimes) this multidimensional, complex social phenomenon, we suggest a graph-based approach that provides more detailed information than simple indices: The concentration profile graphically conveys information about how evenly a population group is distributed over the study region (Poulsen et al. 2002), and the spatial proximity profile depicts the degree of clustering across different threshold levels. These graphs can also be summarized into single numbers for comparative purposes, but the interpretation can be more accurate by inspecting the additional information. In the next two sections, we will demonstrate the advantages of the proposed approach over the existing single-figure indices.

3 Alternative approach

3.1 Concentration profile

The concentration profile is a recent addition to the segregation literature, which was, to the best of our knowledge, first introduced by Poulsen et al. (2002). It is a useful tool to inspect the evenness aspect of segregation, but it has not been widely used in part because its strength has not been clearly stated. In this section, we attempt to provide a more formal definition of the concentration profile and then demonstrate its advantages.

The concentration profile is a graphical expression showing the proportions of an individual ethnic group's population at some predefined threshold levels (Poulsen et al. 2002). At each threshold interval, t, it calculates how many of the group members are concentrated in the areas where they comprise at least the chosen "threshold" proportion in the local demographic composition:

$$\upsilon_t = \frac{\sum_{i=1}^{k} \omega_i g(t, i)}{\sum_{i=1}^{k} \omega_i}$$

where k denotes the number of census tracts, and ω_i the number of the minority population in the census tract i. In the equation, $g(t, i)$ is a logical function that is defined as:

$$g(t, i) = \begin{cases} 1 & \text{if } \frac{\omega_i}{\Omega_i} \geq t, \\ 0 & \text{otherwise} \end{cases}$$

where Ω_i is the total number of the population in the census tract i. If the number of thresholds used is large enough, a smooth curve, or *a concentration profile*, that depicts the group's clustering level can be obtained by plotting and connecting the resulting values of υ_t.

To help the illustration of this method, three idealized concentration profiles were constructed (Fig. 3). The dotted line represents a group that is evenly distributed across the study region: the entire group members (i.e., y axis) are located in the areas where they make up 40 % or more of the local populations (i.e., x axis). The dashed line, on the other hand, indicates that the minority population is not residentially mixed with other ethnic groups—complete segregation. The solid line displays a randomly distributed group, which comprises 40 % of the total population in the study region.

The concept underlying this graphical representation of segregation is similar to that of the traditional Lorenz curve in a sense, as the concentration profile also displays the (reverse) cumulative distribution of a population group[1]. Therefore, as we can derive the Gini coefficients from the Lorenz curve (Dorfman 1979), it is also possible to summarize a given concentration profile into a single-figure index for comparative purposes.

To derive a numerical summary statistic, one needs to calculate the area between the actual concentration profile and the line that represents a uniform distribution (i.e., shaded area in Fig. 3), and then divide it by the area above the line of no segregation. This normalized area, R, can be written in mathematical terms as follows:

[1] In the segregation literature, the Lorenz curve is often constructed by plotting the cumulative proportion of one population group against that of the other group. The concentration profile is different from the Lorenz curve in the sense that it plots the cumulative proportion of the population group against their relative demographic share in geographic units.

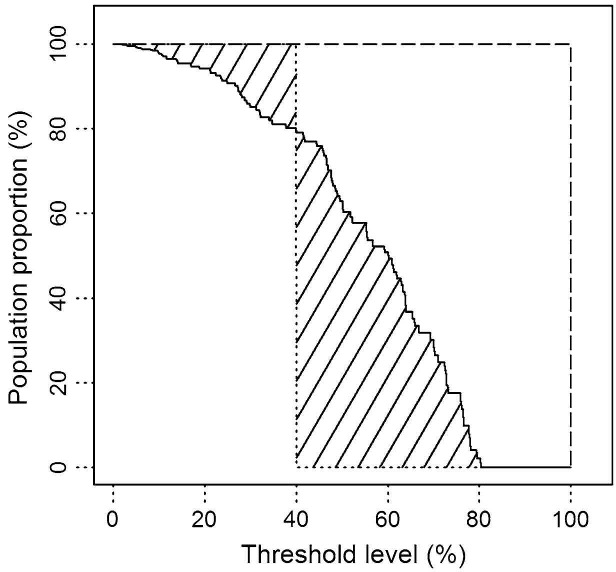

Fig. 3 Concentration profiles representing no segregation (*dotted line*), complete segregation (*dashed line*), and random distribution (*solid line*)

$$R = \frac{p - \left(\int_{t=0}^{p} \upsilon_t \mathrm{d}t - \int_{t=p}^{1} \upsilon_t \mathrm{d}t\right)}{1 - p}$$

where p is the proportion of the minority group in the regional population composition. Since the denominator is always greater than or equal to the numerator, the value of R ranges between 0 and 1 and can be interpreted in a similar manner to the index of dissimilarity: A small value indicates that the group comprises similar proportions of the local population in all census tracts, and a large value implies a high degree of residential concentration.

Figure 4 displays that the normalized areas, R, are strongly correlated with the index of dissimilarity, D, in general. It shows the levels of segregation for 100 randomly created data sets, where each group has the population size of about 5,000, and the scatter plots exhibit a strong linear relationship between R and D ($\rho = 0.861$). This suggests that the proposed method is comparable to the widely used measure of segregation, D, to a certain extent, while it offers more details than the simple single-figure indices through the concentration profile.

The main difference between the concentration profile and the Lorenz curve (or the segregation curve in Hutchens 1991), and therefore between R and D, is that the former is dependent only on the distribution of the subject population. This attribute makes R potentially better for capturing the variations within an individual group. For example, while the index of dissimilarity gave the same value for the three different patterns in Fig. 1 ($D = 0.667$), R can distinguish them, providing the highest value for Fig. 1a ($R = 0.467$) and the lowest for Fig. 1c ($R = 0.167$).

