Multi-Scale Segregation: Multilevel Modelling of Dissimilarity – Challenging the Stylized Fact that Segregation is Greater the Finer the Spatial Scale

A very large literature has explored the intensity of urban residential segregation using the index of dissimilarity. Several recent studies have undertaken such analyses at multiple spatial scales, invariably reaching the conclusion that the finer-grained the spatial scale the greater the segregation. Such findings over-state the intensity of segregation at finer spatial scales because they fail to take into account an argument made by Duncan et al. (1961) some seventy years ago that indices derived from fine-scale analyses must necessarily incorporate those from coarser scales, with the consequence that finer-scale segregation is invariably over-estimated. Moreover, most studies ignore stochastic variation that results in upward bias in the estimates of segregation. This paper demonstrates the importance of of a recently developed multilevel modelling procedure that identifies the 'true' intensity of segregation at every level in a spatial hierarchy net of its intensity at other levels, and net of stochastic variation This is illustrated by both a simulated data set and an empirical study of an English city, with the latter raising important substantive issues regarding the interpretation of segregation patterns and the processes underlying them.

Keywords: segregation, dissimilarity, scale, multilevel modelling

There is a very large literature, stretching back to the 1950s, on the measurement of ethnic residential segregation (for recent reviews see Reardon, 2006; Wong,2016; Lloyd et al., 2015); much of it uses the indices of dissimilarity and segregation widely deployed after several classic papers emanating from the University of Chicago (e.g. Duncan and Duncan 1955, 1957; Duncan and Lieberson 1959). For much of that time analysts have been constrained by two key challenges: firstly any measurement of segregation is conditional on the nature of the available data, especially the areal units for which they have been reported; and second, although many papers have sought to discuss the (changing) intensity of segregation such descriptions have been limited by a lack of empirical tests of significance that take account of inherent uncertainty in observed data, so-called natural variation requiring confidence intervals even with full count census data (Eayres, 2008).

Recent advances in computing power and the development of geographical information systems associated with improved data sources – some of them geocoded – have allowed segregation's intensity, and its spatial organisation, to be explored at a range of spatial scales (Wong et al., 2018). Some of the spatial structures deployed are created by census and other agencies; others use bespoke formats created within software systems (see, for example, Lee et al. 2008; Reardon et al. 2008, 2009; Östh et al. 2015; Clark et al. 2016; Randon-Furling et al., 2018). Several of those authors have noted that identification of varying levels of segregation at different scales invites investigations of the multi-scalar location-decision processes in operation, leading to possible evaluations of the relative intensity of, say, macro-, meso- and micro-levels of segregation. As Fowler (2015) argues, there is no 'correct' scale for studying segregation and its changing patterns; people and households make their residential location-decisions at a variety of spatial scales and also experience segregation's impact at several scales – with consequences for the nature of any neighborhood effects (Enos, 2017).

Those who have analysed segregation levels at a variety of scales have mostly concluded that it is greatest at the finest aggregation levels. The coarser the aggregation (i.e. the fewer and, on average, the larger the areal units deployed) the less intense the measured segregation, although the nature of the exact relationship between scale and segregation varies according to the degree of spatial autocorrelation in the distribution of the groups being studied and the particular areal unit aggregation level (Wong 1997). Thus Logan et al. (2015, 1077), for example, wrote that their purpose was 'not to demonstrate that segregation is higher at a finer spatial scale, *which is already well known*' (our emphasis); indeed, this result is found so regularly it can be seen as one of geography's few stylised facts (Bell et al. 2014).

This conclusion reflects a lack of appreciation of how multi-scale geographies should be analysed and the nature of data as they are aggregated, a process which by definition is akin to spatial smoothing. An alternative approach has its main inspiration in pioneering work by the same scholars who popularised the widely-used indices of dissimilarity and segregation (Duncan et al. 1961); several (e.g. Fowler 2015, 2) have since noted that they recommended 'caution about the effect of scale on measures of segregation' but have not followed through on the implications of that argument.

Drawing on Duncan et al.'s (1961) pioneering but almost entirely overlooked caution about the interpretation of dissimilarity and similar indices, a method to tackle the important issue of measurement in multi-scalar situations has recently been introduced (Jones et al., 2015). It argues, with Duncan et al., that almost all attempts to measure segregation at a hierarchy of spatial scales mis-represent the true situation and presents a multilevel modelling approach which remedies that problem. This paper extends that pioneering essay in two important aspects. First, it shows that multilevel estimates of the variance capturing the degree of segregation at one scale net of other scales and net of stochastic variation can be transformed into a measure (Dm) that is directly comparable with, and can be interpreted in the same way as, the commonly-deployed index of dissimilarity (D). Secondly, the essence of that measure, and the problems with the D index, are clarified through the use of a series of simulated cities that clearly illustrate the core argument regarding the necessity of multilevel analyses of segregation. These simulated cities have known patterns whose analysis illustrates the potentially misleading results revealed by computing D indices at three separate scales and comparing them with the modelled alternative Dm measures. Data from one multi-ethnic English city, Leicester, are then deployed to exemplify the interpretation of the multilevel results and compare them with those suggested by the index of dissimilarity, which generates a discussion of what the new measure offers to the understanding of segregation patterns. The paper is deliberately expository and we point to references for those who require technical detail.

