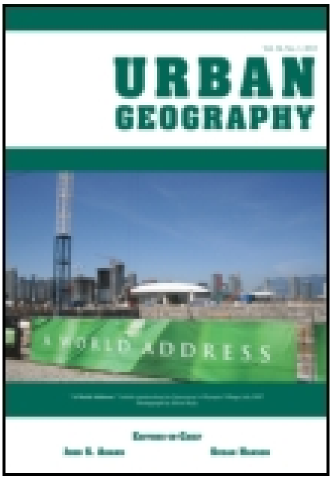

Urban Geography
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rurb20

MEASURING MULTIETHNIC SPATIAL SEGREGATION

David W. S. Wong [a]
[a] Geography and Earth Systems Science, George Mason University, 4400 University Drive Fairfax, Virginia 22030 Tel: 703-993-1212 Fax: 703-993-1216 dwong2@gmu.edu
Published online: 16 May 2013.

To cite this article: David W. S. Wong (1998) MEASURING MULTIETHNIC SPATIAL SEGREGATION, Urban Geography, 19:1, 77-87, DOI: 10.2747/0272-3638.19.1.77

To link to this article: http://dx.doi.org/10.2747/0272-3638.19.1.77

RESEARCH NOTE

MEASURING MULTIETHNIC SPATIAL SEGREGATION[1]

David W. S. Wong
Geography and Earth Systems Science
George Mason University
4400 University Drive
Fairfax, Virginia 22030
Tel: 703-993-1212
Fax: 703-993-1216
dwong2@gmu.edu

Abstract: Several extensively used segregation measures have known limitations in differentiating various spatial configurations of population distribution. Recently, new segregation indices overcoming this deficiency were introduced, but are only applicable to unrealistic two-group situations. In this paper, I attempt to model multiethnic segregation on the basis of the notions that segregation refers to the spatial separation among ethnic groups and that interaction among population groups can reduce segregation. The level of segregation is reduced when interaction among ethnic groups increases. The spatial index introduced here is based on an existing segregation measure for a multigroup setting. I demonstrate that the proposed index is efficient in distinguishing different spatial patterns of multiethnic population and is able to reflect the potential for spatial interaction among groups. A simulation and an empirical study are used to illustrate the utility of the proposed index.

The index of dissimilarity or the segregation index (D) proposed by Duncan and Duncan (1955) has been used extensively to reflect the magnitude of separation between two ethnic groups. In brief, D is defined as

$$D = 0.5 \times \sum_i \left| \frac{b_i}{B} - \frac{w_i}{W} \right| \qquad (1)$$

where b_i and w_i are the Black and White population counts in areal unit i, and B and W are the total Black and White population counts in the entire study area. D ranges from 0 (no segregation) to 1 (a perfectly segregated situation). Given the multiethnic environments in many societies and especially in North America, measuring multigroup segregation is more realistic. Using several 2-group indices, Morrill (1995) compared different ethnic and social groups in a pair-wise manner. Morgan (1975) and Sakoda (1981) introduced an index to summarize multigroup segregation. To differentiate this index from the conventional two-group segregation index, we denote this multigroup index as D(m):

$$D(m) = \frac{1}{2} \frac{\sum_i \sum_j |N_{ij} - E_{ij}|}{\sum_j NP_{.j}(1 - P_{.j})} \qquad (2)$$

Urban Geography, 1998, **19**, 1, pp. 77–87.

where

$$E_{ij} = \frac{N_{i.} N_{.j}}{N}.$$

N_{ij} is the population count of ethnic group j in areal unit i, $N_{i.}$ is the total population count in areal unit i, $N_{.j}$ is the total population count in ethnic group j, N is the total population count in the entire study area, and $P_{.j}$ is the proportion of population in ethnic group j. E_{ij} is the expected population size of ethnic group j in areal unit i under the assumption of a uniform (or proportional) distribution. This index can be interpreted in the same manner as D, and also shares many characteristics with D. As D(m) increases, segregation level increases accordingly. Besides D(m), other segregation measures for multigroup situations include the entropy index (Theil, 1972; White, 1986), the interaction index, and the isolation index (Lieberson, 1981).

