Geographic Scales of Residential Segregation in English Cities

Tian Lan[a], Jens Kandt [b], Paul Longley [a*]

[a] *Department of Geography, University College London, London, UK;* [b] *The Bartlett Centre for Advanced Spatial Analysis, Bartlett Faculty of the Built Environment, University College London, London, UK*

Abstract

The barriers to social integration posed by ethnic residential segregation are currently receiving renewed attention in Great Britain. A common characteristic of past studies of ethnic segregation in Britain is reliance upon aggregated Census data, raising potential issues of ecological fallacy. In this study, we address this challenge by using novel individual-level Consumer Register data for the UK to calculate an entropy-based spatial segregation index. We measure changes in segregation over twenty years and examine the impact of geographic scales upon observed levels of segregation in five policy relevant case study areas. Our results and findings can be used to improve the evidence base on segregation dynamics in the United Kingdom and have methodological implications for the future study of the phenomenon.

Keywords: ethnic segregation; consumer registers; social integration; spatial analysis

1. Introduction

The integration of ethnically diverse communities has provided a recurring focus for policy analysis in Britain, ever since the advent of large scale migration from Commonwealth nations and colonies of the former British Empire to supplement the domestic labour market in the mid-Twentieth Century (Simpson, 2004). Post 2004, the free movement of labour within the enlarged single market of the European Union (EU) led to the immigration of an estimated 1.5 million[1] new UK residents from Eastern European states over a ten-year period. Together, these changes have made the ethnic diversity of local populations a recurring focus of interest (Catney, 2016). At the same time, flashpoints such as periodic disturbances in some English towns have triggered public and political debates around social issues such as residential segregation in ethnically diverse areas (Cantle, 2001; Casey, 2016; Phillips, 2005).

*Contact Paul Longley p.longley@ucl.ac.uk

[1]https://www.ons.gov.uk/releases/noteonthedifferencebetweennationalinsurancenumberregistrationsandtheestimateoflongterminternationalmigration

Following the Casey Review (Casey, 2016) and other evidence, the UK government has published a Green Paper (Ministry of Housing, Communities & Local Government, 2018) in order to seek opinions on challenges to community integration from individuals, communities, and organisations. The Paper identifies residential segregation as one of seven potential barriers to integration; it sits alongside: lack of English Language proficiency, labour market disadvantage, educational attainment, level and pace of migration, lack of meaningful social mixing, and issues arising from religious and cultural practices. Dedicated policy focus has been brought to Black, Asian and Minority Ethnic (BAME) groups, which are understood to be more segregated than other groups. The Green Paper proposes localised implementation of a national framework of policy interventions, to be trialled initially in the "Integration Areas" of five English local authorities. The five areas were assigned this policy status because of the particular integration challenges that they face and the accumulated experience of past initiatives, however, the Green Paper that heralded their introduction acknowledges that many integration metrics are unavailable at local level, or not updated with sufficient frequency to measure the impacts of community integration strategies.

In this paper, we therefore respond to the challenging task of improving the information used in evidence-based policy formulation. In the absence of any administrative name-and-address registration data in the UK, we utilise a novel linked Consumer Register (Lansley et al., 2019) grounded at the level of the adult individual, in order to measure residential segregation aspects of community integration. We draw upon the results of address-level estimation of the ethnicity of residents, using annual Consumer Registers for the period 1997-2016. Our case studies are developed for the five pilot Integration Areas proposed in the Green Paper: Blackburn with Darwen, Bradford, Peterborough, Walsall, and Waltham Forest. We examine the temporal changes in segregation in the five case study areas, as well as the effects of geographic scale upon recorded segregation levels. The paper is structured as follows: we will first set out some relevant debates from the British literature on segregation; second, we will discuss the drawbacks of the aspatial segregation measures and the current spatial segregation measures in the literature; we will then describe the data hardening and pre-processing effort and elaborate on how we formulate the individual level spatial segregation index using Consumer Registers; and finally, we will present the substantive results of the investigation and assess their implications for policy.

2. Scale and spatial segregation

In the contemporary British context, Johnston et al. (2002) have contended that Bradford, Leicester and Oldham manifest American-style minority enclaves, although Peach (2009) has pointed out the seemingly arbitrary thresholds that this work used to define minority "ghettos". Finney and Simpson (2009) have challenged related popular myths about ethnicity and migration, warning that statistics are used in misleading ways to support political arguments. Iceland and Mateos (2011) have compared ethnic residential segregation between Great Britain and the United States and found that black communities in Britain were less segregated than in the United States, while the opposite held for some Asian communities. Catney (2015) has explored national-level changes in ethnic segregation over the 2001-11 intercensal period, and identified increased residential mixing between the White British majority and all other ethnic groups, a finding which is contrary to the assumption that accelerating ethnic diversity is associated with increasing residential segregation. Cantle and Kaufmann (2016) have concurred that some UK ethnic minorities have partially vacated the neighbourhoods in which they first became established, but also contend that, at the same time, segregation between White British and the sum total of all other ethnic minorities, has increased in some towns. Lan et al. (2018) have presented annual small area segregation measures for the ethnic groups defined in the 2011 England and Wales Census and suggested that over-all levels of residential segregation have decreased over recent decades. Others examined segregation along religious lines (Gale, 2013) and in relation to the provision of education (Harris, 2017).

The research findings all share reliance upon aggregated data, typically pertaining to small area census geographies. Reliance upon the UK Census of Population raises important methodological issues. First, although the building blocks of small area census geographies are designed with the relative homogeneity of some population characteristics in mind, the within-zone distributions of ethnic groups are not revealed. This restricts the scale range over which ethnic concentrations can be detected, and potentially renders the results of their analysis vulnerable to scale and aggregation effects. Segregation measures based upon census zones implicitly assumes correspondence of zones with spatial distributions of members of ethnic minorities. In addition, the aspatial nature of segregation measures in previous studies may be criticised for not accommodating local distributions that traverse boundaries, and for assuming uniformity within zone distributions of ethnic groups. The incomplete capture of spatial proximity effects renders analysis vulnerable to the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984) and checkboard problem (Reardon & O'Sullivan, 2004).

To overcome these limitations, spatial segregation measurements have been developed and applied in the international literature. Wong (1999) contributes a novel spatial segregation index that uses standard deviational ellipses to reflect the correlation among ethnic groups. This work is extended in Wong (2002), whereby the analysis uses multiple aspatial segregation measures that incorporate spatial interaction measures across areal unit boundaries by taking into account shared boundary length and geometric considerations. Reardon and O'Sullivan (2004) propose several spatial segregation measurements and compare them with selected aspatial counterparts. O'Sullivan and Wong (2007) use kernel density estimates to accommodate probable within zone heterogeneity in ethnic composition. Similar kernel density surface estimates are used in other spatial segregation studies to incorporae variability in household incomes (Feitosa et al., 2007; Monkkonen & Zhang, 2014). Östh et al. (2015), and Hennerdal and Nielsen (2017) develop a k-nearest neighbour based method to measure exposure dimension among different ethnic groups. These spatial segregation measurements have not yet avoided the MAUP issue completely, which is mostly limited by the availability of ethnicity data at disaggregated level.

The issues inherent in using aggregate data become apparent when they are used to measure segregation across multiple geographic scales. Segregation should be conceived as a multi-scale phenomenon, and measures of it are scale dependent: but where such analysis is founded upon aggregate data, this dependence can only be evaluated over a limited range of standard geographies, such as UK Census Output Areas (OAs), Super Output Areas, Wards, or Districts (Cantle & Kaufmann, 2016; Harris, 2017; Simpson, 2007). Reardon et al. (2008) seek to accommodate fixed scale effects using a kernel based approach that improvises population counts on the assumption of continuous variation between zone centroids – an assumption that is strained or broken by the variegated neighbourhood geographies of many settlements. Similar work (Lee et al., 2008) reveals patterns of residential segregation at different scales for the 100 largest U.S. cities. Catney (2018) uses a similar spatial weighting method to examine the scale effect in England and Wales: her findings indicate that ethnic groups are more segregated at localised neighbourhood scales and less segregated (but to differing degrees) across more extensive regional scales. Further limitations of census-based analyses arise when examining residential segregation trends over time, since census-based analysis is restricted to ten-yearly intercensal periods, and low-level zonal geographies may change between censuses. For example, between the 2001 and 2011 UK Census, 4,354 of 175,434 OAs (2.4%) in England and Wales were either split or merged. Where boundary change occurs, recorded change in segregation levels may be more apparent than real (Simpson, 2007).

Our principal contribution here is to estimate ethnicity for every adult individual in a series of Consumer Registers that have near total population coverage for the UK. Our novel approach infers ethnicity from individuals' names as recorded in these annualised registers. We: (1) make use of the annual updates to record the dynamics of change throughout intercensal periods; (2) calculate address level spatial segregation measures for our case study areas and explore annual changes of spatial segregation measurements; and (3) examine the effects of geographic scale upon our results.