Scale and Segregation Indices

Duncan et al. (1961, 84) were among the first to reach the widely-accepted conclusion that 'in general, the smaller the average size of areal unit, the larger the index [of dissimilarity] value', but they followed that statement with:

> … if one system of areal units is derived by subdivision of the units of another system, the index computed for the former can be no smaller than the index for the latter. Thus the index of concentration on a county basis will exceed the index on a State basis, because the county index takes into account intrastate concentration.

They didn't go further in suggesting how the degree of segregation/concentration could be calculated at each scale independent of the others, but they did conclude their discussion of the patterns in a five-level spatial hierarchy with (pp. 98-99):

> Many researchers on areal differentiation are forced to work with prefabricated areal units which they accept for reasons of convenience and expediency: moreover, as we have indicated, the results of manipulating areal data often are to some degree dependent on the choice of a set of areal units. Consequently, present practice in research can be fully satisfactory neither from the extreme "nominalist" viewpoint (because the description can

> be given in terms of a particular set of areal units) nor from the extreme "realist" viewpoint (since prefabricated areal units are not "real" regions). How this problem may be resolved cannot be foreseen. But it seems that men [sic] trying to develop cogent theories of areal structure will have to reckon with it for some time to come. Meanwhile, students of areal structure must take into account the discrepancy between their hypothetical constructs and their actual results which is generated by the necessity of working with systems of areal units for which data are available.

As noted above, that remained the situation until relatively recently, but scholars who have pioneered multi-scale analyses have not taken the first quote from Duncan et al. into account. Nor have they reflected on a parallel argument made, without reference to Duncan et al., by Tranmer and Steel (2001a, 33) who observed that, in multilevel analyses of UK census data, if one of the levels is omitted from an analysis 'the variation that occurs at the level not included in the models is redistributed to the levels that the models do include'. It is not always clear how much is allocated to which other level but they also report that 'The results suggest that the effects of an omitted highest level will be reflected in estimated components for the highest level included in the model' (see also Tranmer and Steel, 2001b). A single-scale analysis of the US at the county level will therefore incorporate State differences and the two scales will be confounded, which accords with Duncan et. al's insight. Our contention is that while many have claimed to undertake multiscale analyses by varying the size of the areal units, they have actually only undertaken *single*-scale analyses on differentially aggregated data so that their results potentially confound different sources of variation and do not reflect the 'true patterns'at each scale. From this perspective, finding that small scale analysis has greater segregation is a *fait accompli* – nothing else could be found because the smaller scale includes the degree of segregation at higher scales.

Given these strictures regarding the analysis of multi-scale data using indices of dissimilarity and segregation – strictures that also apply to other indices such as those of isolation and exposure – a methodology is needed that separates out the intensity of segregation at each level independent from that at any higher levels. Both Voas and Williamson (2000) and Fischer et al. (2004) sought means of spatially decomposing segregation measures but, like Duncan et al. (1961), neither produced a method that identified its intensity at any one scale net of that at others. However, Wong (2003) has done so in developing what he calls (p184) a conditional segregation measure but this is essentially a descriptive approach and does not take into account, the stochastic nature of count data, an issue to which we know turn.

Segregation and Small Numbers

Alongside direct concerns about scale when measuring segregation using those two popular indices issues have also been raised about the nature of the indices themselves, especially when applied to data at fine spatial scales when the unit populations analysed may be small (as discussed by Winship, 1977; see also Falk et al., 1977 and Winship, 1978). Carrington and Troske (1997), for example, showed that the combination of structural and stochastic elements in the allocation of individuals to areal units with small average populations is likely to lead to an over-inflation of the indices and hence of the interpreted level of segregation which is known as the upwards bias of the null (Allen et al., 2015). If a minority group comprises 5 per cent of a city's population and that city is split into areas with an average of 100 residents, they show that a random allocation of the minority group's members across those areas would produce an average index of dissimilarity across many simulations of 0.18, and with 50 residents per area the average index would be an even more worrying 0.26 despite the 'true' situation being no segregation at all. The larger the minority group the lesser the problem: nevertheless, if it comprised 30 per cent of the city's population, the average index of dissimilarity produced by their random allocation across areas with 50 residents would be

0.12. The apparent segregation is not real but is bound to occur under conditions of natural variation when fine-grained spatial data and small absolute counts are being analysed, as increasingly they are in many studies given the richness of the available data.

In any city, the distribution of members of a minority group across its neighborhoods, however defined, will reflect not only structural factors – those usually discussed in analyses of segregation such as choice, financial constraints and discrimination in housing market operations – but also stochastic factors – reflecting, for example, the administrative allocation of areal unit boundaries relative to those (often fuzzy) of the neighborhoods into which location-decisions are made. Analyses seeking to identify the intensity of segregation at any scale should partial out those stochastic elements in order to bring the structural causes into clear focus. As Mazza and Punzo (2015, 81) have shown analytically, indices of segregation will be upwardly biased even when they are computed using full-count census data, the bias being greatest when the minority population is relatively small and an area's total population is small in absolute terms. Wong's (2003) descriptive conditional segregation approach will be affected by this bias.

A further issue regarding the use of dissimilarity and similar indices is the lack of any formal means of comparing one with another to assess whether a difference in their magnitude is statistically significant – i.e. not likely to have occurred by chance. This issue is relevant to any comparative studies but is especially so when assessments are being made of changes over time (e.g. Peach, 2009). Some essays into this problem have been undertaken (e.g. Allen et al., 2015; Lee et al., 2015) but most studies make no such formal assessments of differences.

