



Delft University of Technology

Modelling residential segregation as unevenness and clustering

A multilevel modelling approach incorporating spatial dependence and tackling the MAUP

Jones, Kelvyn; Manley, David; Johnston, Ron; Owen, D.

DOI

[10.1177/2399808318782703](https://doi.org/10.1177/2399808318782703)

Publication date

2018

Document Version

Accepted author manuscript

Published in

Environment and Planning B: Urban Cities and City Science

Citation (APA)

Jones, K., Manley, D., Johnston, R., & Owen, D. (2018). Modelling residential segregation as unevenness and clustering: A multilevel modelling approach incorporating spatial dependence and tackling the MAUP. *Environment and Planning B: Urban Cities and City Science*, 45(6), 1122-1141. <https://doi.org/10.1177/2399808318782703>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Modelling residential segregation as unevenness and clustering: a multilevel modelling approach incorporating spatial dependence and tackling the MAUP

Kelvyn Jones¹, David Manley^{1,2}, Ron Johnston¹ & Dewi Owen¹

¹Geographical Sciences, University of Bristol, UK

²OTB, TU Delft

Abstract

A model-based approach to measuring residential segregation is further developed by explicitly including spatial effects at multiple scales. This model distinguishes segregation as unevenness and as spatial clustering in the presence of stochastic variation. An accompanying badness-of fit measure allows the identification of the scale and zonation where the spatial patterns come into focus thereby potentially transcending the modifiable areal unit problem. The model is applied to Indian ethnicity in Leicester UK finding segregation as unevenness and as spatial clustering at multiple scales.

Keywords: segregation, modifiable area unit problem, Bayesian modelling, spatial dependence, multiscalar analysis

Introduction

The measurement and understanding of segregation in cities is a major topic in urban geography, for policy makers and for governments alike. Segregation is routinely cast as a major obstacle to a well-functioning society and has been placed centre of the explanation for many social problems including recent UK race riots. However, the measurement of segregation often relies on individual data aggregated into pre-given areal units of varying size and extent. Unfortunately, the spatial mesh that is deployed in this aggregation can define the results we find - the modifiable areal unit problem (MAUP; Openshaw, 1984). Much of the literature has described these changing results as unpredictable, pervasive and unsolvable. However, a smaller body of work views the outcome of the MAUP as a function of genuine spatial structures within the georeferenced data that are aggregated (Manley et al., 2006). Consequently, the MAUP ceases to be intractable but becomes an opportunity to identify the structures of cities; it becomes an analytical tool that can inform.

This paper is not concerned with showing yet again that different zonations at different scales produce different results. Instead, we recast the MAUP as model misspecification in which the spatial structure of the data (at potentially multiple scales) needs to be incorporated into the model. An explicit framework is developed to distinguish patterns from stochastic noise and to compare the badness-of-fit of alternative specifications of spatial structure. This allows the identification of the spatial configurations that bring the patterns into sharpest focus. Moreover, this model distinguishes and assesses segregation as both unevenness and as spatial clustering (Massey and Denton (1988).

We begin by setting out a methodological agenda through the review of previous work; we then consider the data to be analysed as an exemplar- that for self-declared Indians in Leicester in 2011. The existing hierarchical multilevel approach to segregation is considered and extended to the spatial case for multiscale analysis. Results and discussions follow and we conclude with an agenda for investigations of segregation and the MAUP in general.

Setting the methodological agenda

The methodological agenda is informed by the substantive processes that are operating in the city, combined with the nature of the data to be analysed, and the current tools for analysis. Socio-spatial data analysis requires a recognition that observed outcomes are likely to be the result of multiple processes, which can be divided into two broad groups: stochastic and structured. Stochastic processes refer to those unexplainable elements that drive random differences in population residential location whereas the structural refer to processes behind segregated outcomes such as discrimination and exclusion from certain housing markets, and attraction to like-minded neighbours. Even in situations where there is no structural segregation, the population is unlikely to be evenly dispersed across residential space because of stochastic processes. Methodologically, the finer the spatial scale and the fewer the number of people living in an area, the greater the probability for chance to affect results. Many traditional segregation indices assume that there are no stochastic variations and provide upwardly-biased estimates with apparent segregation when there is none (Carrington and Troske, 1997). It is thus necessary to adopt inferential modelling to assess segregation *net* of stochastic variation as pioneered by Leckie et al. (2012). Segregation is characterised as an estimated variance around a mean. If all places have the same share of a subgroup the variance estimate will be zero, and any apparent differences are merely due to chance. However, an estimated variance whose uncertainty intervals do not include zero indicates genuine segregation in the form of unevenness. Importantly they show that several well-known indices are an increasing monotonic function of the variance. Consequently, the same conclusions will result

whichever are used to compare different groups or groups over time. They also provide a method of transforming the variance and accompanying uncertainty into these indices if that helps comparison.

Mobility research finds that movers choose (or are forced to choose) a broad zone of a city and then decide at a micro scale where to locate within that zone (Johnston, Forrest et al. 2016). There can be multiple processes operating at multiple scales and a group can be concentrated in zones and within certain subareas within zones. Methodologically, the scale at which segregation analysis is conducted is critical and can determine the outcomes found. Previous research has resulted in the stylised fact that the most marked segregation is at the finest scale and declines at higher spatial scales. However, there has been much misunderstanding about analysis at multiple scales. Typically (e.g. Lee et al. 2008) the analysis moves through a sequence of ever-growing scales finding the maximum segregation at the finest scale. These analyses are problematic as the finest scale implicitly includes segregation at any relevant higher scale (Duncan et al., 1961) and the apparent decline is likely to be an artefact of the aggregation process whereby data are smoothed as they are combined into larger units. Fortunately, the modelling approach can be extended to analyse multiple scales simultaneously where unevenness is characterised by the variance at one scale *net* of the variation at other scales (and net of stochastic variation). Several studies using this approach have found that segregation is greater at the *macro* scale, as for example for London (Johnston, Jones et al., 2016; Jones et al. 2015).

Processes of structured segregation are likely to produce not just unevenness of distribution but clustering whereby similar groups are spatially concentrated in certain parts of the city and separated from dissimilar groups. Methodologically, however, most segregation analyses are inherently aspatial merely analysing the frequency distribution of minority rates and ignoring location. What is needed is an assessment of the degree of spatial dependence of the rates – the degree to which similar rates are adjacent to each other. The developing modelling approach in partitioning the variance at different levels implicitly models the dependency at each scale below the topmost one (Jones et al. 2015); larger variances indicate both greater differences between and equivalently greater similarity within areas (Bullen *et al.*, 1997). But this crude approach has no explicit parameter for the degree of spatial dependence; the variance measures only unevenness and there is a need for a modelling approach that additionally assesses spatial clustering as well as aspatial unevenness.

