The impacts of aggregation and spatial non-stationarity on migration models

Galen J. Maclaurin, Stefan Leyk and Lori M. Hunter

ABSTRACT: Demographic data are typically collected at various levels of spatial aggregation (i.e. census unit, county, or village), resulting in the modifiable areal unit problem (MAUP), with significant implications for analysis. This study uses demographic survey data from an impoverished area in rural South Africa to investigate aggregation (or scale) effects attributed to MAUP in global and local modeling approaches for migration outcomes. Aggregation effects in migration models are not well understood and interactions between aggregation and effects of spatial non-stationarity have not been studied to-date. This study makes use of spatially referenced household-level data to develop a spatial permutation approach to systematically aggregate analytical units. The random generation of aggregated units is repeated 500 times at each of nine levels of aggregation and global and local Generalized Linear Models (GLM) are run on each permutation to produce robust results. The local GLM model is based on a random region permutation algorithm which allows for robust analysis and diagnostics with statistical inference thus overcoming limitations of existing local estimators.

Results show that some variables, such as socio-economic status (SES), have very high operational scale sensitivity in that they represent highly significant predictors for migration at only (or close to) the operational scale of the process (household level). As units of analysis are aggregated these variables become non-significant and local models show little significant variation in local relationships. Other variables seem to be less sensitive to aggregation, such as 'Level of Household Education,' which remains a significant predictor across all levels of aggregation. This study shows that spatial aggregation reduces spatial variation in migration-related associations but affects variables differently.

KEYWORDS: MAUP, cross-scale spatial non-stationarity, spatial migration models, random spatial permutation, local coefficient maps

Introduction

Demographic data are collected at various levels of aggregation (i.e. census unit, county, village, etc.) and this aggregation's structure has significant implications for analysis and interpretation (Flowerdew et al., 2002). This phenomenon has been explained as the modifiable areal unit problem (MAUP) as it applies to the analysis of geographic information (Openshaw, 1983). The delineation of aggregated units (e.g., census tracts, zip codes) is usually an administrative and non-data driven process, and thus any analytical procedure, be it spatial or aspatial, will be influenced and confounded by the nature of aggregation. Wong (1995) divides MAUP into two sub-problems: scale effect and zoning effect. The former addresses the scale of aggregation (i.e. the number of observations being grouped into each aggregated unit) and implies that different results might be obtained by analyzing the same data at different levels of aggregation. As such, any interpretation would have to address the scale of analysis. The latter, zoning effect, refers to the form or shape of the partitioning of observations (i.e. the different ways in which observations can be aggregated into *x* number of units). This effect implies that the way the atomic units of the underlying process (i.e. individuals or households) are grouped could produce different results. The zoning effect is compounded by successive data aggregation, which in turn exacerbates the scale effect acting on results at different aggregation levels. Therefore it is very difficult to separate the impacts of these two different effects.

The effects of MAUP on regression models using demographic data have been studied extensively at nested levels of aggregated units (i.e. census blocks, groups and tracts) (Openshaw and Taylor, 1979) but usually only a few such levels are available. These studies have shown that coefficient estimates can fluctuate significantly and even exhibit sign changes at different scales. More recent research has addressed such aggregation effects within multiple regression models (Fotheringham and Wong, 1991), effects of random aggregation (Flowerdew et al., 2001), and effects on population - environment interactions (Walsh et al., 2004). To that limit, the effects of MAUP on spatially aggregated census data are well understood (Wong, 2009). However, most demographic processes studied, such as migration, have a very different—and much finer—operational scale (i.e., the individual or household) on which data should be ideally analyzed. Data are rarely available at such atomic units and if so (e.g., in U.S. census microdata) spatial identifiers are typically not readily available; this can be seen as one main reason why the impact of MAUP on migration models is widely unrecognized and understudied to date.

Another phenomenon that has recently gained increasing interest in demographic research involving geographic information is spatial non-stationarity. Local estimators have been proposed to examine the effects of spatial non-stationarity in statistical associations, showing that these effects can confound the results of an aspatial modeling approach that essentially ignores spatial relationships. A number of local estimators have been developed, with geographically weighted regression (GWR) (Fotheringham et al., 2002) being the most commonly used. However, recent research has shown that the GWR method lacks robustness for statistical inference (O’Sullivan and Unwin, 2010) and has further been argued to induce multicollinearity (Griffith, 2008, Wheeler, 2007) and patterns of spatial heterogeneity in coefficient surfaces (Cho et al., 2010). In response to some of these shortcomings an alternative modeling framework has been proposed that allows for more robust analysis and diagnostics when modeling local migration-environment associations (Leyk et al., 2012). However, since MAUP has not been investigated in migration research, the interaction between effects of aggregation and spatial non-stationarity is widely unknown although it can be expected to confound the analysis in complex ways.