However, D and D(m), as well as several other popular measures of segregation, are nonspatial in two respects. First, both D and D(m) can be regarded as the effort required to move the ethnic populations among areal units in order to achieve an even distribution among groups. Jakubs (1979, 1981) argued that the effort of moving population indicated by D fails to include the distance of relocation and thus does not account accurately for the effort of movement. Therefore, he proposed a measure including a distance component to indicate the effort required to relocate population to achieve minimum segregation. White's (1986) index of spatial proximity and Morgan's (1983) interaction index use a similar approach. Second, both D and D(m) fail to differentiate the distribution of population in various spatial configurations. Morrill (1991) and Wong (1993) suggested new indices for two-group situations by incorporating spatial autocorrelation measures. These measures implicitly account for the potential of spatial interaction among groups (see Getis, 1991, for a discussion) and can differentiate spatial patterns of ethnic mix efficiently. However, all of the above measures of segregation, which treat spatial structure of population explicitly, share a major limitation with other segregation measures. These indices can be applied only to dichotomous situations, whereas in reality we always have to deal with multigroup settings.

This paper proposes a spatial measure of segregation that shares some features with the spatial indices proposed by Morrill (1991) and Wong (1993), but is applicable to multigroup situations. The starting point is the multigroup index introduced by Morgan (1975) and Sakoda (1981). The proposed measure is based explicitly on the notion that segregation implies spatial separation of population groups (Newby, 1982). Thus, the magnitude of segregation among ethnic groups is inversely related to the opportunity of interaction among ethnic groups.

A MULTIGROUP SPATIAL SEGREGATION MEASURE

In reality, population groups defined by ethnicity or socioeconomic criteria in an area can interact with population of other groups in neighboring areas. D does not take into account this aspect. To overcome this limitation, Morrill (1991) and Wong (1993) included a spatial term capturing topological relationships among areal units in their indices to account for interaction of ethnic groups among areal units. In its simplest form, the spatial term is denoted as c_{ij}, where i and j represent areal units, and $c_{ij} = 1$ when i and j are neighbors, and 0 otherwise. The resultant indices can distinguish different patterns of ethnic distribution.

D(m) shares the same limitation with D. The index implicitly assumes that different ethnic groups in different areal units cannot interact to reduce segregation. Therefore, in Equation 2, the difference between N_{ij} (the observed population count) and E_{ij} (the expected population count) in areal unit i will not be compensated by the differences found in the surrounding areal units and thus is regarded as the segregated population. However, people move across enumeration boundaries and mingle with people of other ethnic groups in neighboring units. Therefore, when the observed population count of a particular ethnic group exceeds the expected in areal unit i, the excess population can still mix with other groups in the neighboring units and thus are not isolated spatially. If neighboring areal units of i have population counts of that particular ethnic group below the expected counts, then the excess population of a particular group in i can balance the "deficit" of the same group in the neighboring units. Thus, the "deficit" population counts of an ethnic group in one areal unit may be compensated by the "surpluses" from the neighboring units.

The basic premise of the spatial segregation measure for a multigroup setting is that the population of an ethnic group in areal unit i has access to the total population of other ethnic groups in all surrounding units. In order to reflect that populations of different ethnic groups in an areal unit i can mix with populations of different ethnic groups in the surrounding areal units in the calculation of D(m), a composite population count is proposed. The composite population count for areal unit i for ethnic group j is defined as

$$CN_{ij} = \sum_k d(N_{kj}) \tag{3}$$

where d(.) is a function defining the surrounding units of areal unit i and k, which can be equal to i, refers to areal units 1, 2, 3, . . . n. In addition, CN_{ij} should not be calculated if $N_{i.}$, which is the total population in areal unit i, equals 0. This composite population count includes population counts of the same ethnic group in all surrounding units. Then, the multigroup spatial index of segregation SD(m) can be defined as