The issues of multiscalar analysis, of bias associated with small absolute numbers, and of lack of an inferential framework are interrelated. A detailed, fine-scale, single-scale analysis not only confounds what may actually be variation at a higher scale but is also most likely to be prone to bias due to small denominators, while confidence intervals, if calculated, would be narrow due to the number of units involved. An aggregated, coarser-scale, analysis at a single scale would smooth out stochastic variation as the larger areas would usually result in larger denominators leading to less bias but wider degrees of uncertainty should accompany the results as there are fewer units in the analysis. All three problems therefore need to be tackled simultaneously.

Modelling Segregation

Three issues have been raised with regard to the use of dissimilarity indices in the analysis of residential patterns at a range of spatial scales, therefore:

- The need to separate out the intensity of segregation at each scale, net of its level at all others;
- The need to separate out the structural and stochastic components of any observed patterns through the treatment of small populations in component areas; and
- The need for a rigorous statistical analysis of observed patterns that allows for formal evaluations of the extent of differences in their measured intensity.

A recently-devised multilevel modelling strategy meets all three criteria (Leckie et al 2012) for the two ethnicity situation using a binomial model while Jones et al., 2015) develop a Poisson model for multigroup analysis. [1]A key feature of the former paper is that it allows a between area variance (on a logit scale) to be transformed into modelled indices of dissimilarity (Dm), which are directly comparable to, and interpreted in the same way as, the commonly-deployed D index. Its usefulness is evaluated here by using a series of four simulated data sets constructed to highlight the

[1] A detailed algebraic specification of the logit model in the context of residential segregation is given in Jones et al (2018b) which also includes code in MLwiN ,Stata and R.

disadvantages of the standard methodologies which the modelling strategy eliminates, and then through its application to data for one English city with a large multi-ethnic population, which introduces a discussion of how segregation patterns develop.

That not taking scale-confounding effects into account can lead to misleading results is first illustrated by three simple examples reported in Figure 1. Each refers to a city divided into six wards each of which is divided into four census tracts; each tract has a population of 100, all of whom are classified as either W or B.

In the first city, all 600 of the B population are concentrated into two of the wards, where they form 75 per cent of the population in each of the tracts; the other four wards are 100 per cent W. There is clearly segregation at the ward scale with the B population concentrated in a minority of them. At first glance there is also segregation at the tract scale, with the B population concentrated in eight of the twenty-four tracts: this is shown by the single-scale D indices, which are 0.890 for each scale. However, closer inspection suggests that there is no segregation at the tract scale additional to that at the ward scale; across the two wards with B residents there is no difference in their percentage contribution to each tract's population, and nor is there any in the other four wards which have no Bs in any tract. To project the 'true' situation correctly, therefore, the D value at the tract scale would be 0.0 if segregation at the larger scale were taken into account; as Duncan et al. (1961) pointed out, the value of D at the smaller scale must be at least as large as its value at the larger scale but, as in this example, that can be very misleading.

In the second city, the 600-strong B population is concentrated into just half of the wards, with the other half containing W residents only. Here there is segregation at both scales. The B population is concentrated into half of the wards and within each they form a larger share of the population (60 per cent) in two of the tracts than they do in the other two (40 per cent). The single-scale D indices are the same for both scales – 0.665. But does the value at the smaller scale over-estimate the degree of segregation there once segregation at the ward scale is taken into account?

Finally, the third city has no segregation at the ward scale since each contains 100 B residents. But there is segregation at the tract scale, since in each ward there are two tracts with no B residents and one each with 60 and 40. The single-scale D index at the ward scale is 0.0 whereas at the tract scale it is 0.665 – which appears to be correct.

So what are the 'true' levels of segregation at each scale, when the levels at other scales have been partialled out? To address that question we adopt a multilevel model-based inferential approach, which models patterns at several scales simultaneously, rather than a descriptive single-scale analysis. The model's basic form – a binomial logit – is set out in detail in Leckie et al. (2012) where, for example, the response variable is the proportion of the population claiming Black ethnicity in a particular tract in a particular ward with a total population of White plus Black as the denominator. The log of the odds of being Black is modelled as a function of an overall city-wide average, a differential for each ward (the highest area in the hierarchy) from that average, plus a tract differential (the next highest unit in the hierarchy) from the ward differential in which that tract nests. This use of differentials at each scale distinguishes the degree of patterning at each scale net of the other; it is possible to have a negative differential for both so that compared to the city the ward and the tract have low log-odds of being Black. Two positive values indicate high ethnicity concentrations in some areas at both levels. It is also perfectly possible to have low odds of being Black in a tract whose surrounding ward has high odds – and indeed vice versa. A differential of zero indicates the ward has the same proportion Black as the city average; a zero for a tract indicates a typical tract within a ward (i.e. it has the same proportion Black as the average across all tracts in that ward; because the assessment at each scale is made relative to the larger units within which the finer-grained units are nested the resulting indices are not aspatial but relate to clustering around a local mean).[2] These differentials are assumed to be Normally distributed (very plausible on the log scale) and are summarised by variances so that it is possible to distinguish between-ward variation as well as within-ward, between-tract differences. These are our primary measures of segregation; if there is no segregation at a particular level in the hierarchy then the variance at that level will be zero.