3. Data and method

We develop address level ethnicity information of individuals from two data assets: annual Consumer Registers for 1997-2016 and Ordnance Survey AddressBase.[2] The first of these data assets is held securely by the Consumer Data Research Centre (CDRC),[3] and records individual surname, forename, residential address and postcode, with near universal population coverage of the entire UK (Lansley et al., 2019). Researcher access to the datasets is available at three UK secure labs via the CDRC secure service subject to project approval requirements. Consumer Registers are compiled by third-party data companies from disparate data sources: they comprise full versions of annual Electoral Registers for 1997-2003; and for 2003-2016, they are composed of both the public Electoral Registers and various consumer data sources, which are employed to include the population who "opted-out" of the publicly available electoral roll. Ordnance Survey AddressBase Premium is the most comprehensive available register of the 28+ million postal addresses in Great Britain over this period: it is linked to the Royal Mail Postcode Address File (PAF) and includes precise geographic coordinates of each address.

[2] https://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-products.html
[3] https://www.cdrc.ac.uk

3.1. Data 'hardening'

Although more detailed, disaggregate and more frequently updated than the conventional sources used in segregation studies, Big Data sources such as Consumer Registers are not of known provenance. Such data sources have been described as 'soft' by (Goodchild, 2013), and here we summarise the procedures of data 'hardening' used to pre-process and clean the data in order to establish and confirm their fitness for purpose in segregation analysis. We used an extensive global names dictionary (O'Brien & Longley, 2018) and standardised addresses through linkage using AddressBase using the fuzzy string match algorithm developed by Lansley et al. (2019). Residential address changes and population counts from Consumer Registers were further validated with external aggregated sources—specifically the 2011 UK Census, successive Mid-year Population Estimates from the UK Office for National Statistics (ONS), and Land Registry records of individual property sales. Table 1 presents the over-all correspondence between Consumer Register counts of adults and ONS Mid-year Population Estimates. The adult population captured in Consumer Registers broadly correspond to the numbers of adults from the ONS Mid-year Population Estimates, albeit that the ONS source is also deemed likely to be increasingly inaccurate with time elapsed since the most recent (2011) Census. Our view is that the greatest source of bias in the later registers is likely attributable to failure to replace all of the individuals who 'opt out' of inclusion in the public Electoral Register with consumer data sources. The heavy reliance upon the public Electoral Register is likely to bias inclusion towards individuals enfranchised to vote in local, national or EU elections, but it should be noted that we do not calculate segregation of any non-voter ethnicities from 1997 to 2003 in our analysis, since the linked Consumer Register only captures registered voters in these years. Linkage to AddressBase is used to validate addresses and to assign precise geographic coordinates to every individual in the Consumer Registers.

Table 1: Comparison of the adult population (17 and plus) between Consumer Registers and Mid-year Population Estimate (MYPE) 1997-2017 (Source: Authors' calculation and published ONS statistics)

Year	Consumer Registers	Mid-year Population Estimates (MYPE)	% of MYPE
1997	45,128,535	45,560,428	99.1%
1998	46,100,649	45,739,580	100.8%
1999	46,207,147	45,951,062	100.6%
2000	46,302,578	46,200,136	100.2%
2001	46,542,177	46,488,614	100.1%
2002	46,561,516	46,809,778	99.5%

2003	46,982,475	47,113,733	99.7%
2004	47,218,924	47,455,211	99.5%
2005	47,234,395	47,949,873	98.5%
2006	47,269,670	48,371,924	97.7%
2007	47,382,612	48,833,940	97.0%
2008	47,722,362	49,313,815	96.8%
2009	49,181,334	49,717,852	98.9%
2010	49,578,070	50,160,114	98.8%
2011	49,971,711	50,634,451	98.7%
2012	50,578,970	50,952,203	99.3%
2013	50,862,893	51,274,613	99.2%
2014	51,622,350	52,101,602	99.1%
2015	51,637,091	51,687,804	99.9%
2016	52,109,264	52,525,330	99.2%

3.2. Name-based ethnicity inference

We use the forename-surname pairing of each record of Consumer Registers to estimate the most probable ethnicity of their bearer. Ethnicities are ascribed to individuals named on the Consumer Register using outputs obtained from the ONS Virtual Microdata Laboratory as described in Kandt and Longley (2018). These authors describe how such assignment is an error-prone process, particularly for 'hard-to-reach' groups such as Black Caribbeans and individuals of mixed races, or where very common names are shared across multiple groups. However, the ethnicity estimation method is reported to have a success rate of 88% in predicting which of the 12 ethnic categories individuals assigned themselves to when responding to the 2011 Census (Kandt & Longley, 2018). The software is made available to approved research users, free of charge, following successful application to CDRC[4]. Similar name-based ethnicity inference has been used in many other applications (Lan et al., 2018; Lansley & Li, 2018; Petersen et al., 2011).

The software outputs are provided for the following categories used in the 2011 Census: Bangladeshi, Black African, Black Caribbean, Chinese, Indian, Other Asian, Pakistani, White British, White Irish, Other White and Any Other. This categorisation was developed by the ONS for use in the 2011 Census in consultation with the key users of Census data (Office for National Statistics, 2009). Although the Census categories have been criticised for the arbitrary and imprecise definition based on skin colours (e.g. Black British), and for combining diverse groups into "pan-ethnic" classes (e.g. Other White) (Aspinall, 2002;

[4] https://ee.cdrc.ac.uk/

Berthoud, 1998; Simpson, 2004), use of the Ethnicity Estimator outputs requires that we adopt the Census ethnic categorisation in this study.

We compare the estimates for the adult (16+) population with benchmark Census data for 2001 and 2011 (see Table 2 and 3). This reveals strong correspondence, particularly for the Indian, Pakistani, Other White and White British groups. Occurrences of some groups, specifically the Chinese and Black Caribbeans, are underestimated. The White Irish group is over-enumerated, possibly reflecting lack of self-identification with this group in the Census of individuals who are long settled in the UK. Thus, White British and White Irish are combined in our study in view of the inherent ambiguities in self-assignment to these groups and their marginal relevance to segregation debates. Population growth of minority ethnic groups over the 2001 - 2011 period is well reflected in the Consumer Register estimates. We find that the White British population has fallen over the 2001 - 2011 period in all Integration Areas except for Peterborough, while Pakistani, Bangladeshi, Indian, and Black African minority populations increase in size. Significant increase in the "Other White" population is observed in Peterborough and Waltham Forest, probably following 2004 and 2007 European Union enlargement.

Table 2: Comparison of 2011 adult population ethnicity enumerates from Consumer Registers (CR) with the 2011 Census. (Source: authors' calculations and ONS 2011 Census Table DC2101EW)

	Blackburn with Darwen		**Bradford**		**Peterborough**		**Walsall**		**Waltham Forest**	
	CR	**Census**	**CR**	**Census**	**CR**	**Census**	**CR**	**Census**	**CR**	**Census**
Bangladeshi	926	856	7,206	5,696	658	147	2,371	2,844	2,619	3,509
Chinese	171	548	585	1,789	426	701	378	770	1,013	2,197
Indian	9,042	13,710	11,976	10,671	3,702	3,618	13,331	12,828	7,555	7,435
Pakistani	11,272	11,374	60,181	67,690	8,113	7,535	9,049	9,059	20,686	18,765
Other Asian	339	1,161	1,356	5,821	920	2,692	564	3,037	3,882	8,914
Black African	1,009	425	2,698	3,786	1,574	1,634	1,330	1,333	9,832	12,977
Black Caribbean	172	180	1,158	3,138	456	1,020	637	2,643	2,455	15,307
White Other	2,718	2,450	13,514	12,903	14,318	15,752	3,643	3,186	26,135	31,710
White British	62,798	79,612	230,708	273,267	88,407	106,135	147,225	169,854	76,925	79,082
White Irish	5,892	771	17,108	2,444	4,816	1,199	6,424	1,119	6,684	3,737
Any Other	548	2,035	1,466	12,916	796	4,062	455	6,450	3,060	19,498
Total	94,887	113,122	347,956	400,121	124,186	144,495	185,407	213,123	160,846	203,131

Table 3: Comparison of 2001 adult population ethnicity enumerates from Consumer Registers (CR) with the 2011 Census. (Source: authors' calculations and ONS 2001 Census Table ST101)

	Blackburn with Darwen		**Bradford**		**Peterborough**		**Walsall**		**Waltham Forest**	
	CR	**Census**	**CR**	**Census**	**CR**	**Census**	**CR**	**Census**	**CR**	**Census**
Bangladeshi	710	281	5,541	2,862	350	73	1,585	1,453	1,430	1,437
Chinese	132	309	436	1,587	231	818	291	706	788	3,270
Indian	7,063	9,485	10,091	9,352	2,123	2,178	9,896	10,193	5,616	5,834
Pakistani	7,460	7,260	35,964	42,232	3,319	4,226	5,216	5,790	12,362	11,716
Other Asian	191	806	561	1,980	200	655	233	651	1,980	3,669
Black African	678	150	893	793	274	444	263	308	4,823	8,873
Black Caribbean	173	92	1,036	2,555	336	947	462	2,231	2,001	13,697
White Other	1,894	1,155	9,647	6,177	4,940	3,978	2,299	1,426	11,150	11,911
White British	70,627	81,543	248,933	284,149	87,104	105,678	159,380	172,843	96,020	100,717
White Irish	6,206	1,134	17,181	3,316	4,413	1,597	6,154	1,386	8,177	4,743
Any Other	407	591	985	3,347	325	1,180	246	1,443	1,786	5,554
Total	95,541	102,806	331,268	358,350	103,615	121,774	186,025	198,430	146,133	171,421

3.3. Individual level spatial segregation measure

Following Reardon and O'Sullivan (2004), the ethnicity data are used to formulate an individual level spatial segregation measure. We use Theil's (1972) information theory index H to measure the spatial inequality of residential distributions of different ethnic groups. Perceptions of ethnic residential segregation reflect the degree of disparities in ethnic compositions among each individual's local environment. We thus develop the individual level spatial segregation index in two steps: (1) defining bespoke neighbourhoods for each individual; and (2) quantifying the degree of disparities across each individual's ethnic neighbourhood.