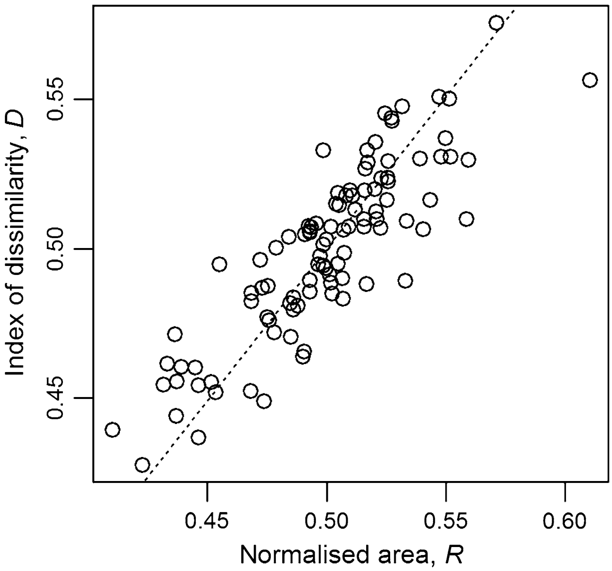

Fig. 4 Positive linear relationships between the normalized area, R, and the index of dissimilarity, D

3.2 Spatial proximity profile

As described in the previous section, the concentration profile evaluates the degree to which a population group is unevenly, in proportional terms, spread across census tracts in a similar manner to the index of dissimilarity. The spatial proximity profile complements its aspatial characteristic and addresses another dimension of segregation—clustering.

Let x_{ti} denote the ith census tract where the population of interest comprises more than or equal to a given "threshold" proportion, t, in the local demographic composition. Then for each threshold value, we can have a group of census tracts, $X_t = \{x_{t1}, x_{t2}, \ldots, x_{tk}\}$, and the distance between one geographic unit, x_{ti}, and another, x_{tj}, for all possible pairs in X_t can be calculated. If this distance adequately represents the degree of social and physical interactions between individuals living in the corresponding census tracts, a measure of clustering can be defined as:

$$\eta_t = \frac{k^2 - k}{\sum_i \sum_j \delta_{ij}}$$

where k refers to the number of census tracts in X_t and δ_{ij} is the distance between i and j.

The distance between the geographic units can be measured in a variety of ways. One way of determining δ_{ij} would be to use a spatial structure matrix, $\mathbf{W}$, whose elements describe the spatial relationship between all pairs of X. If $\mathbf{W}$ is a binary contiguity matrix, such that each value in the matrix is:

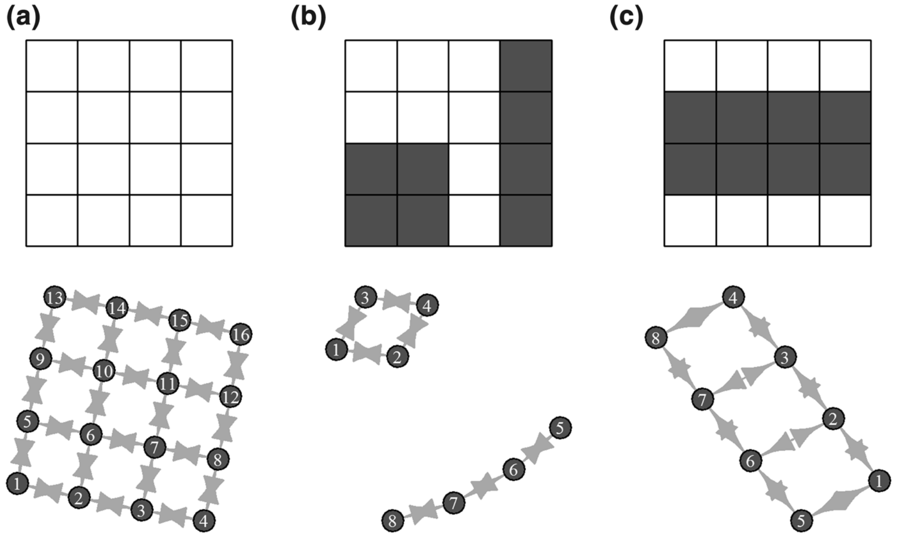

Fig. 5 4-by-4 grids and connectivity graphs: **a** The graph, G, describes the spatial relationship between all cells in the grid; **b** the *black-colored* cells in the grid form two separate clusters, and their connectivity graph, G_1, also consists of two subgraphs; **c** the *black-colored* cells are completely clustered together, and their connectivity graph, G_2, is fully connected

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are adjacent,} \\ 0 & \text{otherwise} \end{cases}$$

and if none of the census tracts are completely separated from the rest, a connected graph, G, can be constructed by linking adjacent pairs (Fig. 5a). A subgraph of G, G_t, contains only those in X_t, and this may and may not be connected, as shown in Fig. 5b, c. If two census tracts, x_{ti} and x_{tj}, are somehow connected in G_t, then δ_{ij} can be set to 1, indicating that the two are clustered; otherwise, the order of the neighborhood is obtained from $\mathbf{W}$, and it can be used as δ_{ij}. For example, two census tracts, x_1 and x_2, that do not have a common boundary but both are adjacent to the same unit, x_3, are second-order neighbors, so δ_{12} becomes 2.

In this case, η_t equals 1 only when all the census tracts in X_t are completely clustered (i.e., $\delta_{ij} = 1$ for all combinations of i and j), and it moves toward 0 as the census tracts become separated and further apart from each other. Similar to the concentration profile, if the number of thresholds used is large enough, a smooth curve, or *a spatial proximity profile*, can be constructed by plotting and connecting η_t.