Modelling Segregation

An important element of this model is that random variation is catered for by a binomial disturbance term so that stochastic variation will be greatest when the underlying modelled proportion is close to 0.5 and the denominator is small. Consequently, the variances not only measure segregation at one scale net of another but also net of natural variation that accompanies count data, thereby tackling the problem of the upwards bias discussed earlier Moreover, the model is now cast in an overall inferential framework. An estimated variance whose uncertainty intervals do not include zero indicates genuine segregation in the form of unevenness – there is systematic segregation beyond the stochastic element. The model is estimated in a Bayesian framework using MCMC procedures (Jones, 2018b, gives a complex example with accompanying code) so that the measures of uncertainty around the estimates are credible intervals and may be asymmetric. An important aspect of the Leckie et al. (2012) paper is that they show that a number of well-known indices (D dissimilarity, Gini, Isolation and Theil's entropy) are all a monotonic increasing function of the modelled variance. They provide a method of transforming the variance and accompanying uncertainty into these indices if that is needed to compare with previous work and other forms of analysis, and we use the modelled D (Dm) below to compare with traditional, and potentially biased, descriptive D.

A Synthetic Example

This first evaluation of the methodology uses a simulated city comprising 625 micro-scale areas (each with a population of 300) nested within 125 meso-scale areas (i.e. five micro-scale areas in each meso-scale area) with these in turn nested within 25 macro-scale areas (with five meso-scale areas in each macro-scale area). The city has an ethnic minority of 30,000 individuals (i.e. 16 per cent of the total population of 187,500). These were randomly allocated to each micro-scale area with a maximum of 60, a minimum of 25, and a median of 48 per area, in four different scenarios: (I) the variance in the distribution of the minority group is equally allocated across the three scales; (II) most of the variance is at the micro-scale, with small amounts at each of the other two; (III) most of the segregation is at the meso-scale; and (IV) most of the segregation is at the macro-scale (Table 1). Realistically each count in each micro-area is simulated with natural variation according to a binomial distribution. The numerator and denominator of these data at the micro-scale in each scenario were then aggregated, as is done in standard multiscale analysis, to the meso-scale and then to the macro scale, and the resultant proportions calculated

For each scenario the index of dissimilarity (D) between the members of the minority group and the remainder of the population was derived for each scale (Table 2). The D indices for each scenario show the same trajectory across the three scales – smallest at the macro-scale and largest at the micro-scale – for each scenario, despite their differences in the allocation of the variance in the simulations (Figure 2a). In particular, although in scenario I the variance was equally distributed across the three scales, nevertheless the D values suggest that segregation was substantially greater

[2] The implicit nature of the model's handling of spatial dependence is considered in Jones et al. (2015) while the logit multilevel model is extended to an explicit spatial model at multiple scales in Jones et al (2018b). Despite the title, Wong (2003) does not explicitly handle spatial dependce.

at the micro-scale than at either of the other two and also that segregation was greater at the meso- than the macro-scale, an overall picture that is inconsistent with the 'real' (i.e. known) level of segregation. Because the measured value of D at the micro-scale must be at least as great as that at the larger scales within which its areas are nested, then any further segregation at the micro-scale must mean that its D is greater than that for the coarser scale.

Table 3 gives the results of the multilevel modelling of those data. For each scale at each scenario it gives the D index in Table 2 together with the modelled Dm value and its 95% CIs. The penultimate column compares the two indices (Dm-D) and the final column indicates whether there is no difference between the two or whether the D value is an over-estimate or under-estimate compared to Dm. For scenario I, where the variance was equally apportioned across all three scales, the three Dm values fully reflect that; there are small differences between the three values but the CIs overlap so there is no evidence of any significant difference. An equal intensity of segregation at each scale was created and the modelling replicates that (Figure 1b): by comparison, the D indices slightly under-estimate its intensity at the micro-scale but substantially over-estimate it at each of the other two.

In all three other scenarios the relative size of the Dm values faithfully reflects the relative allocation of the variance to the three scales. Thus in scenario II, where most variance was allocated to the micro-scale, this is reflected not only in its much larger (0.53) Dm value there compared to the other two scales (0.13 for each) but also in the clear indication that the former value is significantly larger than the other two (Figure 1b). Similarly, in scenario III the largest index (which has the same value as in scenario II) is correctly placed at the meso-scale, and it is significantly larger than the Dm values for the other two scales, which are not significantly different from each other. Finally, the highest level of segregation in scenario IV is correctly located at the macro-scale, with again no significant difference between the modelled Dm values for the other two scales. In all four scenarios, therefore, the Dm values fully replicate the simulated allocation of variance across the three scales, but the D values do not.

A Substantive Empirical Example

An empirical example of the modelling approach uses 2011 census data for Leicester, an English city with 328,839 residents. It had a very large Indian ethnic population (93,335) and several other substantial ethnic minority groups of which five – Pakistani (8,067), Chinese (4,245), Arab (3,311), Black African (12,480) and Black Caribbean (4,790) – are studied here. For census reporting purposes Leicester was divided into 969 Output Areas (OAs) designed to maximise internal homogeneity and maximise external heterogeneity in terms of dwelling type and tenure (Cockings et al., 2011). These, with an average total population of 370, were nested into 192 Lower Layer Output Areas (LSOAs) on the same criteria, which in turn were nested into 37 Middle Layer Output Areas (MSOAs).

For illustrative purposes we compare five pairs of distributions at each of the three scales:

- Indian: Non-Indian;
- Indian: Pakistani;
- Chinese: Non-Chinese;
- Arab: Non-Arab; and
- Black African: Black Caribbean.