Initially, the modelling approach was applied to educational segregation which analyses data for pre-existing higher-level entities (children in classes in schools). For residential segregation, this is not the case and the areas of interest must be defined, thereby confronting the MAUP. There are two aspects to this: the scale problem is that results may depend on the number of units studied; while the zonation problem is that different results can be found for a constant number of geographical units but with differing spatial arrangements. To date the model-based approach has taken a strictly hierarchical approach where the scales and zoning are taken as given and fixed. Thus, a model-based study of segregation in London (Johnston, Jones *et al* 2016) used census Output Areas at the finest scale which are exactly nested in Middle Layer Super Output Areas which are in turn nested in Boroughs. It may be that this specific mesh is determining the results so we need to research the MAUP to gauge potential sensitivity of results to different spatial architectures. This problem is closely bound up with spatial dependence for if the rates are truly without map pattern then differential aggregations and zonings are unlikely to determine the results (Wong 2009, 114).

Views on MAUP range from an 'essentially unpredictable' intractable problem (Fotheringham and Wong, 1991, 1025) to a 'very powerful analytical device' (Openshaw, 1984, 7). Another striking feature is that much research simply concludes that results differ and gives little insight as to why.

There are several reasons for this. Some studies use real data and then create zones from these – the million correlation coefficients of Openshaw and Taylor (1979) are based on re-grouping areal voting data for Iowa. It is then difficult to know what the ‘true’ relationship is and to isolate what is producing the results. Other studies use simulated data of known properties but are unrealistic in not considering different forms of spatial configuration in determining the results (Amrhein, 1995). A more pervasive problem is that studies have not been set in a general modelling framework (Wrigley, 1995) resulting in no goodness-of-fit measure for determining what is the ‘best’ zonal arrangement. Here we see the MAUP as an opportunity to vary scale and zonation to bring the patterns into focus and thereby gain insight to the underlying processes producing segregation. However, we lack a statistic to provide guidance when this focus has been achieved. This is especially difficult in real data as there may be multiple and unknown configurations where spatial processes are operating; working at a single scale may miss what is going on.

Three other issues are methodologically important for us. Firstly, in assessing change there is a need to incorporate statistical uncertainty. A decreasing population per unit area will result in smaller numbers which will emphasize stochastic variation upwardly biasing segregation estimates. Such analysis of change must be multiscalar as countervailing tendencies (polarization at one scale alongside dispersal at another) may cloud the picture of what is happening (Haggett, 1965). Secondly, ethnicity in the 21st city is not a binary variable, with multiple ethnicities characterising contemporary society. This diversity requires the simultaneous modelling of multiple groups which results in smaller numbers as ethnicity is disaggregated. Thus, the composition of the urban environment makes a modelling approach particularly valuable. The third aspect is that there may be multiple influences on location decision-making so that residential segregation may be differently driven by ethnicity and class. It is vital to have a methodology for assessing which are the key influences on segregation and to do so for multiple groups at multiple scales. We have addressed all three in previous work (Johnston, Jones et al., 2016; Jones et al. 2017) and here we concentrate on spatializing the modelling approach to measure clustering while examining sensitivity to the MAUP for a single time cross-section for just two ethnic groups.

Our agenda should now be clear. A modelling approach is required that can handle and separate stochastic variation from ‘true’ pattern. It must work at multiple scales simultaneously distinguishing unevenness and spatial clustering. It should allow sensitivity analysis to assess which alternative spatial arrangements bring the patterns into focus at one scale net of others. This needs to be set in an inferential framework where the uncertainty of parameters is evaluated and where appropriate measures of goodness of fit allow the preferment of some spatial arrangement over others. The method of model comparison must consider model complexity in coming to a parsimonious judgement on models with many potential parameters. The method must also be computationally feasible on realistically-sized data sets.

Setting and data

As an exemplar for this work we have chosen one UK city – Leicester – as our setting and one population group – those self-identified as Indian at the 2011 Census – as our one ethnic group (the largest non-White group in the city at 28.3%) compared to all other ethnic groups including the White majority. The city is a medium-sized one with a population of some 330,000 located in the English East Midlands.¹ Those of Indian ethnicity are clustered in the eastern parts of the city (Byrne,

¹ Hennerdal and Nielsen (2017) contend that there is an element of the MAUP due to the choice of area of reference; we can overcome this in the modelling approach by constraining the overall mean of evenness to say the all England value instead of the overall value for Leicester.

1998). Ethnic counts are available for census Output Areas which are created post enumeration to be meaningful units based on homogeneity of housing characterises using the AZTool (Martin,2002). These OAs are very fine-grained with a median of 330 individuals. We have further created a higher geography of Zones using the same AZTool based on a target population of 8500. The zones are composed of contiguous OAs but the proportion of Indians has not been used in their definition and represent one possible macro geography of the city. Their median population size is 8450 and they contain on average 25 OAs. At the outset, we have a three-level strict hierarchy with individuals at level one nested in 969 OAs nested in turn within 39 Zones at level 3.

Classification diagrams of the structures to be analysed

To implement the modelling approach, we use multilevel modelling (Duncan and Jones, 2000) and the levels involved are most easily conveyed in a classification diagram (Browne et al. 2001). Figure 1 gives three types of structure: (a) is a strict hierarchy with three levels: individuals are nested in OAs which are nested in Zones. Alternatively, Zones classify OAs and OAs classify individuals. An enclosing box signifies a classification and a single arrow indicates a strict hierarchy where a lower level OA unit belongs to one only higher-level unit. This strict hierarchy has been a basis of modelling segregation since Leckie et al. (2012). Figure (b) represents a non-strict hierarchy. Individuals are nested in two separate classifications. In one individuals are strictly nested in one classification of their OA of residence while the second is a neighbouring ‘patch’ of surrounding OAs say the nearest three (not including the focal OA). The cross-classification is indicated by a lack of linkage between boxes and the double arrows indicate multiple membership. The patch is akin to a ‘moving window’ that hovers over each lower level unit to form a bespoke neighbourhood. defined by the analyst. Figure 1c represents multiple processes operating at different scales. Individuals belong to three separate classifications. A strict hierarchical relationship exists as individuals are nested in their own OA and two multiple membership relations; the local patch based on the nearest three OAs; and the more zone-like arrangement where individuals belong to a seven OA grouping. Figure 1a is space invariant while b and c are explicit spatial structures if the patches are defined on contiguity.