3.3.1. Defining bespoke neighbourhoods

Past ethnic segregation studies have been limited in the delineation of neighbourhoods to the size and extent of available geography aggregations, such as UK Census Output Areas. This has been criticised for its lack of social meaning (Logan et al., 2011). In this study, we define bespoke neighbourhoods as circular regions focused upon each individual's address p with a pre-defined radius r (Figure 1). The radius of these neighbourhoods can be adjusted to represent, say, a 500-metre walking distance, or a 1000-metre radial school commute, or a more extensive activity space with a radial distance of 3,500 metres, albeit these are very crude abstractions of individuals' meaningful neighbourhoods. The flexibility makes it possible to explore the scale effect of neighbourhoods by changing the radial bandwidth.

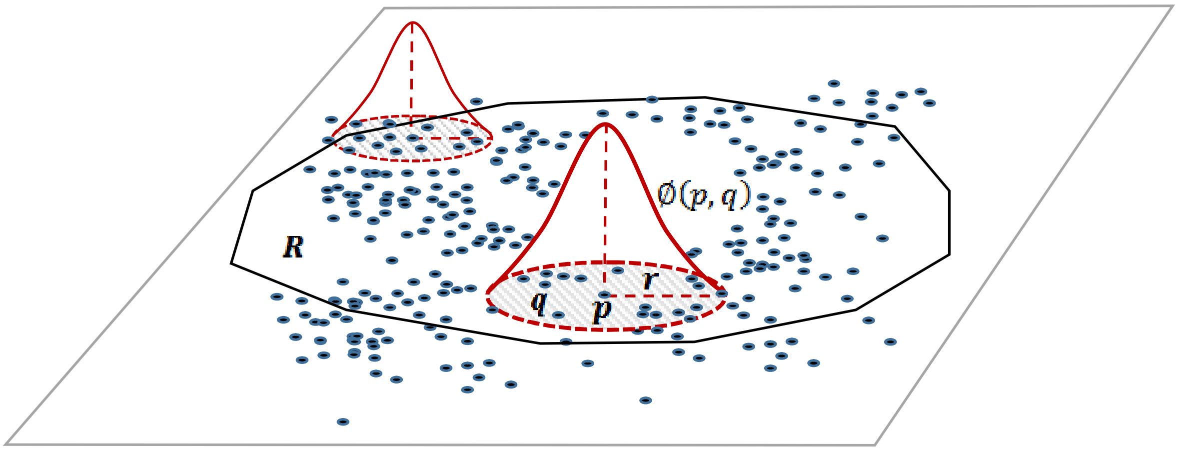

Figure 1. An illustration of bespoke neighbourhoods and weighting function in the study area R

To incorporate spatial proximity, we weight the ethnic composition within each individual's neighbourhood by the distance between the individual and other residents. Equation (1), $\tilde{\pi}_{pm}$ represents the proportion of group m in the neighbourhood around p among the total population in the same neighbourhood, where τ_{qm} and τ_q denote the population density of group m and the population density of all groups at all other locations q, which fall into the radius r from p. We incorporate the distance decay effect using the function $\emptyset(p, q)$ in Equation (1), assuming that nearer residents contribute more to segregation than ones that more distant ones. Different functional forms of distance decay might be posited under different spatial interaction scenarios such as the quadratic kernel used by one of us for school catchment representation (Singleton et al., 2011) or the negative exponential shape for commuting studies or the inverse power function for migration modelling (Longley et al., 2015; O'Kelly & Horner, 2003; Östh et al., 2016). Here, we follow the established practice of previous residential segregation studies (Catney, 2018; Monkkonen & Zhang, 2014; Reardon et al., 2008), by adopting the bounded quadratic form of the distance decay function defined in Equation (2). As such, this decision is based upon choice, convention and compatibility with previous research: other decay functions could be used, but this lies beyond the scope of the present paper.

$$\tilde{\pi}_{pm} = \frac{\int_{q \in R} \tau_{qm} \emptyset(p,q)\, dq}{\int_{q \in R} \tau_q \emptyset(p,q)\, dq} \quad (1)$$

$$\emptyset(p,q) = \begin{cases} \left[1 - \left(\frac{d(p,q)}{r}\right)^2\right]^2, & d(p,q) \le r; \\ 0, & d(p,q) > r. \end{cases} \quad (2)$$

3.3.2. Quantifying degree of disparities

The definition of the individual level spatial segregation index H is given in Equation (3), following the work of Reardon and O'Sullivan (2004). Here, T denotes the total population of the study area; E denotes the overall entropy of all neighbourhoods in the study area; and τ_p is the population density at location p. Equation (4) defines the entropy value around the neighbourhood of p over all M ethnic groups. The entropy-based information theory index H can be interpreted as a function of disparities between a weighted average of within-neighbourhood ethnic diversity among individuals and the over-all ethnic diversity of the entire study area. It thus measures the evenness dimension of residential segregation. Similar to other segregation measurements, larger index values denote higher degrees of segregation with a usual upper limit of one. In the most extreme case, if each neighbourhood is fully occupied by a single ethnic group, the entropy value of each neighbourhood will equal 0, which leads to a completely segregated scenario with an index value of 1.

$$H = 1 - \frac{1}{TE} \int_{p \in R} \tau_p \tilde{E}_p \, dp \quad (3)$$

$$\tilde{E}_p = \sum_{m=1}^{M} \tilde{\pi}_{pm} \log_M \frac{1}{\tilde{\pi}_{pm}} \quad (4)$$

As stated above, we take the five Integration Areas chosen in the Green Paper (Ministry of Housing, Communities & Local Government, 2018) as our case study areas: Blackburn with Darwen, Bradford, Peterborough, Walsall, and Waltham Forest. The selected Integration Areas are five UK local authority districts located within different Government Office Regions in England (see Figure 2). We use subsets of the national Consumer Registers corresponding to the five Integration Areas to calculate the individual level spatial segregation index from 1997 to 2016 at a series of discrete scales ranging from 500 metres to 3,500 metres at 500-metre intervals. In addition, to cope with the edge effects of local authority district boundaries, we set a 4,000-metre buffer around each district boundary. Residents in the buffered areas are only taken into consideration when they are located within neighbourhoods of residents from within the five study areas.

