Segregation through the multiscalar lens

Madalina Olteanu[a,b,1], Julien Randon-Furling[a], and William A. V. Clark[c]

[a]Statistique, Analyse et Modélisation Multidisciplinaire, Université Paris 1 Panthéon Sorbonne, 75013 Paris, France; [b]Unité de Recherche Mathématiques et Informatique Appliquées du Génome à l'Environnement, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas, France; and [c]Department of Geography, University of California, Los Angeles, CA 90095

We introduce a mathematical framework that allows one to carry out multiscalar and multigroup spatial exploratory analysis across urban regions. By producing coefficients that integrate information across all scales and that are normalized with respect to theoretical maximally segregated configurations, this framework provides a practical and powerful tool for the comparative empirical analysis of urban segregation. We illustrate our method with a study of ethnic mixing in the Los Angeles metropolitan area.

segregation | spatial statistics | multiscalar analysis

Indices of spatial dissimilarities continue to be at the center of studies of ethnic (and other forms of) segregation in large cities (1–13). However, single-number indices as well as single-scale indices cannot capture the complexity of spatial patterns that arise in actual cities. Only recently, this has led to the exploration of multiscalar approaches (14–23), including multidimensional, matrix indices (24). We present in this paper a framework that allows one to carry out fully multiscalar analyses of spatial dissimilarities and to define multiscalar null models. This framework provides powerful visual representations of segregation phenomena, and it has the capacity to characterize segregation phenomena in a fashion very similar to what may be perceived in actuality.

Our starting point is indeed the common experience that any individual may make of segregation on multiple scales. Consider a city, the population of which comprises k groups. Imagine that an individual explores the city from her home, looking first at her direct neighbors and then at the building next door, and so on, in an ever-expanding visual scale. She will gradually meet the whole city, but at every step, the proportions of the groups in the population that she has encountered so far produce a sequence that yields detailed information on the starting point, in multiscalar relationship with the whole city. We show in Methods how this information may be extracted and analyzed mathematically. In Ethnic Segregation in Los Angeles, we illustrate our method by conducting a study on ethnic mixing in the Los Angeles area.

Methods

We formalize the idea of exploring a city at every scale, from every possible starting point. This is a generalization of a method introduced in ref. 23: Given a starting point and an aggregation procedure (e.g., a nearest-neighbor rule), one may sequentially aggregate all individuals in the city. The aggregated sequence then encodes the order in which someone, starting from that particular point, would encounter the city's inhabitants.* If to each individual in the city one associates a characteristic of interest (ethnicity, income, . . .), then one may also compute the statistical distribution of that same characteristic at every "instant"—in terms of population—of the aggregated sequence. At the first point of the sequence, one has the distribution within the starting spatial unit, whereas at the final point, one has the distribution of the entire city, since all individuals have been aggregated. Instead of considering sequences of statistical distributions computed on ever-larger aggregated populations, one may measure the difference between these and that of the reference population in the entire city. Hence, for each starting point, we shall obtain a dissimilarity trajectory that eventually converges to zero. Furthermore, the trajectories will vary from one starting point to another, reflecting the fact that individuals have different experiences of the city according to where they live: The way each trajectory converges to 0 encapsulates, for each starting point, all

information at every scale, from the most local one to the metropolitan one. In this sense, the set of trajectories obtained starting from all points constitutes a fingerprint of the city for the characteristic under consideration (see Fig. 2).

We are interested here in a particular type of distributions, namely, multinomial. That is, we suppose we have k groups in the population, $A_1, A_2, A_3, \ldots, A_k$. To measure the difference between distributions, we use the Kullback–Leibler (KL) divergence.

KL Trajectories. The KL divergence (25) between two multinomial distributions $(q_1, \ldots, q_k)$, $(r_1, \ldots, r_k)$ is defined as

$$d_{KL}(q \mid r) = \sum_{i=1}^{k} q_i \ln \frac{q_i}{r_i}, \qquad [1]$$

with the convention $0 \ln 0 = 0$. The KL divergence quantifies the nonlinear variation in entropy incurred when substituting distribution q for the reference one, r. Note that computing the KL divergence for each unit and then averaging over all units (weighting each unit by its size) leads to the H index (8)—which is the so-called KL information (26). One may plot the value of the KL divergence for each spatial unit in a region. See, for instance, Fig. 1, where we have done so for the Los Angeles area (at the tract level).