Model specification

The specification of the strictly hierarchical model

We begin with an aspatial model of the form that has been used in recent work. It has three classifications – individuals signified by the subscript i which are classified by the OA in which they are resident ($OA(i)$) and by the larger zonal area ($Zone(i)$). The dependent variable (Y_i) is the observed binary individual outcome of Indian origin or not. This is treated as an underlying Binomial distribution with the underlying proportion of occurrence of π_i . It is not this value but it is the expectation of logit of this value $E(\log_e(\pi_i/(1 - \pi_i)))$ that is modelled. The logit is logarithm of the ratio of the probability of self-declaring as an Indian to the probability of being a non-Indian. This prevents impossible predicted proportions outside 0 to 1 and makes the Normality assumption for Zone and OA differences more likely to be fulfilled although results are typically robust (McCulloch and Neuhaus, 2011).

$$\begin{aligned}
 Y_i &\sim \text{Binomial}(n_i, \pi_i) \\
 E\left(\log_e\left(\frac{\pi_i}{1 - \pi_i}\right)\right) &= \beta_0 + \mu_{Zone(i)}^{(3)} + \mu_{OA(i)}^{(2)} \\
 \mu_{Zone(i)}^{(3)} &\sim N(0, \sigma_{\mu^3}^2) \\
 \mu_{OA(i)}^{(2)} &\sim N(0, \sigma_{\mu^2}^2)
 \end{aligned}$$

$$Var(Y_i|\pi_i) = \frac{\pi_i(1 - \pi_i)}{n_i}$$

Prior specifications

$$p(\beta_0) \propto 1$$

$$p(1/\sigma_{\mu^3}^2) \sim \Gamma(0.001, 0.001)$$

$$p(1/\sigma_{\mu^2}^2) \sim \Gamma(0.001, 0.001)$$

These log-odds have an overall mean (β_0) plus a differential for Zones ($\mu_{Zone(i)}^{(3)}$) and for OAs within Zones ($\mu_{OA(i)}^{(2)}$). Importantly the OA differential is thus net of the Zone differential. A positive value for both represents an OA with a high log-odds of Indians compared to its Zone that is also high compared to the city; two negatives indicate a low OA Indian ethnicity in a Zone of low ethnicity. A zero indicates the Zone is the same as the city average; a zero for an OA indicates a typical OA within a Zone. Assuming these differentials are Normally distributed they are completely summarised by variances so that $\sigma_{\mu^3}^2$ represents the between-Zone while $\sigma_{\mu^2}^2$ represents the within-Zone between-OA differences. These are our primary measure of segregation; if there is no segregation the variance for a classification will be zero. At the lowest classification, there is a Binomial distribution with the variance dependent on the modelled rate (π_i) and its denominator n_i which is the OA Indian plus non-Indian total. In practice, there is the same set of units – the OAs – at level 1 and 2 (each level 2 unit is composed of exactly one level 1 unit). This views the aggregate proportions at level 2 as consisting of replicated binary responses for individuals at level 1 (Browne et al., 2005). This allows the separation of variation into exact Binomial at level 1 and over-dispersion at higher levels so that the higher-level variances summarize ‘true’ differences between areas in excess of that from chance. The model is completed by the specification of the prior distributions (see later).

These variances in ratio form give the degree of similarity of outcome within an area. The intra-Zonal correlation is the degree of dependence between individuals within the same Zone:

$$\frac{\sigma_{\mu^3}^2}{\sigma_{\mu^3}^2 + \sigma_{\mu^2}^2 + \sigma_e^2}$$

where σ_e^2 is the Binomial variance of 3.29, the variance of a logistic distribution (Jones and Subramanian, 2013). A high value for this ratio implies that if you picked pairs of people at random from the same Zone (a clustered sample therefore) if one of the pair was Indian the other of the pair would also likely be Indian. The intra-OA correlation, the similarity in outcome between individuals within the same Zone and OA, is

$$\frac{\sigma_{\mu^3}^2 + \sigma_{\mu^2}^2}{\sigma_{\mu^3}^2 + \sigma_{\mu^2}^2 + \sigma_e^2}$$

Finally

$$\frac{\sigma_{\mu^3}^2}{\sigma_{\mu^3}^2 + \sigma_{\mu^2}^2}$$

gives the similarity of OAs within the same Zone. A high value indicating that knowing the proportions of Indian ethnicity in one OA in a Zone is informative of the proportion in another OA in the same Zone. This hierarchical specification models the degree of dependence (autocorrelation)

but rather rudimentarily. The OAs are clustered within their Zone but the Zonal analysis is aspatial for there is nothing that accounts for which Zones are contiguous. Moreover, while OAs are defined in terms of the Zone to which they belong there is no further taking account of their spatial arrangement – adjacent OAs could be in different Zones thereby imposing an unrealistic rigid boundary. These models assume that the results are invariant to location (Elffers, 2003). We can rearrange Zones and OAs within Zones without affecting the variances which measure the degree of segregation. These models therefore essentially assess segregation as departures from evenness albeit decomposed into various (spatial) scales.

The specification of the multiple membership cross-classified multilevel

Spatial segregation models can be specified as multiple membership cross-classified models (Fielding and Goldstein, 2006) where rigid zones are replaced by flexibly-defined patches. In this example individuals are nested in an OA, in a small-scale patch of three OAs and a larger neighbourhood of seven OAs.

Opmerking [RJ1]: Aren't they really four and eight, including the OA at the focus of each neighbourhood?

$$Y_i \sim \text{Binomial}(n, \pi_i)$$

$$E\left(\log_e\left(\frac{\pi_i}{1-\pi_i}\right)\right) = \beta_0 + \sum_{j \in 7NHood(i)} w_{i,j}^{(4)} \mu_j^{(4)} + \sum_{j \in 3NHood(i)} w_{i,j}^{(3)} \mu_j^{(3)} + \mu_{OA(i)}^{(2)}$$

$$\mu_{7NHood(i)}^{(4)} \sim N(0, \sigma_{\mu^4}^2)$$

$$\mu_{3NHood(i)}^{(3)} \sim N(0, \sigma_{\mu^3}^2)$$

$$\mu_{OA(i)}^{(2)} \sim N(0, \sigma_{\mu^2}^2)$$

$$\text{Var}(Y_i | \pi_i) = \frac{\pi_i(1-\pi_i)}{n_i}$$

Prior specifications

$$p(\beta_0) \propto 1$$

$$p\left(1/\sigma_{\mu^4}^2\right) \sim \Gamma(0.001, 0.001)$$

$$p\left(1/\sigma_{\mu^3}^2\right) \sim \Gamma(0.001, 0.001)$$

$$p\left(1/\sigma_{\mu^2}^2\right) \sim \Gamma(0.001, 0.001)$$

The response remains a binary outcome (Indian or not) modelled as the underlying log-odds with Binomial level-one variance depending on the total and the modelled proportion in an OA. There are three sets of higher-level departures from the overall mean (β_0) so that $\mu_j^{(4)}$ are the differentials for the 7-member patch for a particular $OA(i)$; $\mu_j^{(3)}$ are the differentials for a 3-member patch, while $\mu_{OA}^{(2)}$ is the (aspatial) differential for each OA. There are three sets of effects for each OA arising from three distinct classification sources. Moreover, all three are estimated simultaneously so all are net of each other and Binomial variation.