This study can be situated within the fields of spatial demography as focused on migration modeling, and geospatial sciences concerned with fundamental concepts addressing MAUP and spatial non-stationarity. It makes use of spatially explicit demographic surveillance data at the household level from a remote rural region of South Africa to investigate the effects of aggregation on migration models across very fine spatial scales. The term *global* in this paper is used to refer to a regression model where all observations are used to return one set of coefficient estimates. In contrast, a *local* model estimates a different set of coefficients for each observation. Since the importance of local migration models has been demonstrated recently, this study will investigate how aggregation affects both global and local statistical models. This will allow evaluation of the interactions between effects of spatial aggregation (scale effects) and spatial non-stationarity to better understand the implications for migration-related associations under both modeling frameworks. The analysis is based on a complex spatial permutation approach in order to handle large volume spatial data and to ensure statistical robustness of the results and identified trends. The availability of household level migration data provides a unique opportunity to compare statistical associations at different levels of aggregation with the associations at the non-aggregated level (households) at which the process of interest operates.

Data

This study employs the 2007 household census conducted at the Agincourt Health and Demographic Surveillance Site (AHDSS) in a rural region of northeastern South Africa, operated by the MRC/Wits Rural Public Health and Health Transitions Research Unit. The dataset consists of 9,374 geo-referenced households representing 38,118 individuals from 21 villages. This dataset allows comparison of model properties (for global and local regressions) between the atomic household unit and increasingly aggregated scales of analysis, thus providing a unique opportunity to study the scale effect of MAUP.

The models we employ estimate the number of temporary migrants (tempmign) as the outcome variable on two independent variables: total years of education of all household members (HHeduc) and socio-economic status (SES). The following control variables were also included in the model: number of household members (HHpop), female head-of-household (Boolean variable—femhead), married head-of-household (Boolean variable—marhead), proportion of household members working (HHwork), proportion of males to females in household (mascprop), dependency ratio of household (number of individuals over 65 years old divided by number of individuals between 15-65 years old—deprop), and indicator of available natural resources for the household (greenness). A temporary migrant is defined as an individual, older than 15 years of age who spends more than six months in a year away from home while remaining linked to the household. The SES variable is a combined measure of household modernization assets, livestock assets, and information about power source used, dwelling structure, and sanitization. The greenness variable was derived from MODIS Normalized Difference Vegetation Index (NDVI) imagery by first taking a 2000 meter buffer around each household. Then the sum of NDVI pixel values within the buffer was divided by the number of households within the buffer, representing an approximation of the communal natural resources available to each household.

Methods

This research provides a modeling framework for investigating the effects of aggregation on local and global regression models of outmigration in a comparative way. The local model as described in more detail below has been adjusted and extended from a method developed by the authors (Leyk et al., 2012). To define benchmarks, the two models are computed at the household level, which is referred to as ‘level one’. To clarify, this is the level at which the process of interest—which is household temporary outmigration—operates (i.e., operational scale), as the choice to send a family member as a temporary migrant is typically a household decision. The household data are systematically aggregated to spatial units. These entities will then be used as analytical units in estimating outmigration at the different levels of aggregation using both the global and local models. This comparative approach allows examination of model behavior across increasing aggregation levels. The dataset used here provides a unique opportunity to investigate fine-scale aggregation effects since the survey is reported at the operational scale (including spatial identifiers) for outmigration and all other scenarios can thus be compared to this original level.