Table 4 gives the descriptive D index for each pair at each scale in an aggregate analysis. All five, as expected given the above arguments, show the same trajectory across the three scales: the finer the scale the larger the index and hence the suggested intensity of segregation (Figure 3a). There is one deviation from that pattern, however; whereas in four of the comparisons the micro-scale D is substantially larger than that for the macro-scale (with the meso-scale value in an intermediate position) the three D values for the Indian: Non-Indian comparison are very similar. The implication of this latter pattern – given the discussion above using simulated data – is that almost all of the variance must be at the macro-scale. Indians are concentrated into certain macro-scale sections of the city relative to the remainder of the population (i.e. they are concentrated into certain MSOAs), but within those sections there is no substantial further segregation of the two groups: they live relatively apart at the macro-scale but not at the micro-scale.

The modelled Dm values provide a very different picture of the patterns of segregation in Leicester (Table 5; Figure 3b).[3] With the Indian: Non-Indian comparison, for which the D indices suggest that segregation was equally intense at each of the three scales, the Dm values indicate that segregation was more than twice as intense at the macro-(MSOA) scale as at either of the other two: the former value (0.52) is statistically larger than the other two – which do not differ significantly from each other. Indians have concentrated into certain macro-sections of the city; within those – and also within those parts with few Indians – the degree of segregation of Indians from Non-Indians varies much less across either the meso- or the micro-scale areas. As the final column of Table 5 shows; the D index substantially over-states the degree of segregation at the two smaller scales, because – as argued and displayed here – that measure does take account of the interdependence of scales. The Dm indices, on the other hand, clarify the situation: segregation of Indians in Leicester is much more the result of macro-scale than meso- and micro-scale residential location decisions.

With three of the other comparisons, rather than the upward trend in the registered level of segregation with decreasing scale as shown by the D indices (Figure 3a), the Dm values in Figure 3b form a V-shaped pattern, with segregation relatively high at the macro- and micro-scales but substantially (and in two of the cases significantly) less so at the meso-scale. Segregation of Chinese from Non-Chinese, of Arabs from Non-Arabs, and of Indians from Pakistanis occurs at both the macro-scale – each group to a considerable extent is concentrated in different parts of the city – and also the micro-scale: within any larger district, group members tend to be located in separate small neighborhoods (OAs). These patterns are not apparent from interrogation of the D indices, in large part because (as shown in the final column of Table 5) they over-state the level of segregation relative to what is shown by the Dm indices.

Only in the final comparison, between the distribution of the two Black ethnic groups – African and Caribbean – is the standard pattern displayed by the D indices reproduced by the Dm values. The relatively small Dm value indicates that these two groups are not substantially separated from each other at the macro-scale; they tend to concentrate in the same major districts of Leicester. But the much larger – and significantly so – micro-scale Dm value shows that within any district there is much more spatial separation of the two groups at the more local, neighbourhood, scale.

These findings for Leicester (which parallel those for other studies in which the earlier version of the modelling strategy was applied: Johnston et al., 2016a, 2016b; Manley et al., 2015) have important implications for the analysis and interpretation of segregation patterns and processes. Most studies – many using the D index – have focused on a single-scale only, usually the finest-grained available, and have interpreted the findings as showing that, to a greater or lesser extent, minority ethnic groups are concentrated, through a combination of both choice and constraint, into small neighborhoods. A multi-scale analysis such as this one of Leicester challenges such an interpretation and instead suggests a variety of patterns. The Indian population there is concentrated in certain macro-scale sections of the city in relative separation from the Non-Indian population, but within those macro-scale sections there is little evidence of further concentration

[3] The varying width of the credible intervals correctly reflects the number of units at each scale in the hierarchy; the more units the narrower the range of the CIs.

into particular smaller-scale neighborhoods. With the two Black ethnic groups, on the other hand, a reverse pattern emerges: a relatively small Dm at the macro-scale suggests that they are concentrated in the same sections of the city, but a much larger micro-scale Dm indicates that within those sections they are concentrated into different neighborhoods. A similar, if less pronounced, pattern characterises the residential patterns of Indians and Pakistanis: they are concentrated in the same macro-scale sections (Dm of 0.27) but occupy separate micro-scale neighborhoods within those sections (Dm of 0.35). Arabs are concentrated at both the macro- and the micro-scales (i.e. clustered within particular neighborhoods within city sections); Chinese are segregated at all three scales – much more so at the meso-scale than any of the other groups.

Conclusions

More than six decades ago Duncan et al. (1961) introduced a major caution to the interpretation of indices of spatial concentration, including those of dissimilarity which have been widely used since then in the analysis of ethnic residential segregation. They offered no statistical resolution to the problem identified, however, and with a very few exceptions nor have any others since. For more than sixty years analysts have made do with whatever, in Duncan at al.'s terms, prefabricated areal units are available with the needed data, from which they have drawn the conclusion that the finer the spatial grain the greater the segregation – implying that the smaller the areas the more exclusive they are in their ethnic composition.

Recent developments in both the nature of available data and the software systems (many of them bespoke) within which they can be analysed have drawn attention away from the single-scale studies of segregation that have long dominated the literature.[4] The growing interest in multi-scale studies has made Duncan et al.'s challenge important, since it suggested that analyses using indices of dissimilarity calculated separately at a range of scales could be open to substantial mis-interpretation: they would almost certainly indicate that segregation was greater (and certainly could not be less) as one moved from one scale to a smaller one, but that may not be a correct interpretation.