$$SD(m) = \frac{1}{2} \frac{\sum_i \sum_j |CN_{ij} - CE_{ij}|}{\sum_j CN \times CP_{.j}(1 - CNP_{.j})} \tag{4}$$

where

$$CE_{ij} = \frac{CN_{i.} CN_{.j}}{CN}.$$

CN_{ij} is the composite population count of ethnic group j in areal unit i, $CN_{i.}$ is the total composite population count in areal unit i, $CN_{.j}$ is the total composite population count of ethnic group j, CN is the total composite population in the entire study area, and $CP_{.j}$ is the proportion of composite population in ethnic group j—that is, $CN_{.j}/CN$. CE_{ij} can be interpreted as the expected composite population of ethnic group j in areal unit i.

The calculation of SD(m) is the same as the nonspatial multigroup index. The only difference is that the population counts of different ethnic groups are not the actual population counts, but rather are derived population counts assuming that population of an ethnic group in one areal unit can mix with population of different ethnic groups in neighboring areal units. Thus, SD(m) and D(m) share almost all mathematical properties. For instance, the range of SD(m) is between 0 and 1, where 0 indicates no segregation and 1 refers to perfect segregation. Also, SD(m) is capable of differentiating between various spatial patterns of ethnic group distribution. This will be demonstrated in the empirical study and the simulation experiment described later.

The function d(.) defining the surrounding units can carry many forms. Griffith (1995) provided justifications for different specifications of the matrix. A widely adopted framework is contiguity. If areal units share a common boundary, then they are weighted equally and grouped together. This concept can be implemented by using a modified binary connectivity matrix *C*. In this matrix, $c_{ik} = 1$ when i and k are neighbors, and 0 otherwise. Normally, $c_{ii} = 0$, indicating that i is not a neighbor to itself. However, when we derive the composite population counts for areal unit i, we should include the areal unit itself. Therefore, the composite population count of area i for ethnic group j using the contiguity framework will be

$$CN_{ij} = \sum_{k}^{n} c_{ik} N_{kj} \tag{5}$$

where $c_{ii} = 1$. That is, CN_{ij} is the sum of population in ethnic group j for all areal units sharing a boundary with areal unit i plus the population of ethnic group j in i. This specification of the spatial weights matrix and hence the composite population counts assumes that there is no obstacle between people in neighboring areal units to hinder the mixing of different ethnic groups across boundaries. The population in one areal unit can interact with the population in neighboring units as if in its own area.

Different ways of specifying the spatial weights structure and the composite population counts reflect different perceptions of how the population is separated spatially. If we assume that the spatial interaction among ethnic groups in different areal units is inversely related to the distance between areal unit i and all other areal units in the system, then we can use the inverse of distance between each pair of areal units when we create the spatial weights matrix. The ijth cell in the spatial weights matrix *W* will be $w_{ik} = 1/d_{ik}$, where d_{ik} is the distance between the centroids in areal units i and k, and $d_{ii} = 1$. We can also follow White (1986) by using a negative exponential function of distance so that $w_{ik} = \exp(-d_{ik})$.

All the above specifications of d(.) use the same matrix for all ethnic groups, assuming that different ethnic groups interact spatially in the same manner and magnitude. It is possible that different ethnic groups interact differently with other ethnic groups. In this case, instead of using the same spatial weights matrix for all ethnic groups, we can use a spatial weights matrix for each group to accommodate the variation of spatial interaction among groups.