Building upon the idea of trajectories previously introduced, we define, for every starting point in the city, its KL trajectory. This is obtained by computing the KL divergence between the distribution on the aggregated population and the reference population. As the aggregation process relies

---

Significance

We develop a method of measuring segregation which captures the multidimensional nature of mixing in metropolitan areas. The use of trajectory convergence analysis provides a flexible way for capturing change across all scales from small spatial units and how the rate of convergence to the citywide average modifies over space. Thus, the method provides an analysis of how far, in spatial terms, any individual or neighborhood is from the citywide multigroup distribution. We use this method to investigate ethnic mixing in the Southern California metropolitan region. The results illustrate excellent visual measures of the patterns of mixing across urban space, and the graphical trajectories reveal the spatial speed at which the process of convergence takes place.

---

*In practice, one is rarely dealing with data at the individual level, so that one is in fact subsampling the original sequence at certain ranks only—those corresponding to the minimal aggregation level in the data (blocks, tracts, . . .). Let us remark here that when working on aggregated data, results may be sensitive both to the scale and the definition of the spatial units. However, we believe that census data at tract or block level is sufficiently "honest" and reasonably fine-grained to guarantee a good performance of the method.
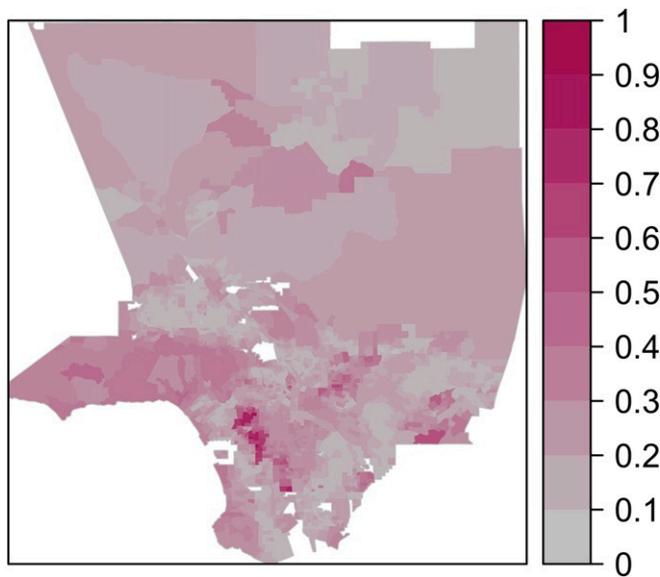
Fig. 1. A map of the Los Angeles area with tracts colored according to the KL divergence between their ethnic composition and that of the whole region. Darker colors represent greater divergence. Values have been normalized by their maximum, so that they scale from 0 to 1. (Four ethnic groups are considered: Whites, Hispanics, Asians, and Blacks. Data are from the 2010 US Census.)

on a spatial nearest-neighbor rule, a KL trajectory is moving from the most local scale through to the whole region. It represents at each step the divergence between the distribution in the population met so far and the distribution in the full region. Mathematically, if $q^{i,1:u}$ is the statistical distribution of the population in the u closest units to i, the KL trajectory associated to unit i is simply the sequence

$$d_{KL}\left(q^{i,1:u} \mid r\right), \quad 1 \leq u \leq U,$$ [2]

where r is the distribution in the full population and U is the number of spatial units in the city.