The use of \in set notation signifies that an OA is an element of the wider patch with the notation conveying that each differential is a weighted sum of a set of random effects. The weights are defined exogenously, typically constrained to sum to one ($\sum w_{i,j} = 1$) and represent the presumed degree of connectivity between OAs in the patch. This could be the inverse squared distance between the focal OA and its neighbours (emphasizing a rapid decline of influence) or simply equal

weights based on the number of members ($1/m$).² To illustrate the latter when m is 7 and 3, the weights will be 0.14 and 0.33 respectively, and the equation for OA1 with its surrounding OAs (with identifiers 2 to 8) is:

$$E(\text{Logit}(Y_1)) = \beta_0 + 0.14 * \mu_1^{(4)} + 0.14 * \mu_2^{(4)} + \dots + 0.14 * \mu_7^{(4)} + 0.33 * \mu_1^{(3)} + 0.33 * \mu_2^{(3)} + 0.33 * \mu_3^{(3)} + \mu_{OA(1)}^{(2)}$$

revealing the computational complexity involved as this is replicated for all 969 OAs.

The differentials are summarised by variances with $\sigma_{\mu_2}^2$ being the unstructured aspatial segregation due to unevenness around the mean, while $\sigma_{\mu_4}^2$ and $\sigma_{\mu_3}^2$ summarise the spatially-based differences of an OA belonging to differently-sized patches. If there is unevenness but no local clustering $\sigma_{\mu_2}^2$ will be nonzero but the two other variances will not. Moreover, because we can specify the number and size of the patches we can change the focus to see at what scale there is noticeable spatial patterning and identify multiscale segregation. The total variance is decomposed into a set of additive components which allows the calculation ratios summarising dependence as for the hierarchical model. However, the nature of the weights must be considered (Fielding and Goldstein, 2006). Taking the 3-unit patch with equal weights the spatial variance is $\sum w_{i,j}^2 \sigma_{\mu_3}^2$ so the sum of the squared weights is involved. Variances as estimated will be too large and cannot be directly compared. However, with an equal number of members (m) that form the patch and equal weights, this can be overcome by dividing by variance by the number of members $\frac{\sigma_{\mu_3}^2}{m}$.

Estimation and model comparison

Likelihood methods could be used (Jones et al., 1998) but this is cumbersome compared to Fully Bayesian MCMC estimation (Browne, 2017). MCMC uses a building-block approach so that an additional set of additive terms (e.g. another patch) can be accommodated without extensive re-writing of the algorithm; a cross-classified model is no more complex than a hierarchical one. Moreover, the inclusion of weights associated with the additive terms does not involve the inversion of the full (969*969) matrix but just those defining each patch. The estimates can be expected to be good ones (Browne and Draper, 2006) as uncertainty in one parameter is considered in estimating all others thereby reducing bias. The distribution of the key variance terms is not assumed to be Normal (unlikely as variances cannot go below zero) with MCMC providing Bayesian credible intervals which give say the 95% probability that the parameter falls between the lower and upper bounds which may be asymmetric as the distribution of estimates may be positively skewed.

An important MCMC by-product is the Deviance Information Criterion (Spiegelhalter et al. 2002). Brunson (2016) argues that this is a different approach to inference, being about model selection not hypothesis testing; there may be no winner but a shortlist of plausible models. Brunson is writing about the likelihood-based AIC which defines model complexity as a function of the number of parameters.³ Our situation is more complex for while the mean and variance parameters as usual are equivalent to consuming one degree of freedom, the differentials also need to be counted but may not each contribute a whole value as they come from a common distribution. With MCMC the effective degrees of freedom is calculated as part of the model fitting process. Thus, it is possible to estimate a non-nested set of models including hierarchical and cross-classified models for the same

² Distance decay weights may be more useful to achieve spatial smoothing (Lawson et al., 2003).

³ The AIC is used by Hirschfield et al. (2014) and Nakaya (2000) in spatial modelling.

outcomes and derive a badness-of-fit measure penalized for model complexity. According to developing practice (Jones and Subramanian, 2017), any model with a lower DIC is an improvement but a model with a difference of 2 still has substantial support and should be kept under consideration. A reduction of 4 suggest that the worse-fitting model has considerably less support, while a difference of over 10 suggest that the model with a higher DIC can be ignored.

In Bayesian modelling the posterior distribution characterises the degree of support for different estimate values. It is obtained by starting with a prior initial guess of the distribution of the estimates and combining this with the likelihood information from the data. The posterior is highly complex as it is the joint distribution of all parameters (means, variances, differentials). Estimation works by making a simulated draw from the marginal distribution of one parameter and feeding this through into simulated draws for other parameters thereby taking account of the full uncertainty of all parameters. The earlier model specifications use weak priors to maximize the influence of the data. The overall mean (β_0) is specified as a uniform distribution so that any value is equally likely. The reciprocal of the variances is assumed to follow a positively skewed Gamma distribution with both the scale and shape parameters set to relatively uninformative values.

In practice, all the models were estimated in MLwiN (Charlton et al. 2017) which is considerably faster than alternatives (Li et al. 2011).⁴ To initiate simulation quasi-likelihood estimates were used as starting values. This was followed by a discarded burn-in of 5000 draws (to escape potentially biased likelihood estimates) followed by a further 250,000 draws for each parameter to characterise the posterior. The trajectories of these draws were inspected to see that there was no trending (that is failure to converge to the equilibrium posterior distribution) and that the effective sample size of each set of posterior estimates was at least equivalent to 750 independent draws. The 2.5 and 97.5 percentiles of the posterior distributions were used as 95% credible intervals while the mean was used for the point estimates.

Results

This study adopts an exploratory approach because of little guidance on the 'right' scale and zonation. Ultimately, we are motivated by earlier research findings that show macro segregation is the norm when using strict hierarchies; a finding that goes against much previous understanding. We begin by applying hierarchical models to real and simulated data then use spatial cross-classified models to examine different scales and zonations as well as multiscale patterning.

Results of the hierarchical models

Table 1 provides the set of results for the strictly hierarchical models. The first major column represents three models for the observed data: null model (no parameter for higher-level differences); a two-level model (individuals within OAs); a three-level one (OAs additionally nested within Zones). The changes in the DIC show that going from a model without unevenness to one with OA differences and then to additional Zonal differences represents a very substantial reduction in the badness-of-fit. The mean estimate for the null model is -0.93 on the logit scale and when converted into percentages indicates that across Leicester some 28% are Indian. In the 3-level model the mean estimate is -1.54 and when converted into percentages, 18% of the adult population is Indian in the median OA in the median Zone while the value is 27% in the mean area. These two different values (the cluster-specific and population-average estimate; Jones and Subramanian, 2013) reflect the positively skewed nature of the underlying uneven modelled rates; areas with distinctively high rates pulling the mean upwards from the median.

⁴ Estimation took some 20 minutes for 250,000 simulations for a spatial cross-classification with 50 members on a Windows based PC (Intel i7 at 3.10Ghz, 16GB RAM, 64 bit).