In the remainder of this paper, I will adopt the binary specification of the weight matrix *C* just for illustrative purposes. Also, using the binary matrix can demonstrate effectively a common problem in using spatial measures. In the next section, we will use a simulation

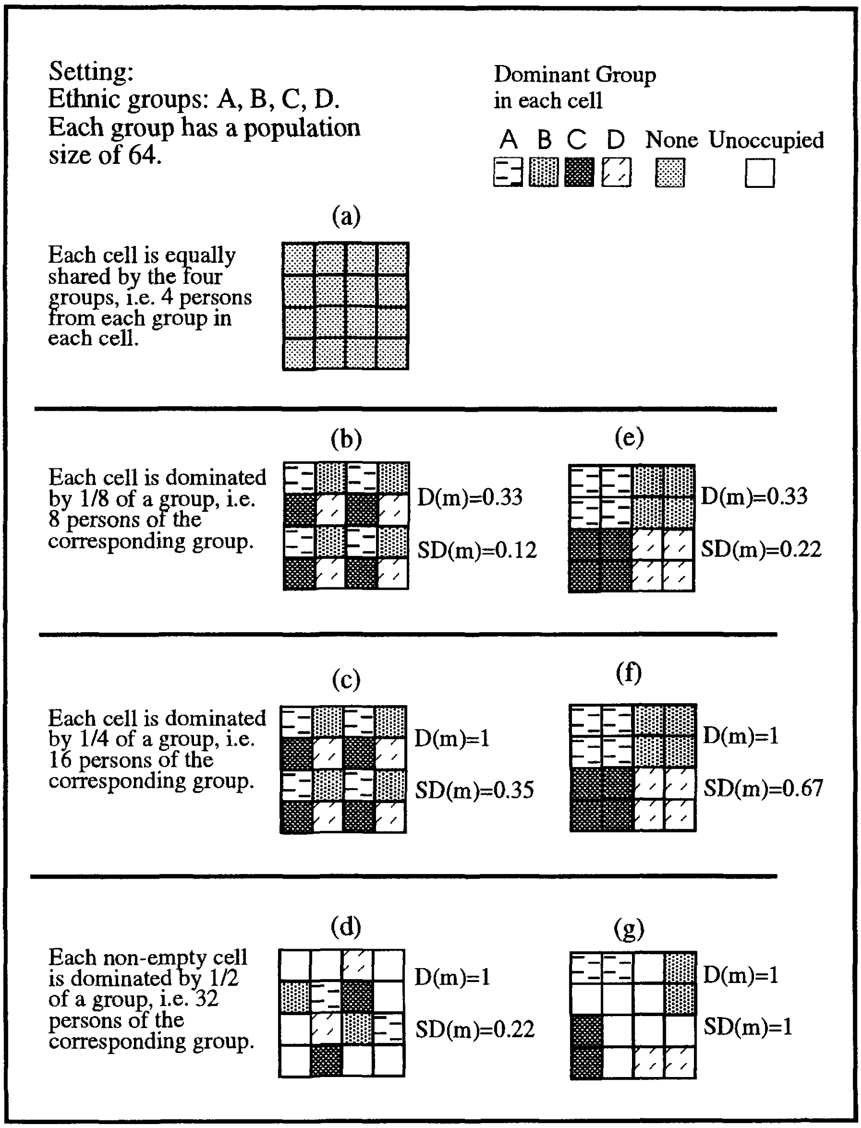

Fig. 1. D(m), SD(m) for simulated configurations.

to demonstrate the strengths of the proposed spatial segregation index for the multigroup situation. Then an empirical example using the state of Connecticut will follow.

SIMULATION

The setting of the experiment is described in Figure 1. Four ethnic groups (A, B, C, and D) are used. The total population of the whole study area is kept constant (256), and each group has equal population size (64). Different patterns are used to indicate the groups dominating a unit.

In the first column of configurations in Figure 1, configurations a, b, c, and d depict either a perfectly even distribution (a) or situations in which different ethnic groups are spatially juxtaposed to facilitate spatial interaction (b, c, and d). The configurations in the second column (e, f, and g) represent situations in which an ethnic group dominates several adjacent cells. These configurations are identical to their corresponding configurations in the first column in terms of proportion of population allocated to each cell. The only difference is that these configurations are created by spatially rearranging cells of their corresponding configurations in the first column so that certain regions in the simulated area are dominated by particular groups. Therefore, configurations in the second column should have higher levels of segregation than their respective configurations in the first column.