By construction, every KL trajectory eventually converges to 0 (Fig. 2). However, trajectories differ widely from one starting point to another. Some converge quickly to 0, corresponding to areas where even relatively small aggregates present a reasonably good picture of the whole city. Other trajectories, on the contrary, converge very slowly to 0, corresponding to
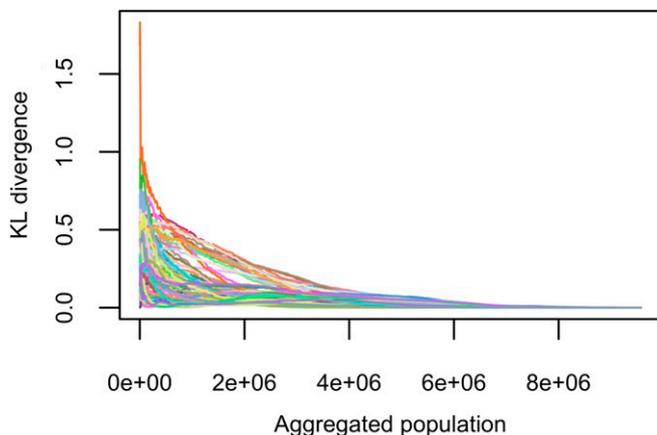
Fig. 2. Trajectories for KL divergence, starting from some (10%) of the tracts in the Los Angeles area. Abscissae are expressed in terms of the aggregated population around the starting point. All trajectories converge to 0, but some do so quickly, while others reach 0 only at much larger scales.
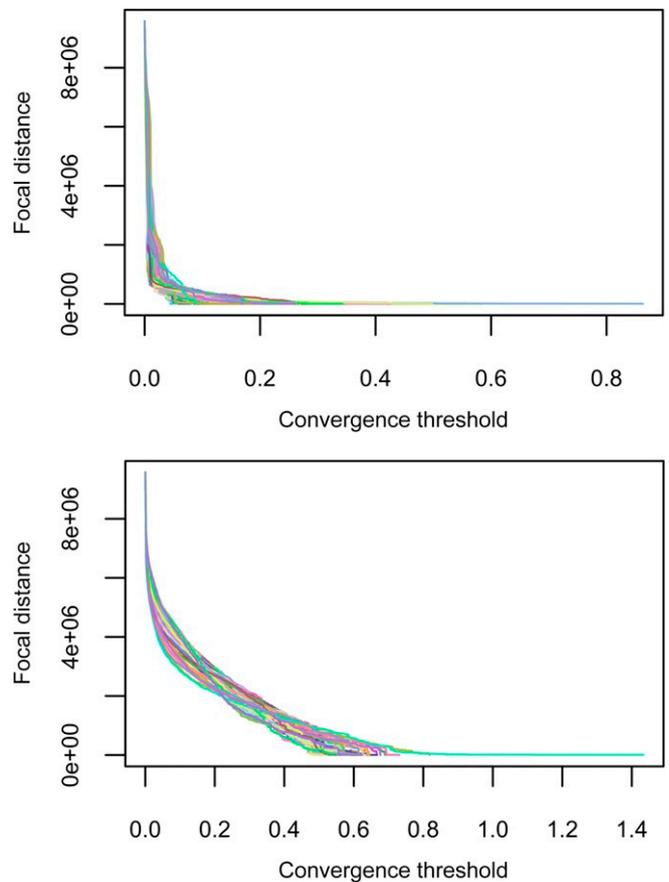
Fig. 3. Focal distance curves, starting from some of the tracts in the Los Angeles area. Ordinates are focal distances expressed in terms of the aggregated population around the starting point. They represent aggregation sizes $f_i(\delta)$ at which the corresponding KL trajectories (Fig. 2) converge, for each value of the convergence threshold δ. We give exemplars of two types of behavior. (*Upper*) Sharply decreasing curves correspond to points with short focal distances, even for small values of δ. (*Lower*) Slower-decreasing curves correspond to points with long focal distances, even for larger values of δ. The distortion coefficient at a given point in the city is then defined (Eq. **4**) to be proportional to the area under the corresponding focal distance curve: The higher the curve, the larger the coefficient.

areas where segregation effects build up across scale and accumulate far beyond the local, tract level.

Focal Distances. We are interested in quantifying the convergence of each trajectory, and the speed at which this occurs, as a mark of segregation of the corresponding starting point. If one fixes a convergence threshold δ, it is then possible to compute, for each trajectory, the instant when it enters (and remains thereafter) in the interval $[0, \delta]$. This instant is the size of the population that one needs to aggregate, for a given starting point, to have a distribution of groups that remains within KL divergence δ of the reference distribution. In other terms, this is, for a given starting point, how far one needs to go in the aggregation process to see (with precision δ) the city's population as it is in its globality—the distance one needs to cover to get a relatively clear picture of the city. We call this convergence size the focal distance. Formally, the focal distance at point i, with precision δ, is defined as

$$f_i(\delta) = \inf_u \left\{ n_{i,1:u} \mid \forall \bar{u} \geq u, d_{KL}\left(q^{i,1:\bar{u}} \mid r\right) \leq \delta \right\},$$ [3]

where $n_{i,1:u}$ is the size of the population in the u closest units to unit i.