To demonstrate the use of η_t and the spatial proximity profile, it is applied to two idealized patterns in Fig. 6. The two patterns have the same population composition ($R = 0.304$) but show different spatial arrangements. In Fig. 6a, the areas with high proportions of the minority population (i.e., black-colored cells) are regularly spread across the grid, whereas in Fig. 6b, they are agglomerated around the core. The degree of clustering is clearly higher in the latter case, and the η_t values reflect this fact well: $\eta_{0.75}$ is 0.126 for Fig. 6a and is 1 for Fig. 6b. A Monte Carlo simulation

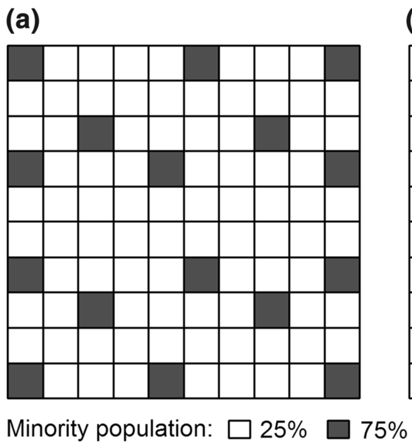

Fig. 6 Two idealized spatial configurations in 10-by-10 grids

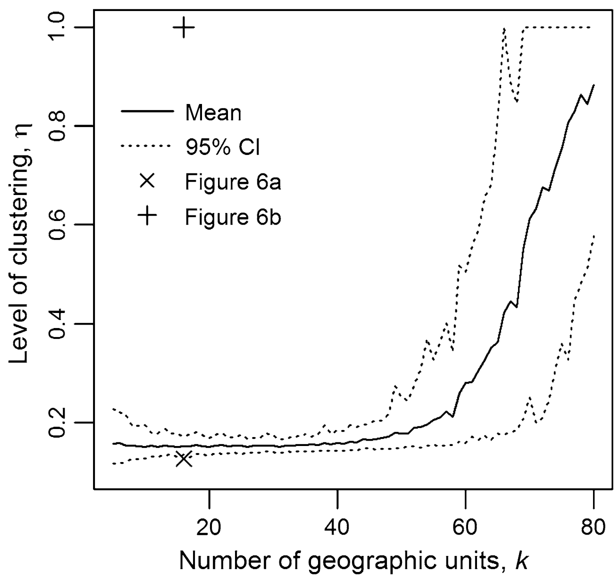

Fig. 7 Mean η_t values from 100 randomly generated patterns in 10-by-10 grids for $k = \{5, 6, \ldots, 80\}$ and its 95 % confidence envelope

with 100 iterations suggests that the $\eta_{0.75}$ value of 1 is very unlikely to be produced by random chance. Figure 7 shows that given the size of $X_{0.75}$ (i.e., $k = 16$), the 95 % confidence interval for a random distribution is 0.135 and 0.172. The observed value is much greater than the expected under the random assumption, indicating that the level of clustering in Fig. 6b is statistically significant.

4 Examples

4.1 Synthetic data

In this section, we apply the two methods described above to two simulated landscapes in Fig. 8. Each simulated data set contains three population groups, A, B, and C, and they are distributed on an idealized 10 km-by-10 km grid of 1-km-square cells. In each landscape, group A comprises 10 % of the total population, and groups B and C constitute 45 % each. Figure 8a shows that group A is highly clustered in the center of the grid, whereas B and C do not exhibit apparent spatial concentrations. Group A is fairly evenly distributed in Fig. 8b, and the centralization of the population disappears. Group B, however, forms a larger cluster in the bottom-right corner of the grid, and consequently, group C is underrepresented in those areas.

The proposed segregation curves, the concentration profiles in Fig. 9 and the spatial proximity profiles in Fig. 10, seem to capture the patterns of each group accurately. Figure 9a depicts that over 60 % of group A is clustered in the cells where they comprise about 40 % of the cell population in Fig. 8a, indicating an extreme level of unevenness, while it suggests that the group is quite uniformly distributed in Fig. 8b (i.e., the concentration profile is similar to the dotted line of no

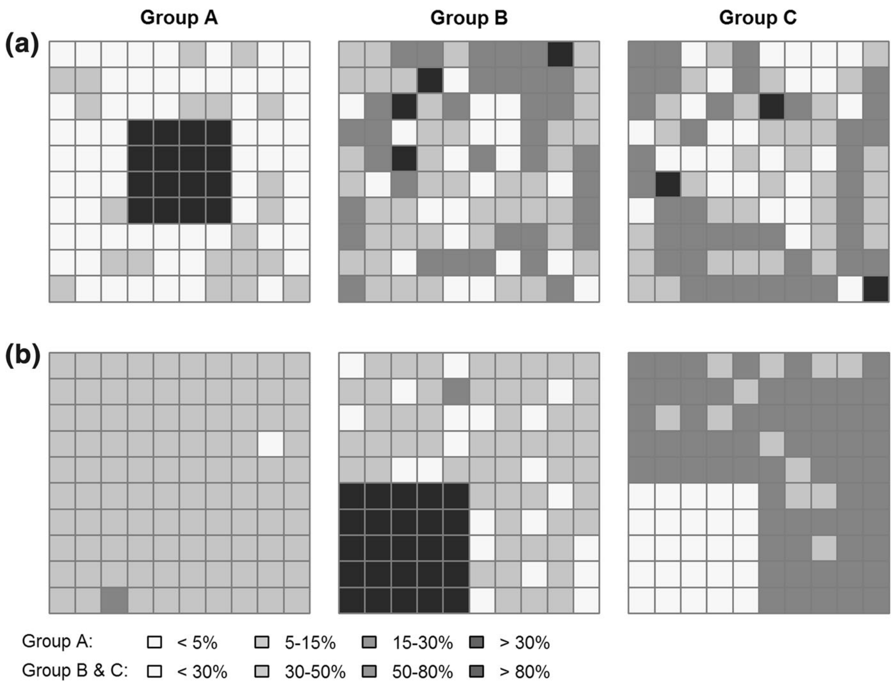

Fig. 8 Simulated landscapes: **a** Group A is clustered in the center of the 10 km-by-10 km grid, whereas groups B and C do not form apparent clusters; **b** group B is overrepresented in the *bottom-right corner* of the grid. Note that group A comprises 10 % of the total population, and groups B and C account for 45 % each