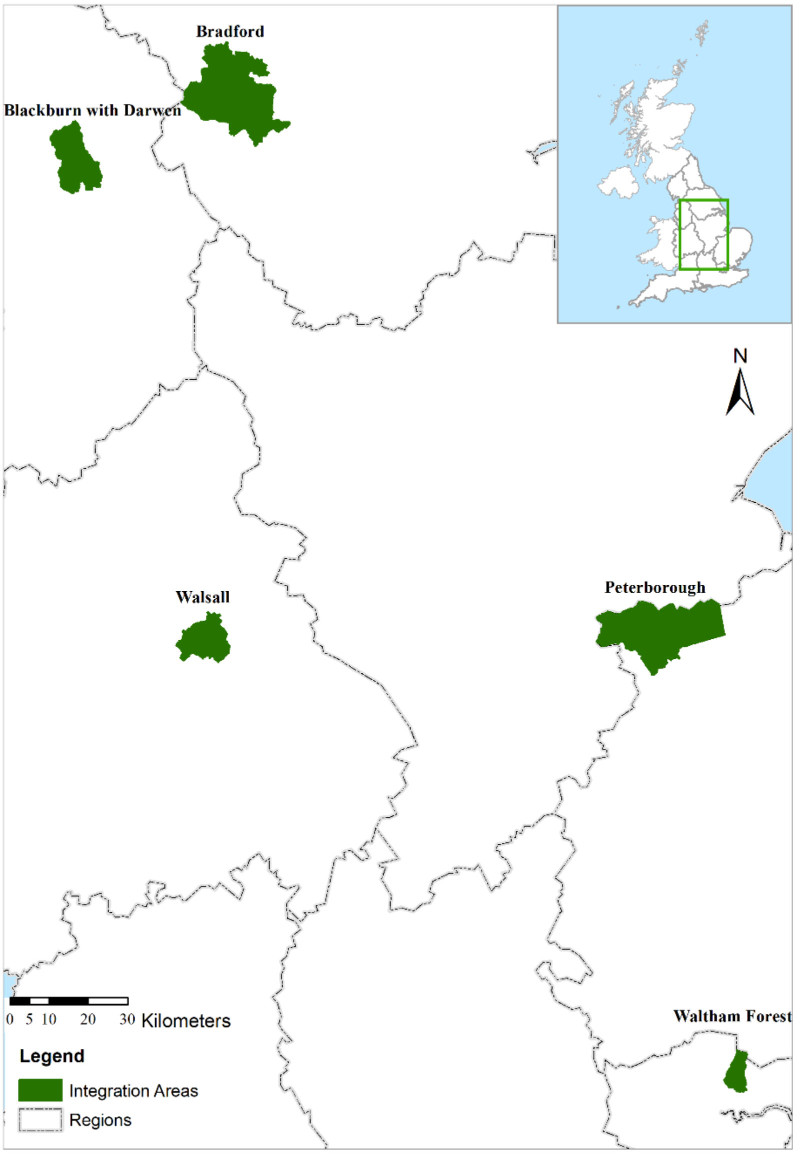

Figure 2. Locations of the five Integration Areas

4. Results

4.1. Changes in segregation, 1997-2016

We first investigate the trend of segregation levels in each of the five Integration Areas over the 20-year study period. We plot the segregation index H for each Integration Area at bandwidths of 500 metres, 1,500 metres, 2,500 metres, and 3,500 metres (Figure 3). As can be seen from Figure 3(A), at the bandwidth of 500 metres, segregation levels of all Integration Areas (except for Waltham Forest) have increased during the first part of the study period before subsequently declining slightly (Waltham Forest also follows this trend of decline). Similar patterns of segregation level changes can be found at the bandwidth of 1500 metres, 2500 metres, and 3500 metres in Figure 3(B), (C), and (D) as well. However, it appears that at the three specified scales, Blackburn with Darwen has become increasingly segregated over the years, which is in contrast to its more granular trend at 500 metres bandwidth in Figure 3(A). Variations of segregation levels in Waltham Forest remain uniformly low—with a standard deviation (SD) of 0.007, while segregation levels of Bradford (SD = 0.017) and Peterborough (SD = 0.015) have declined respectively by 0.044 and 0.045 from 2003 to 2016. The decline should be considered as substantively meaningful changes (the threshold of 0.05) in the temporal dimension as measured using the information theory index H (Reardon & Yun, 2001).

It is quite noticeable in the charts that Waltham Forest stands out amongst the Integration Areas as having the lowest segregation level throughout the study period. This in no small part reflects the continuous nature of its urban development, in contrast to the areas of farmland and open space in the other areas. Waltham Forest also has an evener population distribution across its entire area as well as high ethnic diversity in its population mix. It should also be noted in Figure 3(D) that segregation index values of Waltham Forest in 1997, 1998, 1999, and 2001 are negative at the scale of 3,500 metres. It is mathematically possible that a spatial entropy segregation index takes a negative value, indicative of "hyper-integration" (Reardon & O'Sullivan, 2004) in which the neighbourhood of an individual on average would be more ethnically diverse than the entire region of the population. However, they have also pointed out that this phenomenon has not been empirically observed in their case studies on U.S. cities in another study (Reardon et al., 2009). In addition, the rank order of the segregation levels in the five Integration Areas has barely changed for the past years at the lower scales of 500 metres and 1,500 metres (Figure 3); while at the scales of 2,500 metres and 3,500 metres, several crossovers have been observed among Bradford, Blackburn with Darwen, and Walsall. Bradford has been the most segregated area across the four selected scales all the time, except for early years before 2010, when it had been surpassed by Walsall at the scale of 3,500 metres. This shows the relative changes responding to the different geographic scales vary among these Integration Areas.

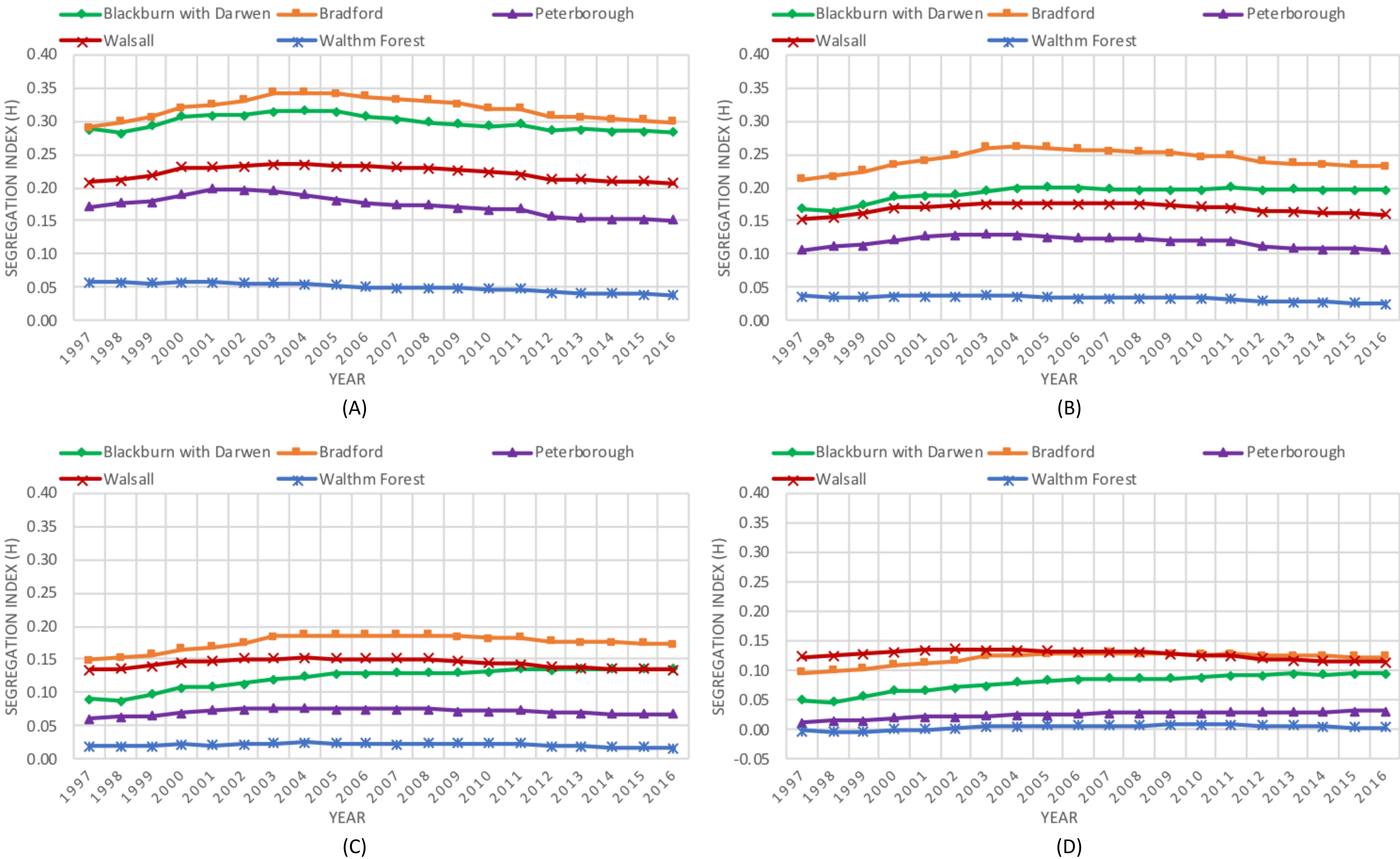

Figure 3. Segregation indices for five Integration Areas (1997-2016) at neighbourhood radius of (A) 500 metres, (B) 1,500 metres, (C) 2,500 metres, and (D) 3,500 metres

4.2. Scale effect on segregation

When comparing the corresponding Integration Areas at different scales in Figure 3, it can be seen that segregation levels are higher at smaller neighbourhood scales. Dramatic temporal variations of segregation levels in Bradford and Peterborough appear to be smoothed by larger neighbourhood radii. The standard deviation of the time series of Peterborough shrunk from 0.015 at the scale of 500 metres to 0.006 at the scale of 3,500 metres. Figure 4 shows the segregation profiles of the five Integration Areas against multiple scales in four selected years. For each Integration Area individually, segregation levels decline as the geographic scales increase from 500 metres to 3,500 metres. It is intuitively plausible that smaller scale areas are more homogenous in terms of ethnic compositions, while larger scale areas tend to be more heterogeneous.