In the two-level model the between OA variance is large (3.44) on the logit scale (equivalent to a D index of 55% meaning that 55% of Indians would have to move OAs to achieve city wide evenness).⁵ This also implicitly includes the between-Zone variance. Indeed, when the Zones are included the majority, 77%, of the variances lies between Zones and only 23% lies between OAs within Zones. These equate to a D index of 51% and 33% respectively. If an incorrect two-level model is fitted, the differences as predicted by Tranmer and Steel (2001) accrue to the lowest included level, the OAs, incorrectly inflating micro-scale segregation. It is the property that has misled analysts about the true scale of macro segregation. These variances allow calculation of intra-unit correlations: the degree of dependence between individuals in terms of Indian ethnicity within the same Zone is 0.40; the similarity between individuals in the same Zone and OA is 0.51; while the typical similarity of OAs within the same Zone is 0.77. There is substantial segregation at both scales but it is the macro geography that is particularly important

Opmerking [RJ2]: We never define D! Not at all clear how you get from a 77:23 split in the variance to a 51:33 ratio in D values?!

The second major column of Table 1 is for simulated data with the same overall Indian ethnicity rate of 28% as the observed data but no genuine differences between OAs and Zones. That does not mean however that there are no differences between the observed OA rates (they range from 20 to 40%) as the data have been created to have Binomial stochastic variation driven by the true varying denominator of the OA. The two and three-level model estimate this correctly with a variance at each level of close to zero. Moreover, the extra complexity of both results in a substantially worse DIC (+34 and +7); OA and Zonal differences are not needed. The logit estimate is the same for all models and this converts to a mean and median of 28% showing no effective segregation. These hierarchical models correctly identify when there is no genuine segregation as unevenness even in the presence of apparent differences based on chance.

The results from the multiple membership cross-classified models

To explore the spatial nature of the structures within the data we have used an approach defining moving windows of 3, 5, 7, 10, 15, 20, 30, 40 and 50 OAs for each of the 969 Leicester OAs. Specifying a cross-classified model, it is necessary to use weights for each OA and given the lack of strong prior information we have used equal weights, which depend on the numbers of members defining the patch (1/m). For each of these scales we have used three zonations to define membership: straight-line distance to form compact patches; fully random without contiguity so that OAs forming the neighbourhood could be anywhere in the city; and 'random distance'. In the latter, for each focal OA a candidate list based on contiguity is produced and one of these is chosen randomly, continuing this process until a zonation based on 50 OAs is achieved. This means that there could be tightly focussed neighbourhoods or more elongated ones.

Given the volume of estimates we have focused on the DIC and the variances choosing to display these graphically so that we can more readily compare. Figure 2a shows the change in the DIC from a model with only unstructured effects (two-level hierarchical) when spatial models with larger and larger patches are included one at a time. None of the random zonation models show any reduction in the badness-of-fit and all the spatial variances have a modal value of zero. The results are effectively the same as a model without any potential spatial effects. In contrast both sets of models where patches are distance-based show a better fit and this is especially the case for macro patches based purely on distance. These results are important in confirming that neighbourhood definition –

⁵ D was calculated using the methods of Leckie *et al* (2012). Another way assessing the logit variance is to transform it into a Median Odd Ratios (Jones *et al* 2015). The MOR can be conceptualised as the increased odds of being and Indian (on average, hence the median) that would result from moving from a lower to a higher area if two areas were chosen at random from the distribution with the estimated variance. A MOR of 1 is no segregation; here the value is 5.87 which is very large.

scale and zonation - is crucial to evaluating spatial segregation and that the DIC can distinguish between alternatives. The underlying similarity of the results suggests the resilience of the findings under different non-random re-arrangements and the importance of explicitly modelling spatial segregation.

The variances and credible intervals for the different size of patches for the best-fitting distance-based model are given in Figure 2b. The initial model without any spatial patches misattributes the variance, substantially overestimating the unstructured OA variance representing unevenness. When varying sized patches are included, the unstructured variance reduces substantially and does not change a great deal whatever size of patch is included. The credible intervals of this variance remain very tight so there is strong evidence of unevenness. When patches are included the spatial variance is large and the credible intervals do not approach the zero value of no clustering; there is strong evidence of spatial dependence in the rates.

As these are one-scale-at-a-time analyses the smaller scale patches may reflect misattributed higher-level variance (Tranmer and Steel, 2001). To address this Figure 3 focuses on the potential multiscale nature of the segregation retaining the unstructured OA variance and the 40 or 50 zone distance-based patches (with similar DICs) in two sets of models which now additionally include smaller patches composed of the 3,5,7 and 10 nearest OAs in terms of distance. In general, and despite its complexity the 50-member model provides the best DIC, additionally including a 5-member patch results in a substantially worse fit, while the best-fitting model of all is one with the additional 7 membership patch. Examining the variances in Figures 3b and 3c for the 40 and 50 neighbourhood models there remains unstructured variance at the OA level beyond Binomial variation; the uncertainty intervals do not approach zero. However, in a model without a small patch the unstructured variance is overestimated and declines when local spatial clustering is included. Most noticeably the largest effect is the spatial segregation at the macro scale and while this attenuates somewhat as a small neighbourhood differentials are introduced, it remains substantially the largest source of segregation. The small neighbourhood patch is also important and we can see that in a model with 7 members this spatial segregation is greater than the unstructured variance but not as important as the macro segregation. These results confirm multiscale segregation: Indians are spatially segregated into large scale areas of Leicester and within these macro areas they are spatially segregated **again**.

Conclusions

In methodological terms we have demonstrated a robust approach to identify multiscale spatial and aspatial segregation. The approach is similar to Lee et al. (2008) in empirically defining neighbourhoods at different scales but distinguishes segregation as unevenness and as spatial clustering net of each other giving estimates at multiple scales simultaneously. The multilevel approach with the innovation of multiple neighbours at more than one scale explicitly models spatial heterogeneity. As the MAUP is produced by changing spatial dependence as areas are combined under different scale and zonations (Wong, 1996), this approach explicitly models that change and so is intrinsically resistant to it.⁶ A complexity-penalized goodness-of-fit measure is used to assess alternative spatial arrangements. This allows refocussing the spatial lens to identify maximal differences between and greatest similarity within and we follow Moellering and Tobler (1972,36) in arguing that this is the geographical arrangement where spatial processes are in 'action'. However, in comparison to their pioneering work we go beyond hierarchical models in allowing each areal unit

⁶ Similarly, Fotheringham *et al.* (2002,144–158) argue that Geographically Weighted Regression is resistant to MAUP as it models spatial heterogeneity.

Opmerking [RJ3]: Here is my main query. All that this conclusion does is confirm our earlier multi-scale work, relevant to the aggregation component of the MAUP. But it says nothing about the zonation component: is the conclusion that different zonations have little or no impact? If so we should say so, or we have not taken the work very far forward empirically – and of course such a conclusion applies to the one case only.

But if I have got it wrong, what is the conclusion regarding the zoning issue – surely we have one?