There is, of course, some arbitrariness in the choice of the convergence threshold δ, but this may be circumvented by considering all its possible values. For any starting point i and for $\delta = 0$, convergence occurs only at the very end of the KL trajectory, so that $f_i(\delta) = N$ (N is the total population).
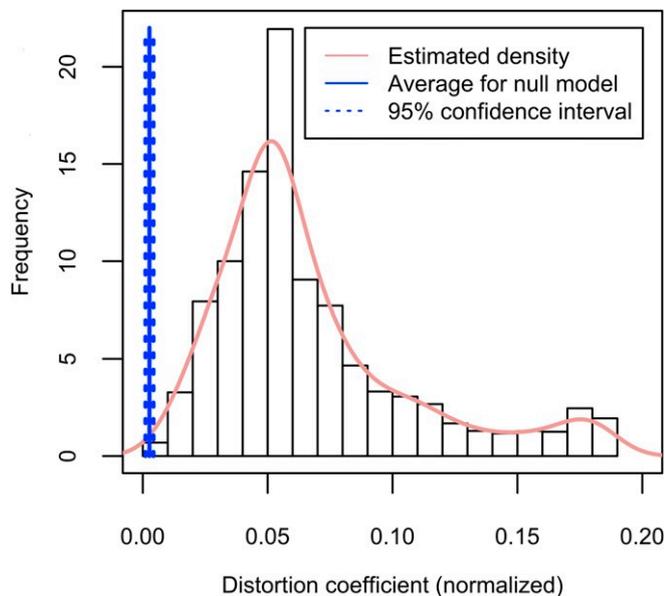
Fig. 4. Histogram and estimated density (pink solid line) of normalized distortion coefficients in Los Angeles area. The solid blue vertical line corresponds to the average value of distortion coefficients in a null model obtained by random (spatial) permutations. The dashed line indicates a 95% confidence interval around the average in the null model.

When δ is large (typically, larger than the maximum of the KL trajectory), convergence occurs immediately so that $f_i(\delta) = 0$. One may then plot the variation of $f_i(\delta)$ as δ increases (Fig. 3). The higher the curve representing $f_i(\delta)$, the longer the focal distances for i, even at large values of δ (i.e., when convergence is easier). This helps identifying points in a city where spatial dissimilarities accumulate on multiple scales to create veritable "hotspots" of segregation. From these points, what one perceives of the city is very much altered, even on large scales, compared with what the city looks like in actuality. We formalize this concept into what we term distortion coefficients.

Distortion Coefficients. We have defined for each point i in the city a trajectory of focal distances $f_i(\delta)$ with $0 \leq \delta \leq \delta_{i,\max}$, where $\delta_{i,\max}$ is the maximum value of the KL trajectory associated to unit i. Summing up focal distances
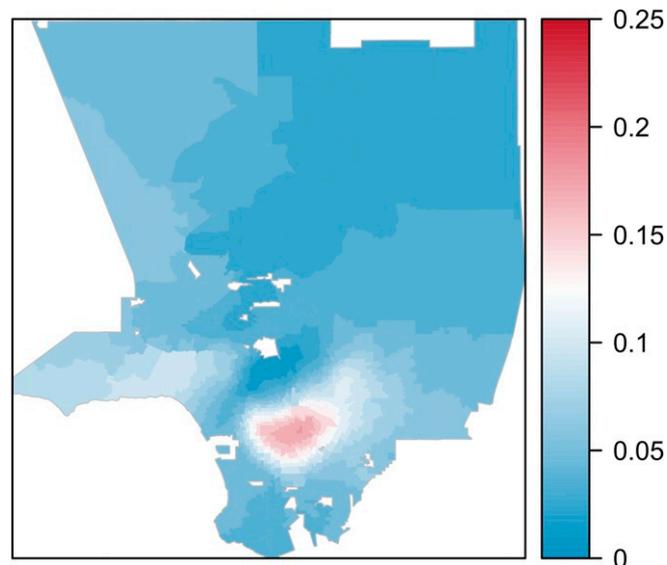
Fig. 5. A map of the Los Angeles area with tracts colored according to their distortion coefficients for the ethnic composition of the population.
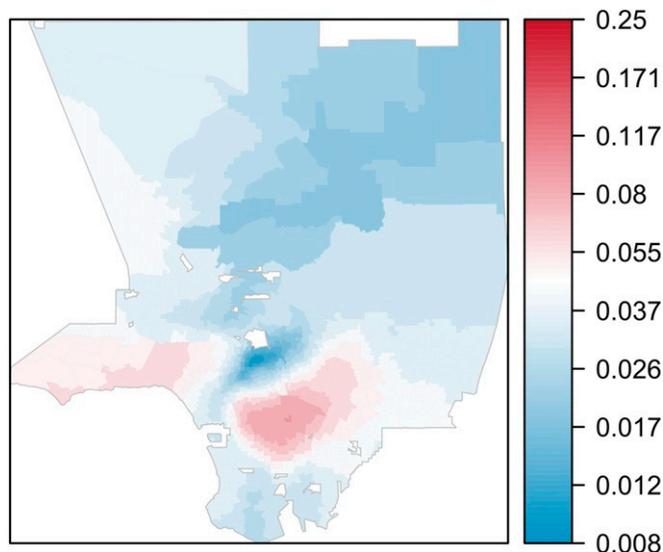
Fig. 6. Same as Fig. 5 but with a logarithmic color scale, allowing one to see finer details in the variation of distortion coefficients across space.