Figure 3 also shows that the ordering of the Integration Areas differs between scales, suggesting that the degree of segregation in these Areas is a scale dependent issue. In Figure 4(A), Bradford is the most segregated Integration Area at scales lower than 1,700 metres in 1997, followed successively in descending order by Blackburn with Darwen, Walsall, Peterborough, and Waltham Forest. Walsall becomes the second most segregated Integration Area when the scale is larger than 1,700 metres and it then becomes the most segregated area when the scale is beyond 2,700 metres. Similar patterns can be observed in 2001, 2011, and 2016 from other sub-graphs in Figure 4. Combining observations from both Figure 3 and Figure 4, it can be said that the scale effect appears to have greater impact on segregation level changes than does the temporal effect.

The slope of segregation curves varies between the five Integration Areas (Figure 4). Blackburn with Darwen, Bradford, and Walsall have relatively flat curves, compared with the two steep curves of Peterborough and Waltham Forest. As the geographic scale increases, segregation levels comparatively converge albeit at different paces. The steep curves of Waltham Forest and Peterborough suggest that the two Integration Areas consist of smaller and homogenous neighbourhoods alongside neighbourhoods with dissimilar ethnic compositions, indicating that there is limited variation in ethnic compositions beyond certain micro scales. In contrast, flat curves are not that sensitive to scale changes. In Blackburn with Darwen, Bradford and Walsall, variations in ethnic compositions are clearly manifest over extensive geographic areas and segregation patterns present macro scales; and consequently, segregation levels remain much higher at or beyond the scale of 3,500 metres than with Peterborough and Waltham Forest.

To depict a more concrete picture of the residential patterns (Figure 5), we map individuals coloured by ethnic group at their addresses across the five Integration Areas in 2016. For disclosure control purposes, we randomise address points within a 200×200 square metres' area. White British/Irish are the majority group in all five Integration Areas, particularly in suburbs and rural hamlets. This group is more spatially mixed in Peterborough and Waltham Forest than in the remaining Integration Areas, where large enclaves of communities from the Indian subcontinent can be observed. These geographic patterns are well mirrored in the cross-scale segregation profiles shown in Figure 4.

Figure 6 shows the changing effects of scale for each Integration Area over time, which provides further insights on how geographic scale affects the evolution of segregation measures. Apart from Walsall, there is some tendency for the curves to become flatter over time, indicating that spatial segregation patterns are evolving from smaller clusters to segregation patterns observed across broader areas. The shift is determined by the relative increase or decrease in the segregation levels of Integration Areas at both small scales and large scales. For example, segregation levels in Bradford are generally declining at smaller scales such as 500 metres over time but are going up at larger scales (3,500 metres) according to Figure 3, which leads to a flatter curve in 2016 in Figure 6. Alternatively, for the case of Peterborough, although both of the segregation levels at smaller scales and larger scales are decreasing over time, the gradient of the curve in 2016 in Figure 6 has declined relative to that of 1997. Macro-scale segregation of an area can change very slowly over time unless population turnover is very rapid (Reardon et al., 2008). It can be observed from Figure 6 that macro segregation levels across more extensive areas (e.g. 3,500 metres) have changed for some Integration Areas, particularly for Blackburn with Darwen, Bradford, and Walsall. This indicates that these areas have experienced rapid population turnover and change in ethnic composition.

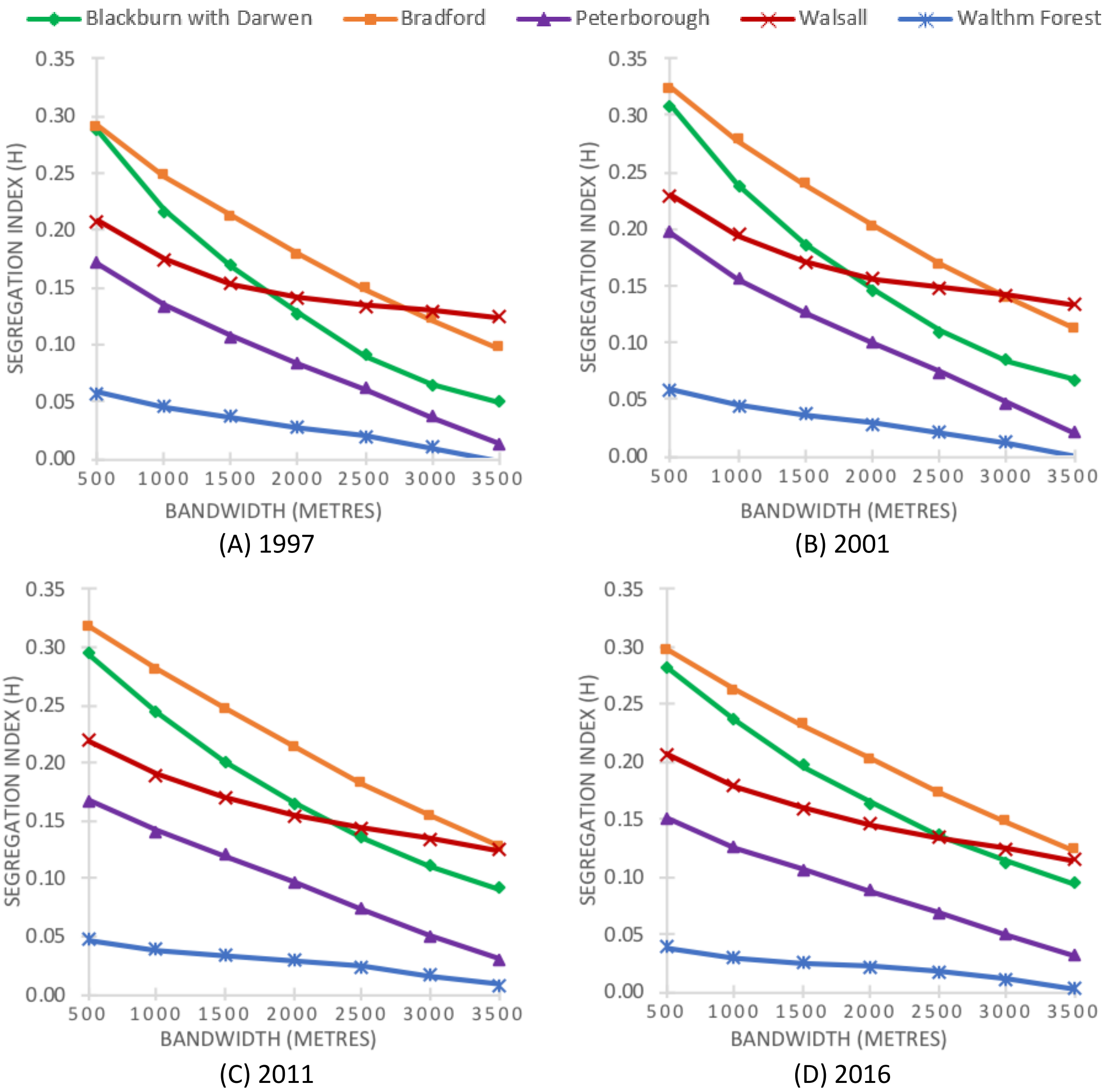

Figure 4. Segregation profile of the Integration Areas at a range of scales in four selected years: (A) 1997, (B) 2001, (C) 2011, and (D) 2016