Opmerking [DM4]: Think I agree ... we do need to mention the zonation as well. We have three (?) different ways of construction zones at the same scale.... So we can say something cutting the results in that dimension as well? Something like:

Once a scale of process has been identified the question of zonation remains. Here, the multiple methods through which the zonations tested were constructed highlighted the importance of linking the boundary divisions to the process as well. Where space was taken into account the results differed demonstrating that zonation within space relies on the spatial structure as heavily as the scale question

to have its own (and differentiated) zonation that explicitly takes spatial dependence into account while also dealing with the discrete outcome of the response variable and the need to take account of varying reliability across the map due to varying denominators. Consequently, the unstructured and spatially-structured variances are net of inherent stochastic variation.

In substantive terms, we have confirmed previous findings on multiscalar segregation at micro, meso and macro scales obtained with strictly-hierarchical models with rigid boundaries. Moreover, and unlike much previous methodologically-compromised results, we have found that the greatest segregation is at the highest and not the lowest scale. The spatial variance terms are important and even higher than the aspatial effects which also remain important. Local neighbourhoods based on straight-line distance are particularly effective in capturing the spatial character of the segregation. There is strong evidence that Indians in Leicester are unevenly distributed and spatially concentrated in certain parts of the city. Indians are clustered macro parts of the city and within those large areas they are clustered again. There is clear scope here to further explore the linkages between residential and neighbourhood choice with the literature on [segregation](#).

This approach sets a large methodological and substantive agenda. A key question is whether the approach can distinguish between different forms of spatial segregation (e.g. ethnoburbs versus classical definitions such as Hoyt and Burgess) under realistic conditions of stochastic 'noise' and confounded variation at multiple scales. The implementation has used specialist software and so an important next step to make these models more generally available and we plan to use the universal statistical software gateway of Stat-JR (<http://www.bristol.ac.uk/cmm/software/statjr/>). The DIC has been undeniably helpful in narrowing the choice of appropriate models but it would be even more useful if it could be more diagnostic between related models. The problem is that the DIC is a single overall goodness-of-fit and it would be benefit from having a separate 'focus' (Spiegelhalter et al. 2014) on both spatial heterogeneity and unevenness.

The model is capable of extension in several ways including the analysis of multiple groups, changes over time and simultaneously analysing multiple sources of segregation (such as class and ethnicity). Moreover random-coefficient models (Jones, 1991) allow the variance to be a function of observed variables and this permits the assessment of the degree to which both the unstructured and spatial variance are influenced by characteristics of people and places moving towards a more 'explanatory' account. It is also possible to include further classifications so that, for example, an analysis of residential and school segregation of children may include a strict hierarchy (children nested in a school), crossed (children belong to neighbourhoods and schools but not everyone from the same neighbourhood goes to the same school) and multiple membership relations (over time children may attend multiple schools and may have lived in multiple neighbourhoods). The random coefficients model could then include differential segregation by child, school and residential characteristics in a model that would examine the changing multi-layered dynamics of segregation in our society. Simpler models may misattribute variance and wrongly characterise what is going on.

Finally, an important limitation is that the OAs as are taken as given and individuals at the lowest level are nested in these pre-existing modifiable units.⁷ Although computationally challenging a possible extension using egohoods (Omer and Benenson, 2016) to assess household-level segregation at the finest possible scale. This would open the possibility of census agencies

⁷ In our defence OAs are meaningful entities designed to maximize within-area similarity, but not specifically for ethnicity (Cockings et al., 2009). In the fitted models having individual data would be no more informative than the proportions (Subramanian *et al.*, 2001).

Opmerking [RJ5]: Again, it would be good to have an empirical conclusion regarding zonation!~

developing bespoke zonations for specific variables to maximize geographical differences between areas. This would identify appropriate scales to assess and display incidence rates and risks (Nakaya, 2000) for if the analysis is too fine the rates will be unstable due to inherent stochastic variation but if the areas are too coarse the results may be over-smoothed and important patterning is lost. It would also allow the development of models with multiple evaluated contexts to tackle the uncertain geographic context problem (Kwan, 2012) of identifying the 'true causally relevant' geographical setting for individual outcomes. Thus 'bespoke' neighbourhoods (Propper et al.,2005) would not to be imposed exogenously but defined adaptively as part of the model building process.

References

- Amrhein C (1995) Searching for the elusive aggregation effect: evidence from statistical simulations. *Environment and Planning A* 27(1):105-119.
- Browne WJ (2017) *MCMC Estimation in MLwiN 3.0*. University of Bristol: Centre for Multilevel Modelling.
- Browne WJ and Draper D (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* 1(3):473-550.
- Browne, WJ., Goldstein, H and Rasbash, J (2001) Multiple membership multiple classification models. *Statistical Modelling* 1(2):103-24.
- Browne WJ, Subramanian SV Jones K and Goldstein H (2005) Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A* 168 (3):599-613.
- Brunsdon C (2016) Quantitative methods II Issues of inference in quantitative human geography. *Progress in Human Geography*. DOI:10.1177/0309132516648020.
- Bullen NI, Jones K and Duncan C (1997) Modelling complexity: analysing between-individual and between-place variation. *Environment and Planning A* 29(4):585-609.
- Byrne D (1998) Class and ethnicity in complex cities – the cases of Leicester and Bradford. *Environment and Planning A* 30(4):703-720.
- Carrington WJ and Troske KR (1997) On measuring segregation in samples with small units. *Journal of Business and Economic Statistics* 15(4):402-429.
- Charlton CJ, Rasbash J. Browne WJ Healy M and Cameron B (2017) *MLwiN Version 3.00*. University of Bristol: Centre for Multilevel Modelling.
- Cockings S, Harfoot, A and Hornby D (2009) Towards 2011 output geographies *Population Trends* 138(1):38-49.
- Duncan C and Jones, K (2000) Using multilevel models to model heterogeneity. *Geographical Analysis* 32(4):279-305.
- Duncan OD, Cuzzort RP and Duncan B (1961) *Statistical geography: Problems in Analyzing Areal Data*. Glencoe,IL:The Free Press.
- Elffers H (2003) Analysing neighbourhood influence in criminology. *Statistica Neerlandica* 57(3):347–367.
- Fielding A and Goldstein H (2006) Cross-classified and multiple membership structures in multilevel models. *DES Research Report No791*. University of Birmingham.
- Fotheringham AS, Brunsdon C and Charlton M (2002) *Geographically weighted regression*. Chichester: Wiley.