For each configuration in Figure 1, we calculate the original multigroup index D(m) and the spatial multigroup index SD(m). The D(m) measures for the two columns of configurations are identical, though we expect that configurations in the second column (e, f, and g) are more segregated than their counterparts (b, c, and d) in the first column. D(m) does not change at all even when cells are spatially rearranged. It is obvious that e is more segregated than b.

In contrast, SD(m) can discriminate between the two groups of configurations very well. The configurations in the second column in Figure 1 always yield a higher SD(m) measure than the corresponding figures in the first column. Therefore, the proposed SD(m) is much more effective than D(m) in discriminating different spatial patterns of population distribution.

Intuitively, configuration a should be the least segregated among all. The results confirm this expectation, as both D(m) and SD(m) equal 0. When cells are increasingly dominated by an ethnic group, the simulation results indicate that D(m) is not a good index for capturing this change, whereas SD(m) is more efficient in capturing the trend. D(m) is 0.33 in both configurations b and e when each cell is dominated by 1/8 of the total population of an ethnic group. D(m) jumps to 1 (a perfectly segregated situation) in configurations c and f when each cell is completely occupied by 1/4 of the total population of an ethnic group. However, according to SD(m), configurations c and f are more segregated than their counterparts in configurations b and e, but they are far from perfect segregation because the spatial configurations allow populations of different ethnic groups in neighboring units to interact. As long as all areal units in the study area are exclusively occupied by one of the population groups, regardless of the spatial pattern of the ethnic distribution, both D and D(m) will equal 1, indicating a perfectly segregated situation.

In configurations d and g, some cells are not occupied and each occupied cell is dominated by a group. Configuration d allows interaction between groups across areal units, while in configuration g, the four groups are spatially separated. Therefore, SD(m) = 1 for configuration g. Not surprisingly, D(m) = 1 for these two configurations, indicating perfect segregation. In configuration d, though areal units are ethnically separated, they are not spatially separated. In fact, the cells are very well connected. All four central cells have three neigboring cells occupied by other ethnic groups, but in configuration c, each cell has neighbors of only two other ethnic groups. Thus, SD(m) in configuration d is lower than 1 and is also lower than SD(m) in configuration c.

TABLE 1.—ETHNIC MIX AND SEGREGATION MEASURES FOR CONNECTICUT (CT)

Counties	D(m)	Ranked by D(m)	SD(m)	Ranked by SD(m)	# Tract	% White	% African	% Indian	% Asian	% Others
CT (counties)					834	87.08	8.32	0.20	1.48	2.89
Fairfield (FAIRF)	0.59	6	0.48	6	213	84.77	9.88	0.12	1.98	3.23
Hartford (HARTF)	0.65	8	0.57	8	228	83.66	10.16	0.20	1.52	4.45
Litchfield (LITCH)	0.38	1	0.23	1	51	97.87	0.89	0.15	0.83	0.23
Middlesex (MIDDL)	0.48	5	0.32	3	34	94.05	4.20	0.16	0.96	0.61
New Haven (NEWHA)	0.62	7	0.52	7	185	85.61	10.15	0.18	1.22	2.83
New London (NEWLO)	0.44	3	0.33	4	69	91.78	4.80	0.56	1.36	1.48
Tolland (TOLLA)	0.41	2	0.24	2	29	95.27	1.99	0.20	1.95	0.55
Windham (WINDH)	0.47	4	0.36	5	25	95.80	1.12	0.36	0.81	1.89

A CASE STUDY

A case study is presented here using 1990 U.S. Census population data for the state of Connecticut. The purpose of this case study is not to provide a thorough investigation and explanation of segregation in Connecticut, but to demonstrate how SD(m) can be implemented in real-world situations. Connecticut has eight counties and 838 census tracts. Using census tracts as the basic areal unit and the five ethnic groups (White; Black; American Indian, Eskimo, or Aleut; Asian or Pacific Islander; and race other than White) defined by the Bureau of the Census (1992), D(m) and SD(m) are calculated for each county in Connecticut. The number of areal units, ethnic mix in each county, and segregation measures are reported in Table 1.