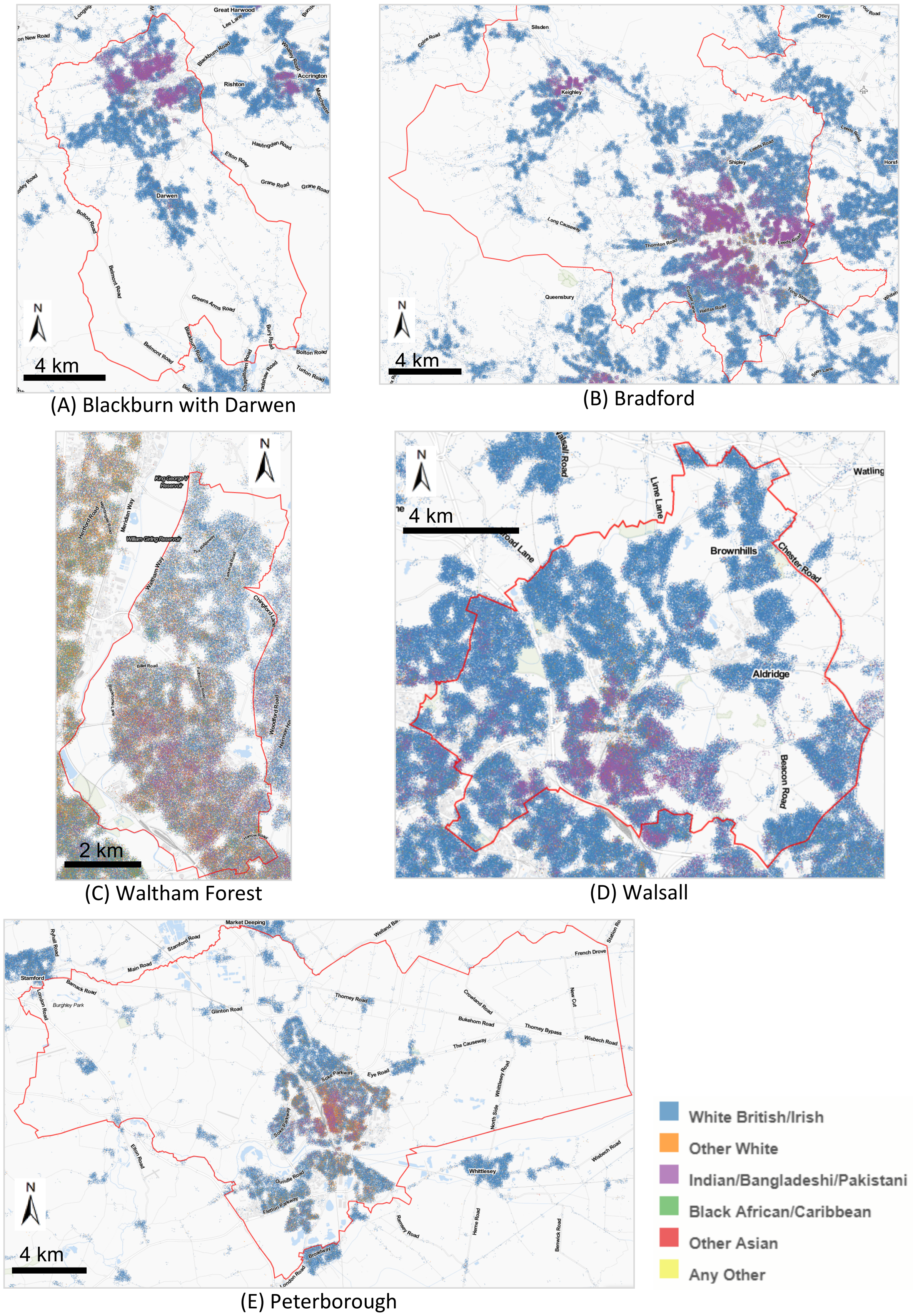

Figure 5. Maps of the five Integration Areas showing population distribution by ethnic groups in 2016; Integration Area boundaries are shown in red lines. (source: the 2016 Consumer Register)

We also adopt a more quantitative measure to demonstrate the steepness of these segregation curves using a macro/micro segregation ratio (Reardon et al., 2008). In our case, we choose 500 metres and 3,500 metres as the respective micro and macro scales of segregation. The macro/micro ratios and the temporal changes of the five Integration Areas can be seen in Figure 7. For the negative segregation indices of Waltham Forest in 1998, 1999, and 2001, we set their values to zero. In Figure 7, ratios of macro to micro scale segregation in Walsall are larger than ratios of other Integration Areas. For instance, a macro/micro ratio of 0.58 in 2016 means variations in ethnic compositions at large scales are more dominant than small scale segregation patterns in Walsall. Larger macro/micro ratios usually correspond to the flatter segregation curves shown in Figure 4, while smaller ratios show steeper curves such as Peterborough and Waltham Forest. Moreover, the change in ratios suggests how the macro/micro patterns of ethnic compositions change over time. Over the past twenty years, ratios of Blackburn with Darwen, Bradford, Peterborough, and Waltham Forest have been climbing up, indicating they are drifting towards being more macro-scale segregation dominant areas. In contrast to these four Integration Areas, patterns of micro and macro segregation in Walsall seem to be changing in the opposite direction during the study period.

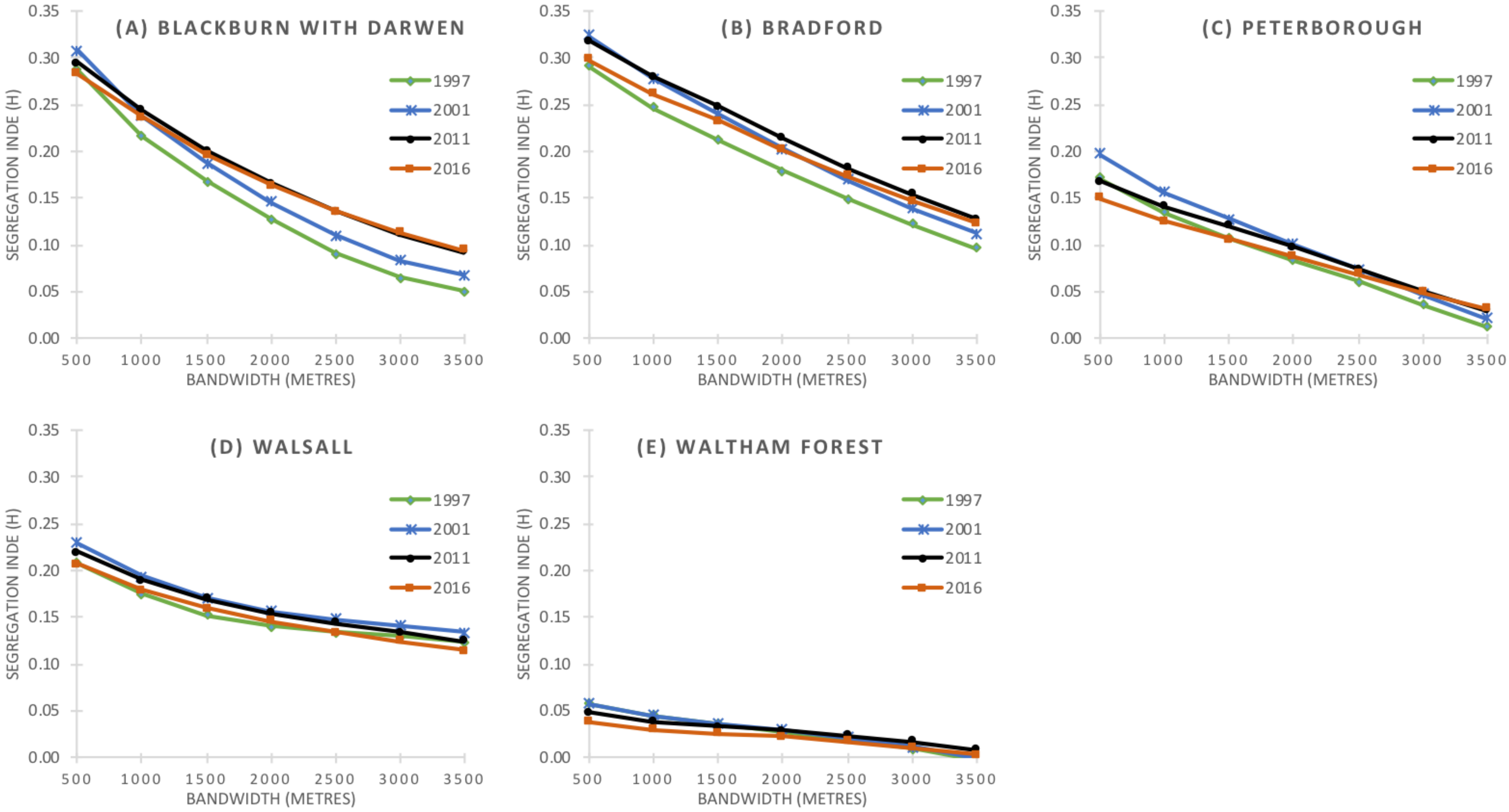

Figure 6. Temporal changes in segregation profiles for each Integration Area

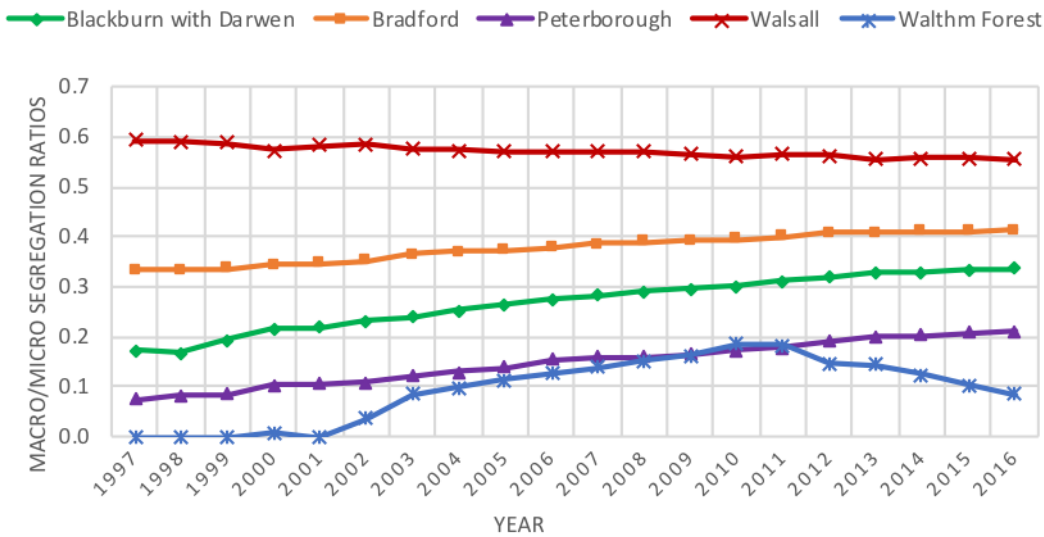

Figure 7. Macro/micro ratios (3,500 metres to 500 metres) of the five Integration Areas from 1997 to 2016

5. Discussion and conclusions

Segregation patterns and their changes result from the interplay of scale effect and temporal effect. The scale effect is likely subject to the geographic environment of the region: its topography (e.g. hills or valleys) and land use morphology (e.g. residential areas, farmland, and parks) can affect the spatial distribution of the population and thus the population captured in radius-based neighbourhood definitions. Waltham Forest and Walsall are examples of Integration Areas that are less sensitive to the scale change, since the Integration Areas are mostly covered by irreversibly urban land use. Given a snapshot of the entire population within these areas during one specified year, an individual's neighbourhood on average would not record dramatic changes when the radius of the neighbourhood is increased. This is because population in urban areas are relatively more ethnically diverse and more evenly distributed throughout space, while for more rural Integration Areas, the population is concentrated within a few villages or hamlets and the ethnic compositions in the rural areas are mostly dominated by White British. In these areas, changes over time are more an outcome of demographic processes such as birth, death, emigration and immigration.