$f_i(\delta)$ for all δ gives a measure of how distorted the perception of the city is, from point i. Formally, we are simply integrating the focal distance curves—that is, we define the distortion coefficient of unit i as

$$\Delta_i = \int_0^{\delta_{i,\max}} f_i(\delta) \, d\delta. \tag{4}$$

To enable comparisons from one variable to another, or from one city to another, distortion coefficients should be independent of the size of the city and also of the average distribution in the city. With this in mind, we compute the normalized distortion coefficients $\tilde{\Delta}_i = \Delta_i / \mathcal{N}$, where $\mathcal{N}$ is a normalizing constant. In practice, $\mathcal{N}$ is chosen as the maximum distortion coefficient in a theoretical extreme case of segregation. Theoretically, the maximal-segregation distortion coefficient is achieved when sorting the k groups into k ghettos, ordered by sizes, and then computing the coefficient for the most isolated person in the smallest group. This person would first meet all of the individuals of his own group, then all those of the second-most infrequent group, and so on, until having seen the entire population of the city. The normalized distortion coefficients $\tilde{\Delta}_i$ take values between 0 and 1 and express the levels of distortion as a fraction of the perspective one has from the theoretical maximally segregated unit (for a given population size and for given group proportions).

To apprehend the meaning of the distortion coefficients, we generate several simple, synthetic configurations. A complete description of these experimental results is provided in *SI Appendix*. It shows how distortion coefficients integrate information across all scales and how this method captures hotspots of multiscalar segregation, as well as fine details of how multiscalar segregation varies across space, from points with low distortion to points with high distortion.

Ethnic Segregation in Los Angeles

As an example, we studied the ethnic distribution in Los Angeles. We worked with 2010 US Census data and analyzed the four largest groups: Whites, Hispanics, Asians, and Blacks. For each tract in the city, the KL divergence between the multigroup distribution inside the tract and the multigroup distribution in the whole city is shown in Fig. 1. We then built the full set of KL trajectories[†] (Fig. 2) and derived the corresponding focal distance curves (Fig. 3). Integrating these curves as in Eq. **4** and normalizing according to the equivalent completely segregated

---

[†]The data were available at tract level, so all trajectories—KL divergence and focal distances—were computed on the aggregated units, by using a nearest-neighbor rule on the centroids as aggregation procedure. Ties were dealt with by using random draws.

four-group population, we obtained the distortion coefficients for each tract, as shown in Fig. 4.

First, let us note the order of magnitude of the distortion coefficients. These reach 0.15–0.20—that is, 15–20% of the maximal value one would obtain by completely separating the four groups into four "ghettos." For comparison, a similar study carried out on different groups in large European cities led to values of only half a percent or a few percent (27). We may thus conclude that ethnic segregation in Los Angeles reaches moderately high levels across mesoscopic and macroscopic scales, not just at the tract level. This conclusion should, however, be considered with caution, since we still need to apply our method to other real data, to get a full perspective of what a highly segregated city should provide as distortion coefficients.