In general, SD(m) is lower than D(m) in each county because SD(m) also accounts for the spatial interaction of population in different ethnic groups among areal units, which lowers the magnitude of segregation. D(m) reflects only the degree of internal homogeneity of census tracts in each county. However, SD(m) is not just a deflated D(m). Though the differences between D(m) and SD(m) are not dramatic (the largest difference is 0.17 found in Tolland County), they do not provide identical rank orders for the eight counties we study here (Table 1). Three counties (Middlesex, New London, and Windham) have different rankings in D(m) and SD(m). Figure 2 can provide some insights in explaining these differences.

In Connecticut, the White population is the largest ethnic group; therefore, other groups may be referred to as minority groups. Figure 2 shows the percentage of minority population in each census tract in Connecticut. Among the three counties having different rankings in the two indices, according to D(m) New London County (NEWLO) has the lowest segregation level and Middlesex County (MIDDL) has the highest. When we

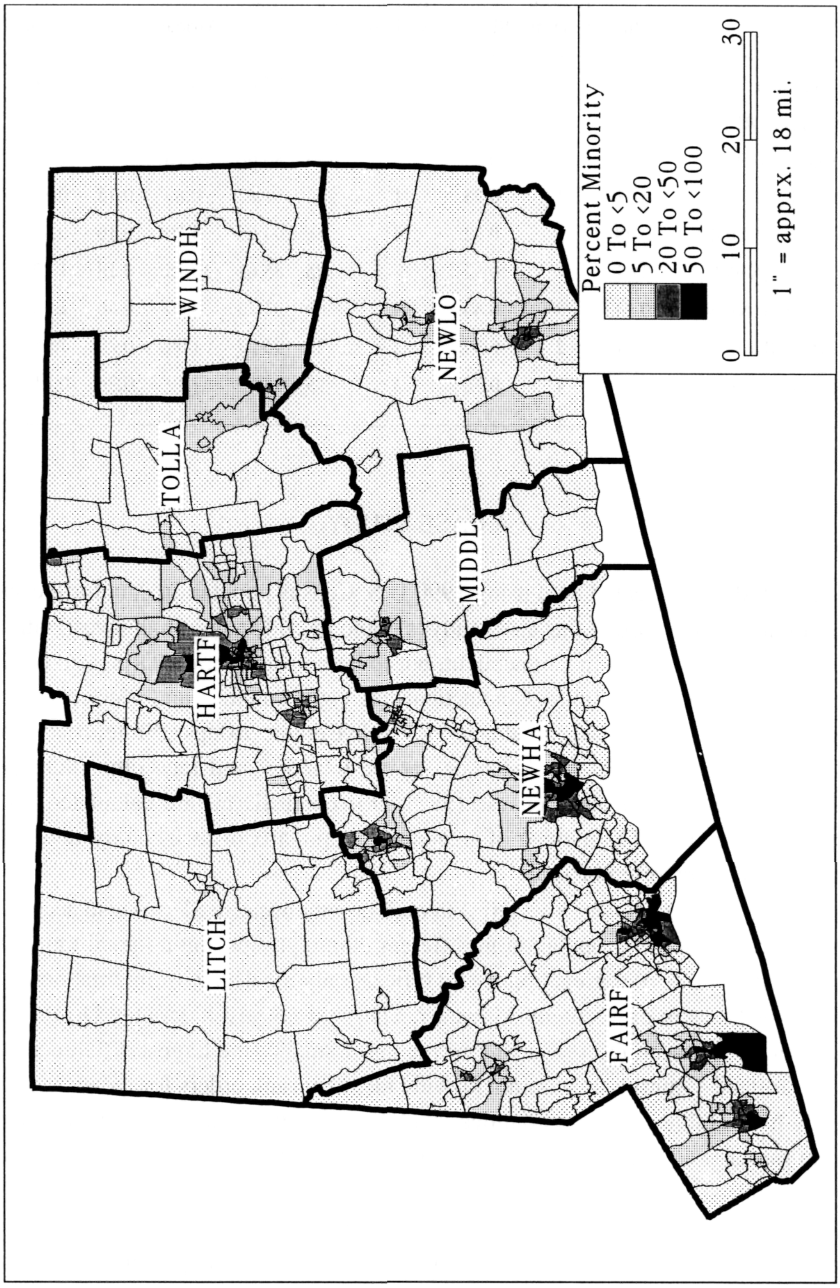

Fig. 2. Percentage of minority population in each tract, Connecticut.

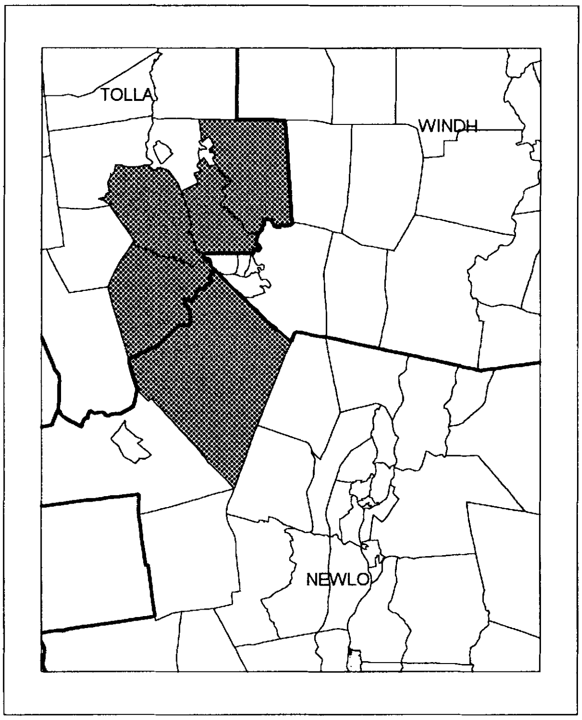

Fig. 3. Tracts that are "annexed" to Windham County to eliminate the boundary effect in SD(m) calculation.

inspect Figure 2 visually, we may not be able to tell the differences in segregation among these three counties. However, when we use SD(m), Windham County (WINDH) is highest and Middlesex County (MIDDL) is lowest among the three. Referring to the formulation of SD(m), SD(m) mixes the population of each areal unit with its neighboring units and then measures the unevenness of population counts conditioned upon the total population count in each areal unit and in each ethnic group. We can also think of SD(m) as a measure reflecting how well we can "spread" populations in different ethnic groups to the surrounding areas to achieve an "even" distribution. Middlesex County has only one area with a high concentration of minority population. New London County has two areas. These high-concentration areas are at relatively central locations in the counties. Therefore, it is relatively easy to "spread" the minority population to the surrounding area units. Furthermore, it is easier to "spread" the minority population in Middlesex County than in New London County, because New London County has two concentration areas. Intuitively, Windham County should have a lower level of SD(m) because it has only one region (southwest corner of the county) with high minority population. However, that high-concentration area borders Tolland County. In the set-up of this case study, population in a county cannot mix with the population in another county. Therefore, among the tracts with a high minority population in Windham County, there are not many with which the minority population can interact, causing the SD(m) for Windham to be relatively high because of the boundary effect (Griffith, 1983). If we permit interaction between the tracts at the southwest corner of Windham County and the adjacent tracts (Fig. 3), the SD(m) of

Windham County will drop from 0.36 to 0.33, which is the same segregation level as Middlesex and New London counties. In short, SD(m) not only consistently produces a lower segregation level than D(m), but also may produce different rankings in segregation level.