The scale effect and temporal effect are intertwined to shape the landscape of residential segregation in these Integration Areas. Segregation levels decrease with increases in radial extents because larger neighbourhoods may incorporate higher ethnic diversity. Segregation levels appear to present fewer variations over time at larger scales than at smaller scales. The rank order of the five Integration Areas has barely changed over time at these pre-defined lower scales; however, the rank order of the five Integration Areas is not consistent across the geographic scales. It can be also observed from Figure 6 that the geographic scale has greater impact on the segregation levels than does the temporal effect.

Our time series analysis broadly supports views that Britain is not experiencing increased ethnic segregation. In methodological terms, our work contributes to the understanding of the spatial granularity at which segregation is manifest. Investigation of temporal changes and the effect of scale upon measured segregation suggest that we cannot simply assert segregation levels for one study area have declined or increased over years, which is the common conclusion of most of the studies in the literature. First, the trends of spatial segregation in the five Integration Areas are not monotonic over time, as exemplified by Bradford. With the finer granular ethnicity data from the annually updated Consumer Registers, we are able to capture the demographic changes between census years. Second, temporal trends of segregation in the five Integration Areas are not consistent across the geographic scales. Changes in segregation levels over years can be contradictory at smaller scales and at larger scales. This finding presents an important caveat to researchers and policymakers: namely that reports on temporal trends in residential segregation need to include a specification of the geographic scale of analysis. The steepness of the curves showing segregation against geographic scales provides further information on segregation profiles that move beyond segregation levels. Flat curves represent macro segregation dominant patterns, and steep patterns represent micro segregation dominant patterns. We plot the macro/micro ratios as a crude measure to show how geographic scales of segregation evolve. By observing macro-scale segregation changes, we find these Integration Areas have experienced rapid demographic change.

From a policy point of view, we may conclude that the challenges posed by residential segregation are not uniform across the different Integration Areas. Therefore, more localised strategies should be considered when tackling residential segregation. Our findings suggest the macro scale segregation is the predominant segregation pattern in Blackburn, Bradford, and Walsall. Thus, regarding the causes of residential segregation, strategies in these areas should be planned and placed within a more holistic policy framework at regional or even national level. This is because macro scale desegregation likely requires extensive cooperation on land use planning, housing policy, and job market opportunities among the government at multiple levels. Policy priorities need to be made to increase economic prosperity, to connect across communities, to establish a more affordable housing market, and to increase the mixing of schoolchildren between different communities. Such measures may include group-specific policy interventions, since the consequences of macro-scale and micro-scale segregation may differ between ethnic groups. Ethnic groups characterised by greater socioeconomic disadvantage may reinforce micro scale segregation because of their positions in the housing market. Some BAME communities, for instance the Pakistanis and Bangladeshis, have younger age profiles, which makes local authorities such as Blackburn and Bradford among the youngest places in England. These areas may need to orient policy to address segregation among younger residents, although over-all daily activity patterns (e.g. with respect to schools attended) may be at least as important as night-time residence. The rapid turnover of some "Other White" populations suggests that Peterborough and Waltham Forest should develop policy focus to support new immigrants from recent EU member states. Rather than treating the "Other White" group as one homogenous group, policy interventions may need to be sensitive to the sub-groups (e.g. the Polish, Romanian, and Czech components). Local plan responses to Integration Area priorities emphasise issues such as improving economic prosperity and improving linkage between both adult and juvenile community members. Residential segregation is but one impediment to these objectives, since communities can also engage through common workplace and leisure activities. The analysis of changing levels of residential segregation at a range of scales is thus strategically important when framing the objectives and successes of these policies.

Patterns, causes, and consequences of segregation are three pillars underpinning the conceptual framework of residential segregation research. Our research spectrum currently centres on measuring patterns of residential segregation, rather than discussing its causes and consequences extensively. To make full use of the information in Consumer Registers, future work can be extended to investigate the possible causes and outcomes. For example, internal migration rates by ethnic groups may be identified from the linkage of the same cohorts of people across Consumer Registers. Such evidence may in explain the transition of segregation patterns. Another possible extension to our current research could be evaluating how different forms of distance decay function would have affected the segregation measurements, although Catney (2018) suggests that the specific form of kernel selected is unlikely to have a major impact on the results. In addition to this limitation, our analysis nevertheless fundamentally remains focused upon the geography of night-time residence (Spielman et al., 2017), and thus does not address questions as to whether or not it is the segregation of daily activity patterns that defines the negative aspects of segregation. In our future work, we hope to develop and adopt consumer data sources that will allow us to identify the activity patterns associated with residence in different neighbourhoods and hence redefine segregation in these terms.

In methodological terms, our motivation is to effect the re-use of consumer data to devise frequently updateable estimates of changes in the ethnic composition of neighbourhoods across a full range of scales. A greater real share of the increased volume of data that are collected about citizens today are assembled by customer-facing organisations, and we believe there to be demonstrable value in re-using these for the social good. We use this new consumer data infrastructure to infer ethnicity from given and family name pairings, using the results of collaborative research with the UK Office for National Statistics. In substantive terms, the grounding of these inferential procedures at the level of the individual makes it possible for us to produce estimates of neighbourhood change not only at more frequent time intervals but also at a full range of spatial scales.

This paper has addressed the challenge of lack of multi-scale and frequently updated data and has provided explicitly scale based metrics for measuring segregation. We have developed a novel means of calculating individual level spatial segregation indices in England. The name-based ethnicity inferences from annual Consumer Registers enable us to monitor annual segregation changes of the five Integration Areas over a twenty-year period. We have made full use of the granularity of Consumer Registers to formulate an entropy-based spatial segregation to avoid the MAUP and “checkboard” issues posed by traditional non-spatial segregation measures. More importantly, by incorporating the spatial proximity, we have developed the capability of changing the ethnic neighbourhood radius to explore the geographic scale effect on segregation levels. Our results suggest that residential segregation is such a complex spatial-temporal phenomenon that no monotonic trend can be generalised simply across the entire range of geographic scales. It should be noted that segregation levels and trends could be meaningful only if they are referenced to specific geographic scales. Given the fact that varied segregation patterns and transitions are uncovered among different Integration Areas, more localised plans need to be implemented when devising community integration strategies. We believe that the proposed method of processing Consumer Registers offers a promising way to inform policy efforts promoting social integration.

Acknowledgement

This research was funded by the Engineering and Physical Sciences Research Council UK Regions Digital Research Facility grant (EP/M023583/1), the Economic and Social Research Council Consumer Data Research Centre grant (ES/L011840/1), and the Economic and Social Research Council grant ‘The Analysis of Names from the 2011 Census of Population’ (ES/L013800/1). The authors would like to thank CACI Ltd and DataTalk Research Ltd for providing the Consumer Register and Electoral Roll data under a special research licence to enable us to carry out this research.

References:

Aspinall, Peter J. (2002). Collective Terminology to Describe the Minority Ethnic Population:The Persistence of Confusion and Ambiguity in Usage. *Sociology, 36*(4), 803-816. doi:10.1177/003803850203600401

Berthoud, Richard. (1998). Defining ethnic groups: Origin or identify? . *Patterns of Prejudice, 32*(2), 53-63.

Cantle, Ted. (2001). *Community Cohesion: A Report of the Independent Review Team*. Home Office, United Kingdom Retrieved from http://tedcantle.co.uk/pdf/communitycohesion%20cantlereport.pdf.

Cantle, Ted, & Kaufmann, Eric. (2016). Is segregation on the increase in the UK? Retrieved from https://www.opendemocracy.net/wfd/ted-cantle-and-eric-kaufmann/is-segregation-on-increase-in-uk

Casey, Dame Louise. (2016). *The Casey Review: A review into opportunity and integration*. London Retrieved from https://www.gov.uk/government/publications/the-casey-review-a-review-into-opportunity-and-integration.