Second, the frequency distribution is particularly informative about spatial heterogeneity. Indeed, if the distribution is narrow and highly peaked at its average, then there is a high level of spatial homogeneity in the city. On the contrary, a broad distribution signals spatial heterogeneity, with multiscalar perceptions of the city varying greatly from one area to another. Frequency distributions may also exhibit more complex features, such as bimodality, suggesting the existence of subsets of segregated areas. In our case, the distribution is markedly skewed, with its peak at 6% of the theoretical maximally segregated value and a secondary mode close to the 20% mark.

Note that finer details and fluctuations of distortion coefficients may be examined by switching to a logarithmic scale, as shown in Fig. 6. The logged representation is particularly interesting when distortion coefficients have a heavy-tailed distribution and span several orders of magnitude. While a linear scale highlights the tail of the distribution and the atypical, highly distorted units, it flattens the rest of the values, for which a logged map provides a better insight. Here, this allows us to access the secondary structure of multiscalar segregation in Los Angeles, with areas like Malibu in the west exhibiting notable levels. We also note that the tracts with the highest distortion and the ones with the lowest distortion appear to both be tightly grouped in different areas: There is a distinct group of contiguous darker blue tracts and a distinct group of contiguous darker red tracts. This is another sign of the level of spatial effects in ethnic segregation in Los Angeles.

Furthermore, one may perform empirical statistical tests and check how much the given city configuration differs from a completely random one. The distribution under the null hypothesis may be easily generated by performing a large number of random permutations and then computing the associated distortion coefficients, their mean value, and a 95% confidence interval around it. According to Fig. 4, all distortion coefficients are greater than the right limit of the confidence interval for the null model, which means that the hypothesis of a completely random distribution of the four communities in Los Angeles is rejected. There is a clustered structure in the data, which means there is segregation, and the maps in Figs. 5 and 6 allow one to identify the most segregated and the most integrated areas.

Conclusion and Perspectives

The method that we have introduced in this paper offers powerful analytical and visual tools to study multigroup segregation in large urban areas.

We have shown here only a few of its capacities. Defining focal distances and distortion coefficients from the convergence of KL trajectories, we have focused on only one type of information that may be extracted from these trajectories. Indeed, other scales of interest may be obtained. Think, for instance, of the scale at which a given KL trajectory attains its maximum—this is the scale of the bespoke neighborhood that is maximally different from the city. On a more theoretical side, we note that a null model may be defined as was done in Ethnic Segregation in Los Angeles: by simply considering random permutations. Establishing the theoretical properties of this null model is not straightforward, as the trajectories obtained when moving from distributions to KL divergence are not readily modeled. In a simpler framework (23), when working with single-group proportions, one may approximate the corresponding trajectories by generalized Brownian bridges (28, 29) and transform the problem into a first-passage one for Brownian motion (30). The general version presented here does not seem amenable to the same techniques and poses an interesting mathematical challenge for future research.

Finally, we believe that this method may open new perspectives on the mechanisms driving segregation, in a well-formalized mathematical framework. Indeed, this technique can be further combined either with theoretical model-based simulations or with a broader empirical context (multiple cities, variables, or time instants), and thus further contribute into a fine understanding of segregation, of how it changes in time or across space, how microscopic and/or macroscopic interventions on the dynamics or on the spatial distribution may lead to significant changes in the trajectories and in the distortion coefficients, where these changes are the most prominent, and how they propagate over space.

1. D. O. Cowgill, M. S. Cowgill, An index of segregation based on block statistics. *Am. Sociol. Rev.* **16**, 825–831 (1951).
2. O. D. Duncan, B. Duncan, A methodological analysis of segregation indexes. *Am. Sociol. Rev.* **20**, 210–217 (1955).
3. R. Farley, K. E. Taeuber, Population trends and residential segregation since 1960. *Science* **159**, 953–956 (1968).
4. C. F. Cortese, R. F. Falk, J. K. Cohen, Further considerations on the methodological analysis of segregation indices. *Am. Sociol. Rev.* **41**, 630–637, (1976).
5. M. J. White, The measurement of spatial segregation. *Am. J. Sociol.* **88**, 1008–1018 (1983).
6. D. R. James, K. E. Taeuber, Measures of segregation. *Sociol. Methodol.* **15**, 1–32 (1985).
7. D. S. Massey, N. A. Denton, The dimensions of residential segregation. *Soc. Forces* **67**, 281–315 (1988).
8. S. F. Reardon, G. Firebaugh, Measures of multigroup segregation. *Sociol. Methodol.* **32**, 33–67 (2002).
9. L. A. Brown, S.-Y. Chung, Spatial segregation, segregation indices and the geographical perspective. *Popul. Space Place* **12**, 125–143 (2006).
10. F. F. Feitosa, G. Camara, A. M. V. Monteiro, T. Koschitzki, M. P. S. Silva, Global and local spatial indices of urban segregation. *Int. J. Geograph. Inf. Sci.* **21**, 299–323 (2007).
11. M. Fossett, *New Methods for Measuring and Analyzing Segregation* (Springer Series on Demographic Methods and Population Analysis, Springer, New York, 2017), Vol. 42.