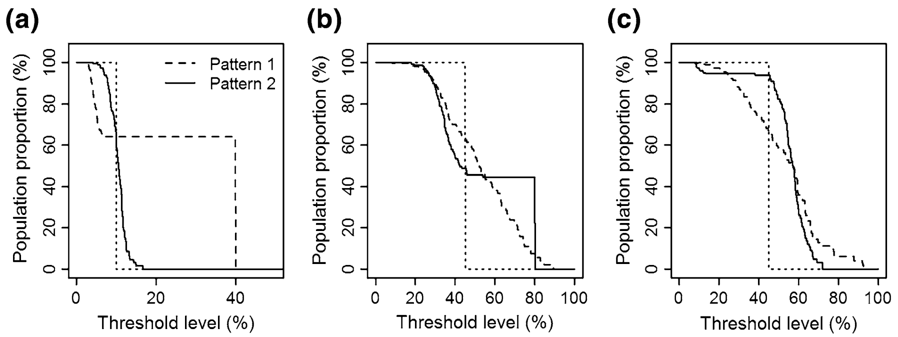

Fig. 9 Concentration profiles for the simulated landscapes: **a** group A in Fig. 8a (*dashed line*) and in Fig. 8b (*solid line*), **b** group B, and **c** the group C. The *dotted line* represents no segregation

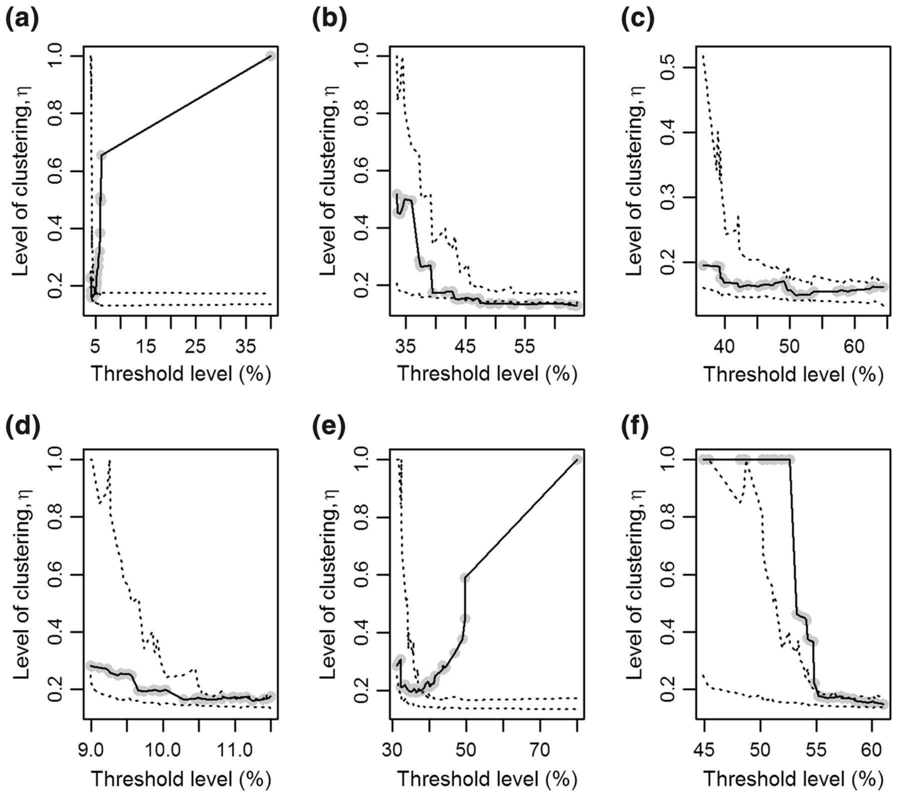

Fig. 10 Spatial proximity profiles for the simulated landscapes: **a**–**c** groups A, B, and C in Fig. 8a, respectively; **d**–**e** groups A, B, and C in Fig. 8b. The *dotted lines* indicate the 95 % confidence intervals

segregation). The spatial proximity profiles (Fig. 10a, d) reveal that the cells where A is relatively overrepresented are much more closely located than the expected in Fig. 8a, but such geographic clustering disappears in Fig. 8b.

Table 1 Spatial information theory index ($\tilde{H}$) and the index of dissimilarity between A and the other groups (D_A), B and the other groups (D_B), and C and the other groups (D_C) for the simulated landscapes

	$\tilde{H}$	D_A	D_B	D_C
Pattern 1	0.074	0.533	0.317	0.328
Pattern 2	0.090	0.091	0.358	0.361

Figure 9b, c indicates that groups B and C are rather unevenly distributed in both simulated landscapes, but the degree slightly increases in the second pattern. More significant changes occur in the level of clustering: Fig. 10b, c shows that the segregation profiles are closer to, or lower than, the lower bound of the 95 % confidence interval, and this suggests that the two groups are more regularly scattered than the expected under the null hypothesis of random arrangements in Fig. 8a. Figure 10e, on the other hand, displays that the observed η_t values are much greater than the expected for $t > 0.50$, indicating the deliberate development of clustering in Fig. 8b. It is noteworthy that, unlike Fig. 10b–d, the spatial proximity profile in Fig. 10f is closer to, or even higher than, the upper bound of the 95 % confidence interval, and this is due to the underrepresentation of group C around the bottom-right corner (and the consequential, relative overrepresentation in the rest of the grid).

Table 1 shows the spatial information theory index ($\tilde{H}$) and the index of dissimilarity for Fig. 8. Despite the presence of the obvious concentration of A in Fig. 8a, the spatial information theory index is larger for Fig. 8b, due to the emergence of the bottom-right cluster (0.074 and 0.090, respectively). This figure may be an accurate description for the overall pattern, as it reflects the changes in the larger group. However, it rendered the residential experience of the small group invisible, and this property is undesirable if the interest lies in the segregation of that group. Of course, the use of additional indices, such as D in Table 1, or the pairwise calculation of $\tilde{H}$ for each possible combination (or for each group against the rest) could allow us to capture the specific aspect of the phenomenon, but nonetheless, it would be difficult to obtain the detailed descriptions drawn from the inspection of the segregation curves.