Creating modifiable units: The aggregation algorithm is based on a so-called binary tree partitioning scheme which recursively divides the household level data into subregions until each of these regions meets a size criterion defined by the number of households. Starting with the original household data (i.e., level one), the algorithm randomly chooses two seed points (household locations) and then groups all remaining households to one of the two seed points based on minimum (Euclidean) distance. This creates two spatially contiguous regions, which are each then subdivided again into two new contiguous regions using the same procedure. This procedure is repeated until each of the resulting subregions has a number of households that is within a set range of thresholds. The lower threshold is given as the minimum number of households for each region, and the upper threshold is twice this minimum, allowing the size of regions to vary in a way that is similar to an administrative unit (e.g., a rural community or a census unit). For example, the first level of aggregation—level two—consists of regions containing between 2 and 4 households. Nine levels of aggregation will be examined; level ten will thus have regions containing between 10 and 20 households. Aggregating beyond level ten would not provide enough sample units on which to run the local model. The households in the determined subregions represent the new units of analysis at each level of aggregation and are spatially referenced by the centroid of the region. Attributes of the households within one region are aggregated by calculating the mean for HHeduc, HHwork, deprop, mascprop, SES and greenness, and the sum for tempmign, HHpop, femhead, and marhead.

Statistical modeling as described below was carried out multiple (500) times for each of the aggregation levels to derive robust results. Thus for each of 500 simulations at the same aggregation level, a new random aggregation process as described above resulted in 500 different aggregation outcomes at each of the nine aggregation levels (two through ten). It is important to note that the regions are partitioned differently for each simulation due to the random choice of seed points, and therefore the centroids determined for the aggregated units at the same aggregation level do not correspond spatially and summarized attributes for the aggregated units vary across the 500 simulations.

Statistical Modeling: For each of the 500 simulations at each aggregation level, we run both global and local regression models. Both models use the same set of dependent and independent variables, which allows direct comparison of global vs. local regression results as well as the corresponding surfaces of residuals. Our dependent variable is counts of temporary migrants per household or per spatially aggregated unit of analysis which follows a Poisson distribution. Thus we employ Poisson-adjusted Generalized Linear Models (GLM). We investigate the spatial structure of regression residuals from both the global and local models using Moran's I to test for global spatial autocorrelation and local indicators of spatial association (LISA) to examine the existence of local clustering of residuals. Local clusters in model residuals often indicate systematic over- or under-prediction resulting in non-random error structures.

Global Statistical Models across different aggregation levels: The global Poisson GLM is run for each of the 500 simulations at each aggregation level (two through ten). Coefficient estimates, corresponding p-values, Moran's I for residuals, and the number of significant LISA clusters ($p<0.05$) from residuals are recorded. Thus, for each level of aggregation 500 sets of global model results are stored. The results are then summarized by aggregation level providing: average coefficients and their proportions of significance tests for each variable ($p<0.05$), Moran's I values and the average number of significant local clusters in model residuals derived from LISA analysis.

Local Statistical Models across different aggregation levels: The local modeling approach for temporary outmigration is based on a random region permutation approach in the same setting. The algorithm computes a Poisson GLM applied to randomly generated sub-regions in a Monte Carlo-like framework which allows for robust analysis of existing statistical relationships at local geographic scales. The advantage of this approach is the model's simplicity; there are no assumptions about the underlying spatial structure inherent in the model, making it more parsimonious than commonly used spatial models (i.e. GWR, spatial lag or spatial error models). A GLM framework also provides for extensive model diagnostics. Spatially contiguous subregions were randomly generated using the same binary tree partitioning algorithm described above but with a different intention: using these subregions as geographic extents across which local models are fit on the units of analysis (i.e., households at level one or aggregated subregions at all other levels). The regions generated for local modeling contain between 100 and 200 units of analysis. The minimum threshold of 100 units is based on a preliminary analysis to minimize prediction error and optimize AIC, while maintaining sufficient degrees of freedom for statistical inference.

To illustrate the procedure of local modeling, when analyzing the level one data the units (households) are randomly partitioned into spatially contiguous regions containing between 100 and 200 households. A Poisson GLM is run on the units within each region. The coefficient estimates, their corresponding p-values and model residuals are stored for all units in each region. This procedure is repeated 500 times, each time randomly generating a different permutation of regions for local modeling. Then for each unit (i.e., households at level one) the mean is calculated across all 500 simulations for the coefficient estimates and residuals. The proportion significant ($p<0.05$) of the p-values for each household across all simulations is calculated and recorded.

The local modeling procedure is then repeated for the nine levels of aggregation; however, the units of analysis are not households but the aggregated units as described above. Everything that applied to the case on the household level is now transferred to the corresponding aggregated spatial units of analysis, and repeated 500 times at each level. Thus subregions for local modeling are created by 100-200 units of analysis, which, for instance, are aggregates of 10-20 households for aggregation level ten. The number of subregions for local modeling is decreasing with increasing aggregation level. This modeling approach results in 2.25 million simulations in total for the local model.