CONCLUSIONS

In a multiethnic environment, segregation can be interpreted as spatial separation of different ethnic groups that inhibits interaction. Adopting this notion of segregation, a spatial index of segregation SD(m) for a multigroup setting is proposed in contrast to an existing multigroup segregation measure D(m) that does not include a spatial component. The calculation of SD(m) utilizes the same idea as D(m). The core idea of SD(m) lies in how the population count in each areal unit is defined. Because spatial interaction among populations of different ethnic groups can lower segregation, the population of an ethnic group of an areal unit is a composite population count including the population of the same group of all neighboring units. We assume that people can interact across areal boundaries. This assumption can be modified, and SD(m) can accommodate changes easily. The proposed index can also effectively differentiate between various spatial patterns of population distribution.

Though the discussion focused on multiethnic segregation, the proposed index is applicable to other types of segregation, such as that found in schools and among social and economic groups. Morrill (1991) proposed a spatial version of D, the index of dissimilarity, which Wong (1993) later enhanced. Morgan (1975) introduced D(m), a multigroup version of D, but it cannot distinguish different spatial patterns of population distribution, which is an underlying structure affecting segregation. SD(m), a spatial version of D(m), is proposed in this paper as a measure of multigroup segregation that can account for spatial structure.

NOTE

[1]The author is grateful for the constructive comments of Carl Amrhein at the University of Toronto, and the editorial efforts of Joseph Wood at George Mason University. The comments from the referees and the editor are also gratefully acknowledged.

LITERATURE CITED

Bureau of the Census, 1992, *Census of Population and Housing, 1990: Summary Tape File 3 on CD-ROM Technical Documentation.* Washington, DC: Author.

Duncan, D. and Duncan, B., 1955, A methodological analysis of segregation indexes. *American Sociological Review*, Vol. 20, 210–217.

Getis, A., 1991, Spatial interaction and spatial autocorrelation: A cross product approach. *Environment and Planning A*, Vol. 23, 1269–1277.

Griffith, D.A., 1983, The boundary value problem in spatial statistical analysis. *Journal of Regional Science*, Vol. 23, 377–387.

______, 1995, Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In S. L. Arlinghaus and D. A. Griffith, editors, *Practical Handbook of Spatial Statistics.* New York, NY: CRC, p. 65–82.

Jakubs, J. F., 1979, A consistent conceptual definition of the index of dissimilarity. *Geographical Analysis,* Vol. 11, 315–321.

______, 1981, A distance based segregation index. *Journal of Socio-Economic Planning Sciences*, Vol. 15, 129–141.

Lieberson, S., 1981, An asymmetrical approach to segregation. In C. Peach, V. Robinson, and S. Smith, editors, *Ethnic Segregation in Cities.* New York, NY: Croom Helm, 61–82.

Morgan, B. S., 1975, The segregation of socioeconomic groups in urban areas. *Urban Studies*, Vol. 12, 47–60.

______, B. S., 1983, A distance-decay interaction index to measure residential segregation. *Area*, Vol. 15, 211–216.

Morrill, R. L., 1991, On the measure of geographical segregation. *Geography Research Forum,* Vol. 11, 25–36.

______, 1995, Racial segregation and class in a liberal metropolis. *Geographical Analysis*, Vol. 27, 22–41.

Newby, R. G., 1982, Segregation, desegregation, and racial balance: Status implications of these concepts. *The Urban Review*, Vol. 14, 17–24.

Sakoda, J. N., 1981, A generalized index of dissimilarity. *Demography*, Vol. 18, 245–250.

Theil, H., 1972, *Statistical Decomposition Analysis*. New York, NY: North-Holland.

White, M. J., 1986, Segregation and diversity measures in population distribution. *Population Index*, Vol. 52, 198–221.

Wong, D. W. S., 1993, Spatial indices of segregation. *Urban Studies*, Vol. 30, 559–572.