4.2 Census data

Now we use the segregation profile curves to measure the levels of segregation for three major ethnic groups, Māori, Pacific peoples, and Asians, in Auckland, New Zealand. The population data have been extracted from the 2006 New Zealand census, which provides detailed demographic and socioeconomic information on the individuals.

In the New Zealand census, the classification system for ethnicity is hierarchical and is composed of five levels. The lowest (coarsest) level classifies an individual's ethnicity into four categories (i.e., European, Māori, Pacific peoples, and other ethnic groups), while the highest (most detailed) level has 231 categories. In this section, we use the second lowest-level data, which have five categories (European, Māori, Pacific peoples, Asian, and other ethnic groups), at the census areal unit (CAU) scale. CAUs are the second smallest geographic units after meshblocks in the New Zealand census, and each CAU is roughly equivalent to a suburb in urban areas. In 2006, the Auckland urban areas consisted of 297 CAUs, and this number should be sufficient to demonstrate how the proposed approach works.

4.2.1 Māori

The New Zealand Māori is the indigenous people of the country, whose ancestors are believed to be Polynesians originating from Southeast Asia. The Māori people, especially young and single, are geographically concentrated in urban areas: In 1936, only 17 % lived in urban areas, but this figure rose to 80 % by the time of the 1986 census. In 2006, 19.6 % of Māori resided in the Auckland urban areas, constituting 10.0 % of the regional population.

Figure 11a shows that although the concentration profile for the Māori population in Auckland (i.e., solid line) is clearly distinct from the line representing complete segregation (i.e., dashed line), it also differs from the idealized line of no segregation (i.e., dotted line). Looking more closely at the figure, it is possible to notice that nearly 20 % of the Māori people live in the areas where they comprise 20 % or more of the local populations and that more than 10 % are located in the CAUs where their co-ethnic residents occupy over 30 % in the population composition. There are, however, no CAUs where this significant, indigenous ethnic group forms the majority of the local populations.

The standardized area, R, is 0.109. This result implies that the geographic distribution of the Māori population is not even in the Auckland urban areas but the

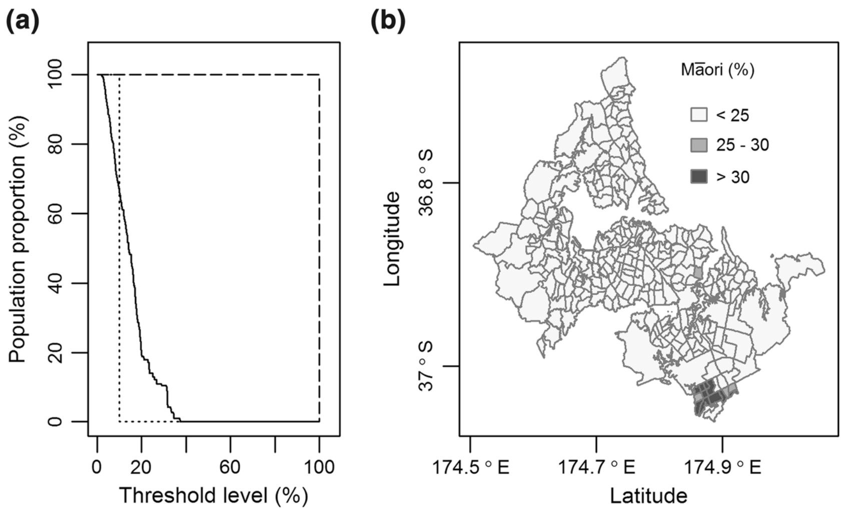

Fig. 11 Māori population in Auckland: **a** concentration profiles representing the group's population distribution (*solid line*), complete segregation (*dashed line*), and no segregation (*dotted line*), and **b** geographic distribution at the CAU level

degree of segregation is not severe. It is indicative of only a modest level of unevenness, and it conforms to the findings from earlier studies that employed, for instance, the index of dissimilarity (Grbic et al. 2010), ordinary least squares and logistic regression (Johnston et al. 2005), and local measures of autocorrelation (Johnston et al. 2011).

Interestingly, however, although the Māori population is not segregated in an absolute sense, the CAUs with relatively high proportions of Māori seem to be clustered in the southern part of Auckland (Fig. 11b). To evaluate the level of spatial clustering, the spatial proximity profile was constructed for $t = (0, 0.344]$, and the 95 % confidence intervals under the null hypothesis of random arrangements were estimated from a Monte Carlo simulation with 100 iterations (Fig. 12). At each trial of the simulation, the same number of CAUs in the corresponding X_t was randomly chosen, and the distances between the selected CAUs were calculated based on a first-order Queen's adjacency matrix, $\mathbf{W}$, each of whose elements is either 1, if two census tracts share at least one vertex, or 0, otherwise.

Figure 12 portrays a clear trend of increasing η_t with increasing threshold level, especially for $t > 0.2$. For the areas where the Māori people comprise <20 % of the local populations, the observed η_t values are generally lower than the upper bound of the 95 % confidence interval, providing little evidence against the null hypothesis of random arrangements. However, the profile indicates that those CAUs where Māori is relatively overrepresented (i.e., over 20 % in the local demographic composition) are more closely located to each other than expected. The degree of clustering for the CAUs with 20 % or more of Māori, $\eta_{0.20}$, is 0.205, which is more

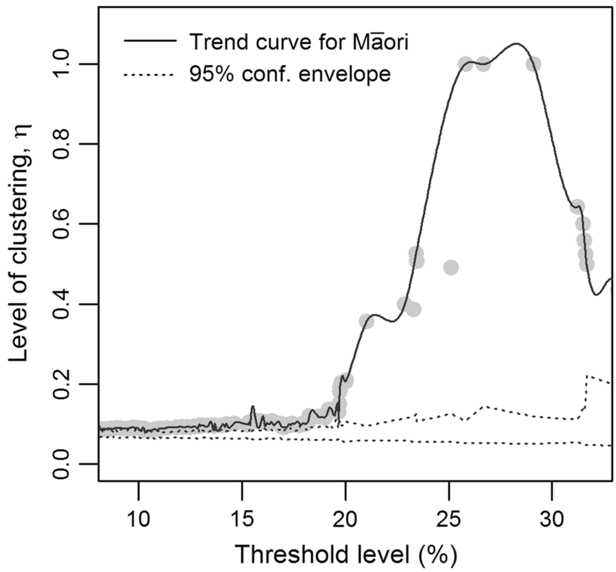

Fig. 12 Spatial proximity profile for the Māori population in Auckland (*solid line*) and the 95 % confidence intervals (*dotted lines*)