- Fotheringham AS and Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* 23(7):1025-44.
- Haggett P (1965) Scale components in geographical problems. In: Haggett P and Chorley RJ (eds) *Frontiers in geographical teaching*. London:Methuen.pp.164–185.
- Hennerdal P and Nielsen MM (2017) A multiscale approach for identifying clusters and segregation patterns that avoids the MAUP. *Annals of the American Association of Geographers* 107(3):555-574.
- Hirschfield A, Birkin M, Brunson C, Malleson N and Newton A (2014) How places influence crime: the impact of surrounding areas on neighbourhood burglary rates in a British city. *Urban Studies* 51(5):1057-1072.
- Johnston, RJ, Forrest J, Jones K and Manley D (2016) The scale of segregation: ancestral groups in Sydney. *Urban Geography* 37(7):985-1008.
- Johnston RJ, Jones K, Manley D and Owen (2016) Macro-scale stability with micro-scale diversity: modelling changing ethnic minority residential segregation. *Transactions of the Institute of British Geographers* NS41(4):389-402.
- Jones K (1991) Specifying and estimating multi-level models for geographical research. *Transactions of the Institute of British Geographers* NS16(2):148-59.
- Jones K, Gould MI, Watt R, (1998) Multiple contexts as cross-classified models. *Geographical Analysis* 30(1):65-93.
- Jones, K, Johnston RJ, Manley D, Owen D and Charlton CJ (2015) Ethnic residential segregation: a multi-level, multi-group, multi-scale approach. *Demography* 52(6):1995-2019.
- Jones, K, Johnston RJ, Forrest J, Charlton CJ and Manley (2017) Ethnic and class residential segregation: exploring their intersections a multilevel analysis of ancestry and occupational class in Sydney, submitted
- Jones K and Subramanian VS (2013) *Developing multilevel models for analyzing contextuality, heterogeneity and change using MLwiN Volume 2*, University of Bristol: Centre for Multilevel Modelling.
- Jones K and Subramanian VS (2017) *Developing multilevel models for analysing contextuality, heterogeneity and change using MLwiN 3.0, Volume 1*. University of Bristol: Centre for Multilevel Modelling.
- Kwan MP (2012) The uncertain geographic context problem. *Annals of the Association of American Geographers* 102(5):958-968.
- Lawson, AB, Browne WJ and Roderio CLV (2003). *Disease Mapping with WinBUGS and MLwiN*. Chichester: Wiley.
- Leckie, G, Pillinger R, Jones K and Goldstein H (2012) Multilevel modelling of social segregation. *Journal of Educational and Behavioral Statistics* 37(1):3-30.

- Lee, BA Firebaugh G Matthews SA Reardon SF Farrell, CR. and O’Sullivan D (2008) Beyond the census tract: patterns and determinants of racial residential segregation at multiple scales. *American Sociological Review* 73(5):766-91.
- Li, B Lingsma HF Steyerberg EW and Lesaffre E (2011) Logistic random effects regression models: a comparison of statistical packages. *BMC Medical Research Methodology* 11(1):77.
- Manley, D Flowerdew, R and Steel D (2006). Scales, levels and processes: studying spatial patterns of British census variables. *Computers, environment and urban systems*. 30(2):143-60.
- Martin D (2002) Geography for the 2001 Census in England and Wales. *Population Trends*. 108:7-15.
- Massey, DS and Denton NA (1988) The dimensions of residential segregation. *Social forces* 67(2):81-315.
- McCulloch, CE and Neuhaus JM (2011) Misspecifying the shape of a random effects distribution. *Statistical Science* 26(3):388–402.
- Nakaya, T (2000) An information statistical approach to the modifiable areal unit problem in incidence rate maps. *Environment and Planning A*. 32:91-109.
- Moellering H and Tobler WR (1972) Geographical variances. *Geographical Analysis* 4(1):34-50.
- Omer I and Benenson I (2016) Investigating fine-scale residential segregation by means of local spatial statistics. *Geography Research Forum* 22:41-60.
- Openshaw S (1984) *The Modifiable Area Unit Problem*. Norwich: Geobooks
- Openshaw S and Taylor PJ (1979) A million or so correlation coefficients: three experiments on the MAUP. In: Wrigley N (ed), *Statistical Applications in the Spatial Sciences*, pp.127–144. London:Pion.
- Propper, C Jones K Bolster A Burgess S Johnston RJ. and Sarker R (2005). Local neighbourhood and mental health: evidence from the UK. *Social Science and Medicine*. 61(10):2065-2083.
- Spiegelhalter, DJ Best NG Carlin BP and Van der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* 64(4):583–640.
- Spiegelhalter, DJ Best NG, Carlin BP and Van der Linde A (2014) The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society B* 76(3):485-493.
- Subramanian, SV Duncan C and Jones K(2001) Multilevel perspectives on modelling census data *Environment and Planning A* 33(3):399-417.
- Tranmer M and Steel DG (2001) Ignoring a level in a multilevel model: evidence from UK census data. *Environment and Planning A*, 33(5):941–948.
- Wong, DWS (1996) Aggregation effects in georeferenced data. In: Arlinghaus SL and Griffith DA (eds) *Practical Handbook of Spatial Statistics*. Boca Raton: CRC Press, pp.83–10.,

Wong, DWS (2009) The modifiable areal unit problem. In: Fotheringham, AS and Rogerson, PA (eds) *The SAGE Handbook of Spatial Analysis*. SAGE, pp.105-124.

Wrigley, N (1995). Revisiting the modifiable areal unit problem and the ecological fallacy. In Cliff AD Gould PR Hoare AG and Thrift NJ (eds). *Diffusing geography*. Oxford: Blackwell.

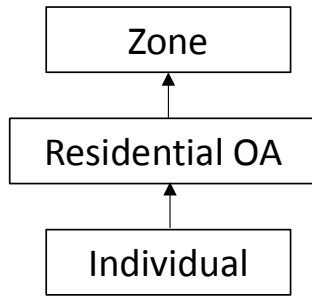


Figures

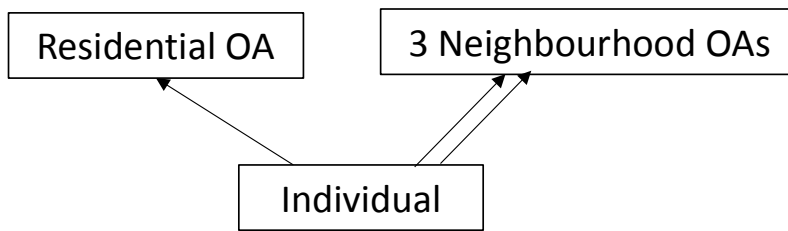
Figure 1 Schematic classifications of the structures used in the analysis

a) Three level strict hierarchy b) multiple membership with two cross classifications c) multiple membership with three cross classification.

a)



b)



c)