Catney, Gemma. (2015). Exploring a decade of small area ethnic (de-)segregation in England and Wales. *Urban Studies, 53*(8), 1691-1709. doi:10.1177/0042098015576855

Catney, Gemma. (2016). The Changing Geographies of Ethnic Diversity in England and Wales, 1991–2011. *Population, Space and Place, 22*(8), 750-765. doi:10.1002/psp.1954

Catney, Gemma. (2018). The complex geographies of ethnic residential segregation: Using spatial and local measures to explore scale‐dependency and spatial relationships. *Transactions of the Institute of British Geographers, 43*(1), 137-152. doi:doi:10.1111/tran.12209

Feitosa, F. F., Camara, G., Monteiro, A. M. V., Koschitzki, T., & Silva, M. P. S. (2007). Global and local spatial indices of urban segregation. *Int. J. Geogr. Inf. Sci., 21*(3), 299-323. doi:10.1080/13658810600911903

Finney, Nissa, & Simpson, Ludi. (2009). *'Sleepwalking to segregation'? Challenging myths about race and migration*: Policy Press at the University of Bristol.

Gale, Richard. (2013). Religious Residential Segregation and Internal Migration: The British Muslim Case. *Environment and Planning A, 45*(4), 872-891. doi:10.1068/a4515

Goodchild, Michael F. (2013). The quality of big (geo)data. *Dialogues in Human Geography, 3*(3), 280-284. doi:10.1177/2043820613513392

Harris, Richard. (2017). Measuring the scales of segregation: looking at the residential separation of White British and other schoolchildren in England using a multilevel index of dissimilarity. *Transactions of the Institute of British Geographers*. doi:10.1111/tran.12181

Hennerdal, Pontus, & Nielsen, Michael Meinild. (2017). A Multiscalar Approach for Identifying Clusters and Segregation Patterns That Avoids the Modifiable Areal Unit Problem. *Annals of the American Association of Geographers, 107*(3), 555-574. doi:10.1080/24694452.2016.1261685

Iceland, J., & Mateos, P. (2011). Ethnic Residential Segregation by Nativity in Great Britain and the United States*. *J Urban Aff, 33*(4), 409-429. doi:10.1111/j.1467-9906.2011.00555.x

Johnston, Ron, Forrest, James, & Poulsen, Michael. (2002). Are there Ethnic Enclaves/Ghettos in English Cities? *Urban Studies, 39*(4), 591-618. doi:10.1080/00420980220119480

Kandt, Jens, & Longley, Paul A. (2018). Ethnicity estimation using family naming practices. *PLOS ONE, 13*(8), e0201774. doi:10.1371/journal.pone.0201774

Lan, Tian, Kandt, Jens, & Longley, Paul. (2018). Ethnicity and Residential Segregation. In P. Longley, J. Cheshire, & A. Singleton (Eds.), *Consumer Data Research* (pp. pp. 71-83). London: UCL Press.

Lansley, Guy, & Li, Wen. (2018). Consumer Registers as Spatial Data Infrastructure and their Use in Migration and Residential Mobility Research. In P. Longley, J. Cheshire, & A. Singleton (Eds.), *Consumer Data Research*. London: UCL Press.

Lansley, Guy, Li, Wen, & Longley, Paul A. (2019). Creating a linked consumer register for granular demographic analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. doi:10.1111/rssa.12476

Lee, B. A., Reardon, S. F., Firebaugh, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). Beyond the Census Tract: Patterns and Determinants of Racial Segregation at Multiple Geographic Scales. *Am Sociol Rev, 73*(5), 766-791. doi:10.1177/000312240807300504

Logan, John R., Spielman, Seth, Xu, Hongwei, & Klein, Philip N. (2011). Identifying and Bounding Ethnic Neighborhoods. *Urban Geography, 32*(3), 334-359. doi:10.2747/0272-3638.32.3.334

Longley, Paul A, Goodchild, Michael F, Maguire, David J, & Rhind, David W. (2015). *Geographic information science and systems* (4th ed.): John Wiley & Sons.

Ministry of Housing, Communities & Local Government. (2018). *Integrated Communities Strategy green paper*. London Retrieved from https://www.gov.uk/government/consultations/integrated-communities-strategy-green-paper.

Monkkonen, Paavo, & Zhang, Xiaohu. (2014). Innovative measurement of spatial segregation: Comparative evidence from Hong Kong and San Francisco. *Regional Science and Urban Economics, 47*, 99-111. doi:10.1016/j.regsciurbeco.2013.09.016

O'Brien, Oliver, & Longley, Paul. (2018). Given and Family Names as Global Spatial Data Infrastructure. In P. Longley, J. Cheshire, & A. Singleton (Eds.), *Consumer Data Research* (pp. 53-67). London: UCL Press.

O'Kelly, M.E., & Horner, M.W. (2003). Aggregate accessibility to population at the county level: U.S. 1940–2000. *Journal of Geographical Systems, 5*(1), 5-23. doi:10.1007/s101090300101

O'Sullivan, David, & Wong, David W. S. (2007). A Surface-Based Approach to Measuring Spatial Segregation. *Geographical Analysis, 39*(2), 147-168. doi:10.1111/j.1538-4632.2007.00699.x

Office for National Statistics. (2009). *Final recommended questions for the 2011 Census in England and Wales: Ethnic group*. Retrieved from https://www.ons.gov.uk/file?uri=/census/2011census/howourcensusworks/howweplannedthe2011census/questionnairedevelopment/finalisingthe2011questionnaire/final-recommended-questions-2011-ethnic-group_tcm77-183998.pdf.

Openshaw, Stan. (1984). The modifiable areal unit problem *Geobooks*. Norwich, England: University of East Anglia.

Östh, John, Clark, William A. V., & Malmberg, Bo. (2015). Measuring the Scale of Segregation Using k-Nearest Neighbor Aggregates. *Geographical Analysis, 47*(1), 34-49. doi:10.1111/gean.12053

Östh, John, Lyhagen, Johan, & Reggiani, Aura. (2016). A new way of determining distance decay parameters in spatial interaction models with application to job accessibility analysis in Sweden. *European Journal of Transport and Infrastructure Research, 16*(2).

Peach, Ceri. (2009). Slippery Segregation: Discovering or Manufacturing Ghettos? *Journal of Ethnic and Migration Studies, 35*(9), 1381-1395. doi:10.1080/13691830903125885

Petersen, Jakob, Longley, Paul, Gibin, Maurizio, Mateos, Pablo, & Atkinson, Philip. (2011). Names-based classification of accident and emergency department users. *Health & place, 17*(5), 1162-1169.

Phillips, T. (2005). After 7/7: Sleepwalking Into Segregation. Speech to the Manchester Council for Community Relations: Manchester University.

Reardon, Sean F., Farrell, Chad R., Matthews, Stephen A., O'Sullivan, David, Bischoff, Kendra, & Firebaugh, Glenn. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Social Science Research, 38*(1), 55-70. doi:https://doi.org/10.1016/j.ssresearch.2008.10.002

Reardon, Sean F., Matthews, Stephen A., O'Sullivan, David, Lee, Barrett A., Firebaugh, Glenn, Farrell, Chad R., & Bischoff, Kendra. (2008). The Geographic Scale of Metropolitan Racial Segregation. *Demography, 45*(3), 489-514.

Reardon, Sean F., & O'Sullivan, David. (2004). Measures of Spatial Segregation. *Sociological Methodology, 34*(1), 121-162. doi:10.1111/j.0081-1750.2004.00150.x

Reardon, Sean F., & Yun, John T. (2001). Suburban Racial Change and Suburban School Segregation, 1987-95. *Sociology of Education, 74*(2), 79-101. doi:10.2307/2673164

Simpson, Ludi. (2004). Statistics of Racial Segregation: Measures, Evidence and Policy. *Urban Studies, 41*(3), 661-681. doi:10.1080/0042098042000178735

Simpson, Ludi. (2007). Ghettos of the mind: the empirical behaviour of indices of segregation and diversity. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*(2), 405-424. doi:10.1111/j.1467-985X.2007.00465.x

Spielman, Seth, Xiao, Ningchuan, Cockings, Samantha, & Tanton, Robert. (2017). Statistical systems and census data in the spatial sciences. *Computers, Environment and Urban Systems, 63*, 1-2. doi:https://doi.org/10.1016/j.compenvurbsys.2017.02.001

Theil, Henry. (1972). *Statistical decomposition analysis*. Amsterdam: North-Holland Publishing Company.

Wong, David W. S. (1999). Geostatistics as Measures of Spatial Segregation. *Urban Geography, 20*(7), 635-647. doi:10.2747/0272-3638.20.7.635

Wong, David W. S. (2002). Spatial Measures of Segregation and GIS. *Urban Geography, 23*(1), 85-92. doi:10.2747/0272-3638.23.1.85