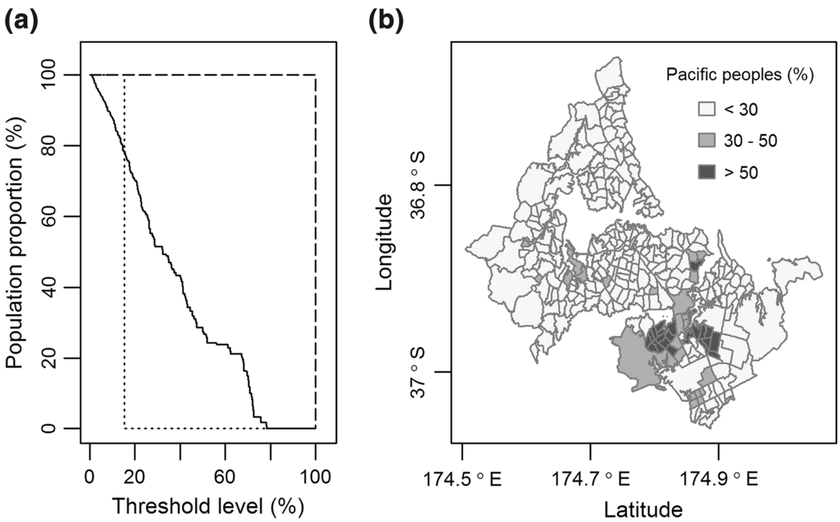

Fig. 13 Pacific peoples in Auckland: **a** concentration profiles representing the group's population distribution (*solid line*), complete segregation (*dashed line*), and no segregation (*dotted line*), and **b** geographic distribution at the CAU level

than double the upper bound of the 95 % confidence interval, 0.091, and at $t = 0.29$, the CAUs are completely clustered (i.e., $\eta_{0.29} = 1$).

These findings together suggest that although the majority of the Māori population is not residentially segregated, about 20 % of them reside in the areas where they are relatively overrepresented, and these areas are geographically agglomerated. Considering their share in the regional population, however, this is not a sign of residential separation between Māori and the other population groups; it only indicates the presence of their main residential areas in the Auckland urban areas.

4.2.2 Pacific peoples

Pacific peoples in the New Zealand census refer to those who identified with at least one Pacific ethnic group[2]. As with the Māori population, the large majority of them reside in urban areas. At the time of the 2006 census, almost four out of every five Pacific peoples lived in one of the three largest cities in the country (i.e., Auckland, Wellington, and Christchurch), and in particular, 63.8 % of them were concentrated in the Auckland urban areas.

[2] In the 2006 census, Samoan was the largest population group in this broad ethnic category (49.3 %), followed by Cook Islands Māori (21.8 %), Tongan (19.0 %), Niuean (8.4 %), Fijian (3.7 %), Tokelauan (2.5 %), and Tuvaluan (1.0 %). Although each of these groups has distinct cultural, linguistic, and historical backgrounds, we analyze the "Pacific peoples" data as a whole to demonstrate the use of the proposed approach. Note that the census respondents were able to choose more than one ethnic group to describe their ethnicity. The sum of the individual proportions in this data set therefore exceeds 100 %.

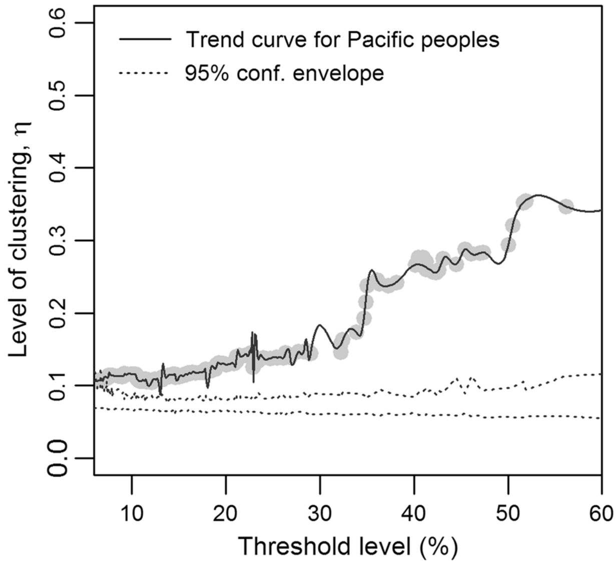

Fig. 14 Spatial proximity profile for Pacific peoples in Auckland (*solid line*) and the 95 % confidence intervals (*dotted lines*)

Pacific peoples constitute about 15.3 % of the regional demographic fabric, and as shown in Fig. 13a, they are somewhat unevenly distributed across the geographic units: Over 50 % of the population lives in the CAUs where they comprise at least 30 % of the local populations, and about 28.6 % reside in the areas where they form the majority. As a result, the standardized area, R, is higher than that for the Māori population ($R = 0.318$), indicating a greater level of unevenness.

In terms of the geographic clustering, they also appear to be highly clustered in the southern part of Auckland[3] (Fig. 13b). Similar to the Māori population, Fig. 14 presents a general trend of increasing η_t with increasing threshold level. The degree of clustering for the threshold of 30 % was 0.146, which is far beyond the 95 % confidence interval under the null hypothesis of randomness (i.e., 0.064 and 0.081), and for the threshold of 50 %, it goes up to 0.293.