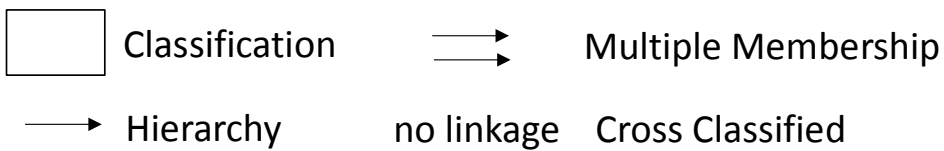
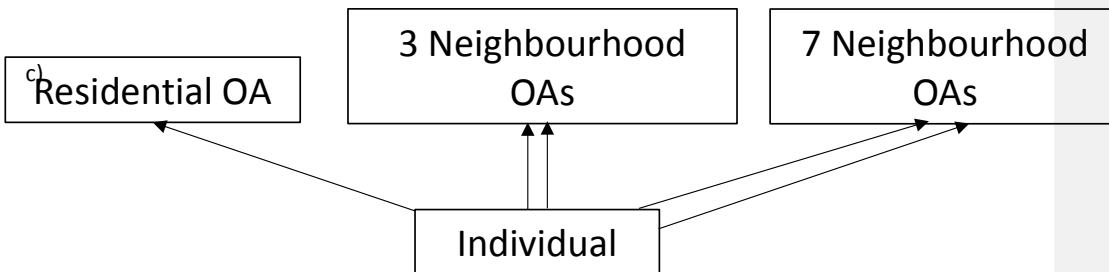


Figure 2 Results for different zonations when there are 3 classifications: individuals, OAs and Neighbourhoods

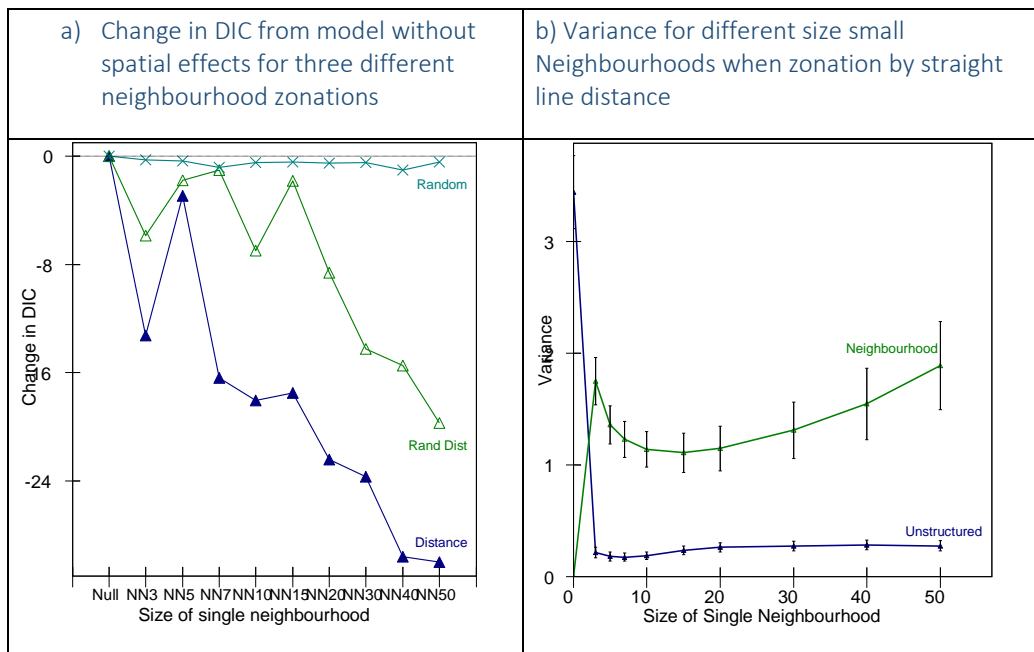
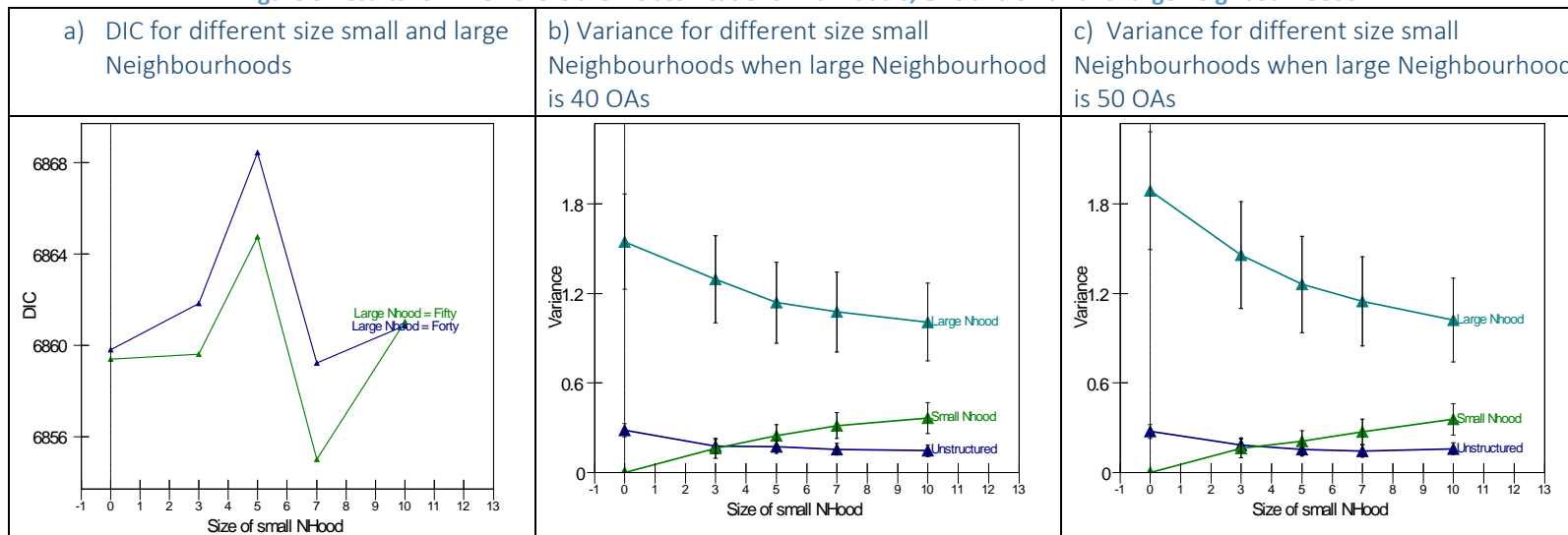


Figure 3 Results for when there are 4 classifications: individuals, OAs and small and large neighbourhoods



Tables

Table 1: The results of a set of hierarchical model fitted to actual and simulated data

Data	Actual					No Segregation						
Model	Null		2 Level		3 Level		Null		2 Level		3 Level	
	Estimate	Estimate	%	Estimate	%	Estimate	Estimate	%	Estimate	%	Estimate	%
Mean	-0.93	-1.68		-1.54		-0.93	-0.93		-0.93			
Between Zone variance				2.68	76.79				0.00	0.00		
Between OA variance		3.44	100	0.81	23.21		0.00	0	0.00	0.00		
Total Variance		3.44	100	3.49	100		0.00	0	0.00	0.00		
DIC	128568	6888		6859		6816	6850		6857			
Change in DIC: Null to 2 level		-121680					+34					
Change in DIC: 2 to 3 level				-29					+7			

Note: Estimate is the estimated parameter mean obtained by MCMC simulation; % represents % higher level variance not including Binomial variance.