12. M. Olteanu, A. Hazan, M. Cottrell, J. Randon-Furling. Multidimensional urban segregation - toward a neural network measure. arXiv:1705.03213 (9 May 2017).
13. R. Harris, R. Johnston, Measuring and modelling segregation–new concepts, new methods and new data. *Environ. Plann. B Urban Anal. City Sci.* **45**, 999–1002 (2018).
14. G. Leckie, R. Pillinger, K. Jones, H. Goldstein, Multilevel modeling of social segregation. *J. Educ. Behav. Stat.* **37**, 3–30 (2012).
15. J. Östh, W. A. V. Clark, B. Malmberg, Measuring the scale of segregation using k-nearest neighbor aggregates. *Geograph. Anal.* **47**, 34–49 (2015).
16. R. Louf, M. Barthelemy, Patterns of residential segregation. *PLoS One* **11**, e0157476 (2016).
17. W. A. V. Clark, E. Andersson, J. Östh, B. Malmberg, A multiscalar analysis of neighborhood composition in Los Angeles, 2000–2010: A location-based approach to segregation and diversity. *Ann. Assoc. Am. Geogr.* **105**, 1260–1284 (2015).
18. K. Jones, R. Johnston, D. Manley, D. Owen, C. Charlton, Ethnic residential segregation: A multilevel, multigroup, multiscale approach exemplified by London in 2011. *Demography* **52**, 1995–2019 (2015).
19. D. Manley, R. Johnston, K. Jones, D. Owen, Macro-, meso-and microscale segregation: Modeling changing ethnic residential patterns in Auckland, New Zealand, 2001–2013. *Ann. Assoc. Am. Geogr.* **105**, 951–967 (2015).
20. G. Leckie, H. Goldstein, A multilevel modelling approach to measuring changing patterns of ethnic composition and segregation among London secondary schools, 2001–2010. *J. R. Stat. Soc. Ser. A* **178**, 405–424 (2015).

SOCIAL SCIENCES

21. R. Harris, D. Owen, Implementing a multilevel index of dissimilarity in R with a case study of the changing scales of residential ethnic segregation in England and Wales. *Environ. Plann. B Urban Anal. City Sci.* **45**, 1003–1021 (2017).

22. B. M. Adrestani, D. O'Sullivan, P. Davis, A multi-scaled agent-based model of residential segregation applied to a real metropolitan area. *Comput. Environ. Urban Syst.* **69**, 1–16 (2017).

23. J. Randon-Furling, M. Olteanu, A. Lucquiaud, From urban segregation to spatial structure detection. *Environ. Plann. B Urban Analytics. City Sci.*, 10.1177/2399808318797129 (2018).

24. C. Ballester, M. Vorsatz, Random walk-based segregation measures. *Rev. Econ. Stat.* **96**, 383–401 (2014).

25. S. Kullback, R. A. Leibler, On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).

26. P. S. Chodrow, Structure and information in spatial segregation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11591–11596 (2017).

27. W. Clark, M. Olteanu, J. Randon-Furling, "Segregation, focal distances & neighbourhood scales for EU and non-EU migrants in European cities" in D4I Final Workshop (EU Joint Research Center, Brussels, Belgium, 2018).

28. B. Rosén, Limit theorems for sampling from finite populations. *Arkiv Matematik* **5**, 383–424 (1964).

29. P. K. Sen, Finite population sampling and weak convergence to a Brownian bridge. *Sankhyā Indian J Stat Ser A* **34**, 85–90 (1972).

30. P. Salminen, M. Yor, On hitting times of affine boundaries by reflecting Brownian motion and Bessel processes. *Periodica Mathematica Hungarica* **62**, 75–101 (2011).

Olteanu et al.