What is different from the residential pattern of Māori is that Pacific peoples appear to be more residentially separated from the other groups in the study region. For the Māori population, about 80 % of them reside in the areas where their co-ethnic residents occupy <20 % in the population composition, and those areas are not geographically clustered (Fig. 12). In case of Pacific peoples, on the other hand, there are only 20 % of them living in the CAUs where they comprise <20 % of the local population, and Fig. 14 suggests that even those CAUs are geographically concentrated in certain parts of the region, the southern Auckland. This result implies that the overall degree of segregation is much higher for Pacific peoples.

[3] While it may be beyond the scope of this paper to explain why Pacific peoples (and the Māori population) are clustered in the southern Auckland, previous studies suggest that the presence of state housing in this locality has attracted low-income Pacific peoples (Johnston et al. 2008).

4.2.3 Asians

The Asian population in New Zealand has increased rapidly over the last two decades, largely as a result of immigration from China, India, and South Korea. In the 2006 census, the Asian people formed about 9 % of the national population, and they made up more than one-fifth of the local population in the study region.

The concentration profile for the Asian population suggests a modest level of residential segregation (Fig. 15a): More than two-fifth of them lives in the areas where they constitute 30 % or more of the population, and there are several census tracts where they even form the majority in the demographic structure. The standardized area of this curve lies between Māori ($R = 0.109$) and Pacific peoples ($R = 0.318$), $R = 0.202$, indicating that the degree of residential concentration for the Asian population is somewhat more severe than that for Māori but not as much as Pacific peoples. Although the traditional index of dissimilarity, D, provides the same relative levels of segregation for the three groups (i.e., D for Māori, Pacific peoples, and Asians are 0.314, 0.520, and 0.324, respectively), it seems to underestimate the Asian ethnic group's residential concentration in the study region.

Compared to the other two groups in the previous sections, the CAUs where the Asian population is relatively concentrated appear to be scattered over the study region (Fig. 15b). The census tracts with high proportions of the Asian people (i.e., >25 %) are separately clustered in several different places, without forming one large cluster, and as a result, the spatial proximity profile in Fig. 16 indicates no statistically significant clustering. This is probably because individual population groups within the broad category of "Asian," such as Chinese, Indians, and Koreans, maintain their own ethnic communities in different parts of the region (see,

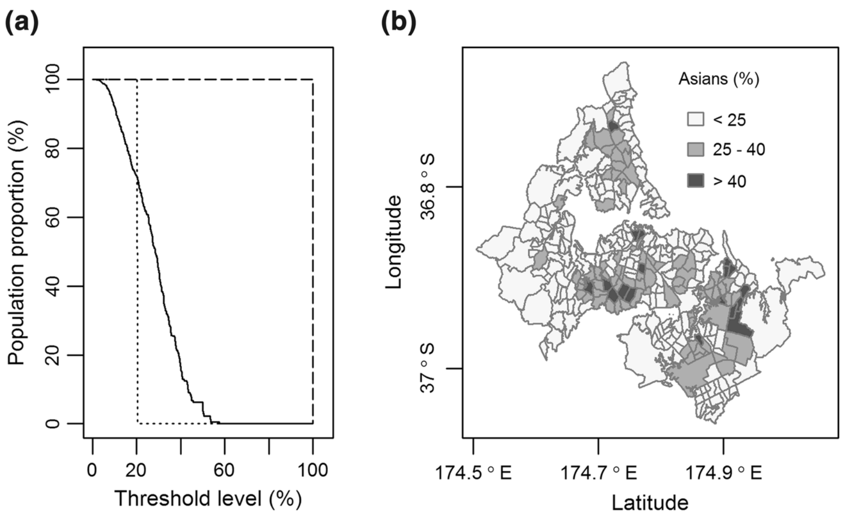

Fig. 15 Asians in Auckland: **a** concentration profiles representing the group's population distribution (*solid line*), complete segregation (*dashed line*), and no segregation (*dotted line*), and **b** geographic distribution at the CAU level

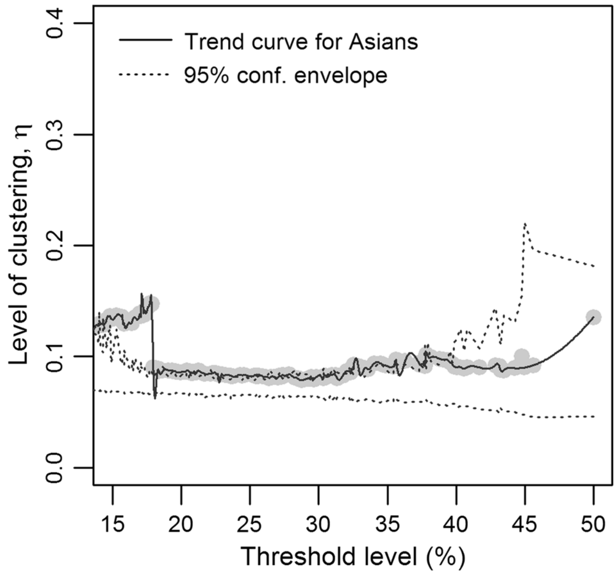

Fig. 16 Spatial proximity profile for Asians in Auckland (*solid line*) and the 95 % confidence intervals (*dotted lines*)

e.g., Friesen et al. 2005 for the case of Indians; Xue et al. 2012 for Chinese; Hong and Yoon 2013 for Koreans). If we constructed a spatial proximity profile for one of the subgroups, the trend would not be much different from those of Māori and Pacific peoples.

It may be worth noting that while the profile curve in Fig. 16 is mostly located within the 95 % confidence envelope, the degree of clustering increases to a significant level when the threshold value is set to around 15 %. This implies that the inclusion of the "mixed" areas (i.e., the CAUs where the Asians comprise 15–20 % of the local population) may connect the separated clusters, making them a single large cluster.

5 Conclusions

This paper presented an alternative, graph-based approach that can depict segregation in more detail. Although the existing indices of segregation have their own merits in summarizing various dimensions of segregation, the reliance on these single-figure indices can result in the misleading interpretation of the phenomenon under examination (Johnston et al. 2010; Wright et al. 2011). We, therefore, introduced alternative methods that provide not only single-figure summary statistics but also a visual platform that enables examining the patterns of segregation in more detail.