The resolution to that issue calls for a modelling strategy that can identify the 'true' level of segregation at any scale, net of that at others – either smaller or larger. Such a strategy has been introduced and illustrated here, presenting a multilevel modelling approach to segregation measurement that not only resolves the scale-confounding issue but additionally overcomes the problem that conventional indices of dissimilarity calculated for small area data tend to over-estimate the intensity of segregation at any scale and lack any statement of the reliability of their values of the sort associated with studies of statistical significance. Application of this approach to four simulated data sets has indicated its power in identifying the 'true' degree of segregation at any scale.

Methodologically much previous analysis has not been multiscalar at all but rather multiple analyses of data aggregated to a single scale. In contrast, this multilevel approach takes the observed data at the finest scale, recognizes their stochastic nature and partitions the 'genuine' variance into specified scales, thereby not pre-determining the results. This modelling approach is capable of numerous extensions. While here we have analysed pairs of ethnic groups at a time, cross-sectionally and essentially aspatially (i.e. no account has been taken of the relative locations of the areas in which the ethnic groups are concentrated), other applications have analysed: multiple groups simultaneously (Jones et al. 2015); changes over time at multiple scales (Johnston et al.

[4] Without individual-level data, analysts remain dependent on the arbitrary sets of areal units deployed by census and other bodies and with any one set are, of course, potentially substantially impacted by the Modifiable Areal Unit Problem (on which see Jones et al., 2018, and Fowler et al., 2018).

2016b); have reported analyses that distinguish unevenness from geographical clustering at multiple scales (Jones et al. 2018b); and have analysed more than one source of residential segregation – class and ethnicity – simultaneously (Jones et al. 2018a).

Finally, the multi-scale findings reported here from the Leicester case study (alongside other recent studies of cities in four countries: Jones et al. 2015; Johnston et al. 2016a, 2016b; Manley et al. 2016) have posed a new set of challenges to students of segregation, of the type essayed by Fowler (2015). The conventional wisdom has been that observed segregation patterns have resulted from residential location decisions either largely made or constrained at local (neighborhood) scales, but the Leicester findings suggest that this may not always be the case. With some groups at least macro-scale location decisions look to be the more important; through choice and/or constraint, the group concentrates into a particular section (or sections) of the city, within which there is less concentration into certain neighborhoods. Such findings indicate that the methodological advance promoted here can have substantial substantive implications for theory and understanding; one size does not fit all.

Literature Cited

Allen, R., S. Burgess, R. Davidson and F. Windmeijer. 2015. More reliable inference for the dissimilarity index of segregation. *The Econometrics Journal* 18: 40-66.

Bell, A., R. J. Johnston and K. Jones. 2014. Stylised fact or situated messiness? The diverse effects of increasing debt on national economic growth. *Journal of Economic Geography* 15: 449–472

Carrington, W. J. and K. R. Troske. 1997. On measuring segregation in samples with small units. *Journal of Business & Economic Statistics* 15: 402-409.

Cockings, S., A. Harfoot, D. Martin and D. Hornby. 2011. Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 census output areas for England and Wales. *Environment and Planning A* 43: 2399-2418.

Clark, W. A. V., E. Anderson, J. Östh and B. Malmberg. 2016. A multiscalar analysis of neighborhood composition in Los Angeles, 2000-2010: a location-based approach to segregation and diversity. *Annals of the Association of American Geographers* 106: 1260-1284.

Duncan, O. D., R. P. Cuzzort and B. Duncan. 1961. *Statistical Geography*. New York: The Free Press.

Duncan, O. D. and B. Duncan. 1955. Residential distribution and occupational stratification. *American Journal of Sociology* 60: 493-503.

Duncan, O. D. and B. Duncan. 1957. *The Negro Population of Chicago*. Chicago: University of Chicago Press.

Duncan, O. D. and S. Lieberson. 1959. Ethnic segregation and assimilation. *American Journal of Sociology* 64: 364-374.

Eayres, D (2008) Commonly used public health statistics and their confidence intervals, Technical Briefing Number 3, Association of Public Health Observatories; available at https://fingertips.phe.org.uk/documents/APHO Tech Briefing 3 Common PH Stats and CIs.pdf.

Enos, R. D. 2017. *The Space Between Us: Social Geography and Politics*. Cambridge: Cambridge University Press.

Falk, R. F., Cortese, C. F. and Cohen, J. 1977. Utilizing standardized indices of residential segregation: comment on Winship. *Social Forces* 57: 713-716.

Fischer, C. S., G. Stockmayer, J. Stiles and M. Hout. 2004. Distinguishing the geographic levels and social dimensions of U. S. metropolitan segregation. *Demography* 41: 37-59.

Fowler, C. S. 2015. Segregation as a multiscalar phenomenon and its implications for neighborhood scale research: the case of South Seattle, 1990-2010. *Urban Geography* 37: 1-25.

Fowler, C. S., N. Frey, D. C. Folch, N. Nagle and S. Spielman. 2018. *Who are the People in my Neighborhood? The 'Contextual Fallacy' of Measuring Individual Context with Census Geographies*. Washington, DC: Center for Economic Studies, Discussion Paper CES 18-11.

Johnston, R. J., J. Forrest, D. Manley and K. Jones. 2016a. The scale of segregation: ancestral groups in Sydney, 2011. *Urban Geography*, 37, 985-1008.