The proposed approach consists of two stages: (1) The concentration profile provides information on how evenly (or unevenly) the population of interest is distributed in the study region and (2) the spatial proximity profile focuses on the spatial dimension of segregation by calculating the overall distance between geographic units where the population of interest comprises at least a given threshold proportion in the local demographic composition.

Although the concentration profile (and other similar threshold approaches) has been used in earlier studies of residential segregation, we attempted to define it in a more formal manner and demonstrated how we can obtain a summary statistic from the profile curve. As demonstrated with the idealized patterns and the New Zealand census data, the normalized area, R, is well capable of capturing the unevenness distribution of a particular population group, and it can also help make more objective comparisons between different graphs.

Because the concentration profile does not address the spatial dimension of segregation—clustering, we developed another segregation profile that describes the degree of spatial concentration for a range of threshold values. These two segregation curves complement each other, and they together can portray the complex, multidimensional phenomenon of segregation more accurately than the traditional single-figure index approach.

It should be worth noting, however, that the proposed methods focus on the distribution of an individual population group. One advantage of this attribute is that while segregation measures that account for multiple groups risk rendering the geographic clustering of small populations undetectable, our approach can more accurately depict the minority groups' residential experience. However, the proposed methods are limited in helping unravel the complex patterns of social and residential interactions between the population groups, and this limitation should be addressed in future research.

Acknowledgments This work was supported by JSPS KAKENHI Grant Number 24-02309. We are thankful to the anonymous referees for their constructive comments.

References

Allen R, Burgess S, Windmeijer F (2009) More reliable inference for segregation indices. CMPO Working Paper Series, vol 216. University of Bristol, Centre for Market and Public Organisation, Bristol

Dorfman R (1979) A formula for the Gini coefficient. Rev Econ Stat 61(1):146–149

Duncan OD, Duncan B (1955) A methodological analysis of segregation indexes. Am Sociol Rev 20(2):210–217

Falk RF, Cortese CF, Cohen J (1978) Utilizing standardized indices of residential segregation: comment on Winship. Soc Forces 57(2):713–716. doi:10.2307/2577693

Friesen W, Murphy L, Kearns R (2005) Spiced-up Sandringham: Indian transnationalism and new suburban spaces in Auckland, New Zealand. J Ethn Migr Stud 31(2):385–401

Grbic D, Ishizawa H, Crothers C (2010) Ethnic residential segregation in New Zealand, 1991–2006. Soc Sci Res 39(1):25–38

Hong S-Y, Yoon H-K (2013) Ethno-economic satellite: the case of Korean residential clusters in Auckland. Popul Space Place. doi:10.1002/psp.1764

Hutchens RM (1991) Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations. Math Soc Sci 21(1):31–51. doi:10.1016/0165-4896(91)90038-S

Johnston R, Poulsen M, Forrest J (2005) Ethnic residential segregation across an urban system: the Maori in New Zealand, 1991–2001. Prof Geogr 57(1):115–129. doi:10.1111/j.0033-0124.2005.00464.x

Johnston R, Poulsen M, Forrest J (2008) Asians, Pacific Islanders, and ethnoburbs in Auckland, New Zealand. Geogr Rev 98(2):214(228)

Johnston R, Poulsen M, Forrest J (2010) Moving on from indices, refocusing on mix: on measuring and understanding ethnic patterns of residential segregation. J Ethn Migr Stud 36(4):697–706. doi:10.1080/13691830903505045

Johnston R, Poulsen M, Forrest J (2011) Evaluating changing residential segregation in Auckland, New Zealand, using spatial statistics. Tijdschr Econ Soc Geogr 102(1):1–23. doi:10.1111/j.1467-9663.2009.00577.x

Massey DS, Denton NA (1988) The dimensions of residential segregation. Soc Forces 67(2):281–315

Morgan BS (1983) An alternate approach to the development of a distance-based measure of racial segregation. Am J Sociol 88(6):1237–1249

Morrill RL (1991) On the measure of geographic segregation. Geogr Res Forum 11:25–36

Openshaw S (1984) The modifiable areal unit problem. Geo, Norwich

O'Sullivan D, Wong DWS (2007) A surface-based approach to measuring spatial segregation. Geogr Anal 39(2):147(122)

Poulsen M, Johnston R, Forrest J (2002) Plural cities and ethnic enclaves: introducing a measurement procedure for comparative study. Int J Urban Reg Res 26(2):229–243

Reardon SF, O'Sullivan D (2004) Measures of spatial segregation. Sociol Methodol 34(1):121–162

Reardon S, Matthews S, O'Sullivan D, Lee B, Firebaugh G, Farrell C, Bischoff K (2008) The geographic scale of Metropolitan racial segregation. Demography 45(3):489–514. doi:10.1353/dem.0.0019

Reardon SF, Farrell CR, Matthews SA, O'Sullivan D, Bischoff K, Firebaugh G (2009) Race and space in the 1990s: changes in the geographic scale of racial residential segregation, 1990–2000. Soc Sci Res 38(1):55–70

White MJ (1983) The measurement of spatial segregation. Am J Sociol 88(5):1008–1018

Winship C (1977) A revaluation of indexes of residential segregation. Soc Forces 55(4):1058–1066. doi:10.2307/2577572

Wong DWS (1993) Spatial indices of segregation. Urban Stud 30(3):559–572

Wong DWS (1997) Spatial dependency of segregation indices. Can Geogr 41(2):128–136

Wong DWS, Reibel M, Dawkins C (2007) Introduction—segregation and neighborhood change: where are we after more than a half-century of formal analysis. Urban Geogr 28(4):305–311

Wright R, Holloway S, Ellis M (2011) Reconsidering both diversity and segregation: a reply to Poulsen, Johnston and Forrest, and to Peach. J Ethn Migr Stud 37(1):167–176. doi:10.1080/1369183x.2011.523865

Xue J, Friesen W, O'Sullivan D (2012) Diversity in Chinese Auckland: hypothesising multiple ethnoburbs. Popul Space Place 18:579–595. doi:10.1002/psp.688