Johnston, R. J., K. Jones, D. Manley and D. Owen. 2016b. Macro-scale stability with micro-scale diversity: modelling changing ethnic minority residential segregation: London 2001-2011. *Transactions of the Institute of British Geographers* NS41: 389-402.

Jones, K., R. J. Johnston, D. Manley, D. Owen and C. Charlton. 2015. Ethnic residential segregation: a multilevel, multigroup, multiscale approach exemplified by London in 2011. *Demography* 52: 1995-2019.

Jones, K., R. J. Johnston, J. Forrest, M. Charlton and D. Manley 2018a. Ethnic and class residential segregation: exploring their intersections a multilevel analysis of ancestry and occupational class in Sydney. *Urban Studies* 55: 1163–1184

Jones, K., D. Manley and R. J. Johnston. 2018b. Modelling residential segregation as unevenness and clustering: a multilevel modelling approach incorporating spatial dependence and tackling the MAUP. *Environment and Planning B: Urban Analytics and City Science*, doi 10.1177/2399808318782703

Lee, D., J. Minton and G. Pryce. 2015. Bayesian inference for the dissimilarity index in the presence of spatial autocorrelation. *Spatial Statistics*, 11, 81-95.

Leckie, G., R. Pillinger, K. Jones and H. Goldstein. 2012. Multilevel modelling of social segregation. *Journal of Educational and Behavioral Statistics*, 37, 3-30.

Lee, B. A., S. F. Reardon, G. Firebaugh, C. R. Farrell, S. A., Matthews and D. O'Sullivan. 2008. Beyond the census tract: patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review* 73: 766-791.

Logan, J. R., W. Zhang and M. D. Chunyu. 2015. Emergent ghettos: Black neighborhoods in New York and Chicago 1880-1940. *American Journal of Sociology* 120: 1055-1094.

Lloyd, C. D., Shuttleworth, I. and Wong, D. W. (eds.) 2015. *Social-Spatial Segregation: Concepts, Processes and Outcomes*. Bristol: The Policy Press.

Manley, D., R. J. Johnston, K. Jones and D. Owen. 2015. Macro-, meso- and micro-scale segregation: modelling changing ethnic residential patterns in Auckland, New Zealand, 2001-2013. *Annals of the Association of American Geographers* 105: 951-967.

Mazza, A and A. Punzo. 2015. On the upward bias of the dissimilarity index and its corrections. *Sociological Methods and Research* 44: 80-107.

Östh, J., W. A. V. Clark and S. Malmberg. 2015. Measuring the scale of segregation using k-nearest neighbor aggregates. *Geographical Analysis* 47: 34-49.

Peach, C. 2009. Slippery segregation: discovering or manufacturing ghettos? *Journal of Ethnic and Migration Studies* 35: 1381-1395.

Randon-Furling, J., M. Olteanu and A. Lucquiad. 2018. From urban segregation to spatial structure detection. *Environment and Planning B: Urban Analytics and City Science* doi 10.1177/2399808318797129

Reardon, S. F. 2006. A conceptual framework for measuring segregation and its association with population outcomes. In J. M. Oakes and J. S. Kaufman (eds.) *Methods in Social Epidemiology*. San Francisco: Jossey-Bass, 169-192 (second edition 2017).

Reardon, S. F., S. A. Matthews, D. O'Sullivan, B. A. Lee, G. Firebaugh, C. R. Farrell and K. Bischoff. 2008. The geographic scale of racial residential segregation, 1990-2000. *Demography* 45: 489-514.

Reardon, S. F., C. R. Farrell, S. A. Matthews, D. O'Sullivan, K. Bischoff and G. Firebaugh 2009. Race and space in the 1990s: changes in the geographic scale of racial residential segregation, 1990-2000. *Social Science Research* 38: 57-72.

Tranmer, M. and D. G. Steel. 2001a. Ignoring a level in a multilevel model: evidence from UK census data. *Environment and Planning A* 33: 941-948.

Tranmer, M. and D. G. Steel. 2001b. Using local census data to investigate scale effects. In N. J. Tate and P. J. Atkinson (eds.) *Modelling Scale in Geographic Information Science*. Chichester: John Wiley, 104-122.

Voas, D. and P. Williamson. 2000. The scale of dissimilarity: concepts, measurement and an application to socio-economic variation across England and Wales. *Transactions of the Institute of British Geographers* NS35: 465-481.

Winship, C. 1977. A revaluation of indexes of residential segregation. *Social Forces* 55: 1058-1066.

Winship, C. 1978. The desirability of using the index of dissimilarity or any adjustment of it for measuring segregation: reply to Falk, Cotese and Cohen. *Social Forces* 57: 717-720.

Wong, D. 1997. Spatial dependency of segregation indices. *The Canadian Geographer* 41: 128-136.

Wong, D. W. S. 2003. Spatial decomposition of segregation indices: a framework towards measuring segregation at multiple levels. *Geographical Analysis* 35: 180-194.

Wong, D W S (2016) From aspatial to spatial, from global to local and individual: Are we on the right track to spatialize segregation measures? In Recapturing space: New middle range theory in spatial demography, ed. M. Howell, R.Porter, and A. Matthews, 77–98. New York: Springer-Verlag.

Yao, J., Wong, D. W. S., Bailey, N. and Minton, J. 2018. Spatial segregation measures: a methodological review. *Tijdschrift voor Economische en Sociale Geografie* doi 10.1111/tesg.12305.

Table 1. The four simulation scenarios

	Description	Variance	Proportion of Variance Micro	Meso	Macro
I	All three scales equal	3	0.33	0.33	0.33
II	Micro-scale dominant	3	0.90	0.05	0.05
III	Meso-scale dominant	3	0.05	0.90	0.05
IV	Macro-scale dominant	3	0.05	0.05	0.90

Table 2. The calculated D index values for the simulated scenarios

Scenario/Scale	Macro	Meso	Micro
I	0.35	0.48	0.56
II	0.13	0.27	0.54
III	0.27	0.52	0.54
IV	0.39	0.48	0.48

Table 3. The Dm values and their CIs for the simulated scenarios

Scenario	Scale	D	2.5%CI	Dm	97.5%CI	Dm-D	Diff.
I	Macro-	0.35	0.28	0.39	0.49	-0.04	Under
	Meso-	0.48	0.31	0.36	0.41	0.12	Over
	Micro-	0.56	0.34	0.37	0.40	0.19	Over
II	Macro-	0.13	0.02	0.13	0.22	0.00	None
	Meso-	0.27	0.01	0.13	0.23	0.14	Over
	Micro-	0.54	0.51	0.53	0.55	0.01	None
III	Macro-	0.27	0.01	0.11	0.25	0.16	Over
	Meso-	0.52	0.49	0.53	0.57	-0.01	None
	Micro-	0.54	0.13	0.16	0.18	0.38	Over
IV	Macro-	0.39	0.44	0.53	0.62	-0.14	Under
	Meso-	0.48	0.13	0.17	0.20	0.31	Over
	Micro-	0.48	0.14	0.17	0.19	0.31	Over

Table 4. The calculated D index values for the Leicester data

Scale	MSOA	LSOA	OA
Indian: Non-Indian	0.52	0.54	0.57
Indian: Pakistani	0.25	0.31	0.39
Chinese: Non-Chinese	0.51	0.57	0.64
Arab: Non-Arab	0.41	0.49	0.61
African: Caribbean	0.20	0.29	0.43

Table 5. The Dm values and their CIs for the Leicester data

Scale	D	2.5%CI	Dm	97.5%CI	Dm-D	Diff
Indian: Non-Indian						
Macro-	0.52	0.44	**0.52**	0.59	0.00	None
Meso-	0.54	0.20	**0.23**	0.26	-0.31	Over
Micro-	0.57	0.23	**0.24**	0.25	-0.33	Over
Indian: Pakistani						
Macro-	0.25	0.21	**0.27**	0.34	0.02	None
Meso-	0.31	0.05	**0.14**	0.19	-0.17	Over
Micro-	0.39	0.32	**0.35**	0.38	-0.04	Over
Chinese: Non-Chinese						
Macro-	0.51	0.32	**0.40**	0.48	-0.11	Over
Meso-	0.57	0.20	**0.28**	0.33	-0.29	Over
Micro-	0.64	0.25	**0.38**	0.41	-0.26	Over
Arab: Non-Arab						
Macro-	0.41	0.35	**0.44**	0.52	0.03	Under
Meso-	0.49	0.16	**0.23**	0.29	-0.26	Over
Micro-	0.61	0.40	**0.43**	0.46	-0.18	Over
African: Caribbean						
Macro-	0.20	0.16	**0.23**	0.29	0.03	Under
Meso-	0.29	0.03	**0.13**	0.20	-0.16	Over
Micro-	0.43	0.37	**0.40**	0.43	-0.03	Over

Figure 1. *Ideal–typical segregation patterns: a city with six wards each containing four tracts. (There are 100 persons resident in each tract; the numbers show the number of B residents in each tract.)*

75	75	0	0
75	75	0	0
75	75	0	0
75	75	0	0
0	0	0	0
0	0	0	0

40	40	0	0
60	60	0	0
40	60	0	0
40	60	0	0
60	40	0	0
40	60	0	0

60	0	0	40
0	40	60	0
40	0	40	0
60	0	0	60
0	0	0	0
40	60	60	40

Figure 2. *The (a) D and (b) Dm values for the simulated data set.*

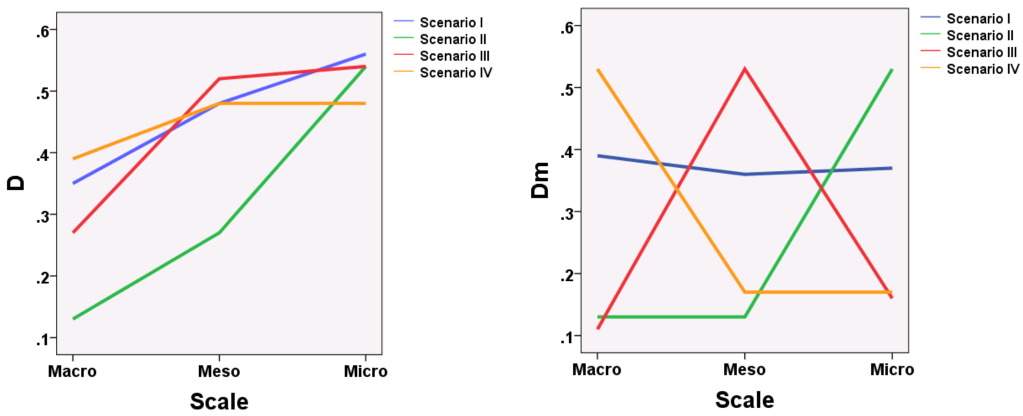

Figure 3. *The (a) D and (b) Dm values for the Leicester data.*