Our overall objective is to contrast results from the global and local models at the same levels of aggregation. This detailed approach will allow improved understanding of MAUP as it manifests within examinations of household-level temporary migration.

Results

Global model results: As seen in Table 1, the coefficients for HHeduc are relatively constant and highly significant on average across all simulations for each aggregation level. This indicates that the level of household education is a strong predictor variable which is stable under various levels of aggregation, and the positive value is in line with substantial migration research linking education to higher migration probabilities (White and Lindstrom 2006). The coefficient values for SES decrease significantly with increasing aggregation and are not significant ($p<0.05$) beyond level four. The consistent decrease in average p-value with increasing aggregation indicates that this variable is a significant predictor only at the household level and the first three aggregation levels. Substantively, SES also exhibits a positive association with temporary outmigration, at least at lower levels of aggregation. This is consistent with research in some regions that demonstrate the necessity of household assets in fueling migration, given that migration entails costs (e.g. Gray and Mueller 2012). Still, just as the variable loses predictive ability at increasing aggregation, the positive substantive association is also not consistent across regional and cultural settings.

Local model results: To visualize the results of the local models across all 500 simulations at each aggregation level, the point vector features (centroids of aggregated units of analysis) were converted to a raster representation with 30m resolution. Such a data reduction and visualization strategy was necessary since the modeling process resulted in extremely large datasets. For example, running 500 simulations at level two alone resulted in 1.6 million points. The aggregated vector units (i.e., points) were converted to raster using a mean function for all points inside a given cell. This was done to create surfaces of regression coefficients and proportions significant for each variable. Coefficient estimates and proportions significant of HHeduc and SES are shown for four aggregation levels in Figures 1 through 4 to represent the trends across all nine aggregation levels.

Table 1. Global and local model residual analysis, and global coefficients and their p-values for HHeduc and SES. Moran's I values, numbers of significant High-High (H-H) and Low-Low (L-L) LISA clusters as well as global coefficient estimates and p-values are shown as means across all 500 simulations for levels two through ten.

Level	Moran's I Global	Moran's I Local	H-H Global	H-H Local	L-L Global	L-L Local	HHeduc Coef.	SES Coef.	HHeduc p-value	SES p-value
One	0.041	-0.008	270	173	85	99	0.038	0.175	1.66e-26	1.38e-12
Two	0.091	0.006	86.986	55.130	84.8	72.5	0.037	0.121	8.52e-09	2.95e-03
Three	0.121	0.018	65.342	40.628	62.9	54.0	0.038	0.101	6.50e-08	2.84e-02
Four	0.145	0.029	50.744	32.546	46.4	37.3	0.040	0.089	3.73e-07	7.22e-02
Five	0.165	0.039	40.404	27.450	39.2	31.3	0.041	0.081	6.04e-06	1.36e-01
Six	0.181	0.049	33.904	24.028	33.7	27.1	0.042	0.072	4.07e-06	1.96e-01
Seven	0.193	0.055	28.046	20.570	29.1	24.7	0.044	0.064	3.89e-06	2.58e-01
Eight	0.204	0.062	23.666	18.450	25.8	22.0	0.044	0.062	1.03e-05	2.99e-01
Nine	0.212	0.068	20.514	17.142	23.1	19.8	0.045	0.055	1.26e-05	3.53e-01
Ten	0.218	0.074	18.246	15.984	20.6	17.5	0.045	0.055	2.67e-05	3.64e-01

The surfaces of HHeduc coefficient values (Figure 1) and the corresponding proportions significant (Figure 2) across all 500 model runs at each aggregation level are highly correlated as can be seen in these spatial distributions. In the upper left map (level one), pockets of the highest coefficient values are overlapping with the highest proportions significant indicating that there is a considerable degree of spatial variation across the study area in model performance and spatial non-stationarity in the underlying process. The results reveal significant change in the coefficient and proportion significant surfaces with increasing aggregation. Changes in the sign of the coefficient values can be seen in a few places between levels one and two. From level two, these effects are smoothed out with increasing aggregation, resulting in a rather regional phenomenon at level ten (bottom right map). There is an increasingly noticeable north-west to south-east gradient in coefficient values and proportions significant with increasing aggregation. However, boxplots of all coefficient estimates from all model runs at each level show that the range of the values is fairly consistent with increasing aggregation (Figure 5), indicating that this variable is fairly stable under aggregation in the local model.

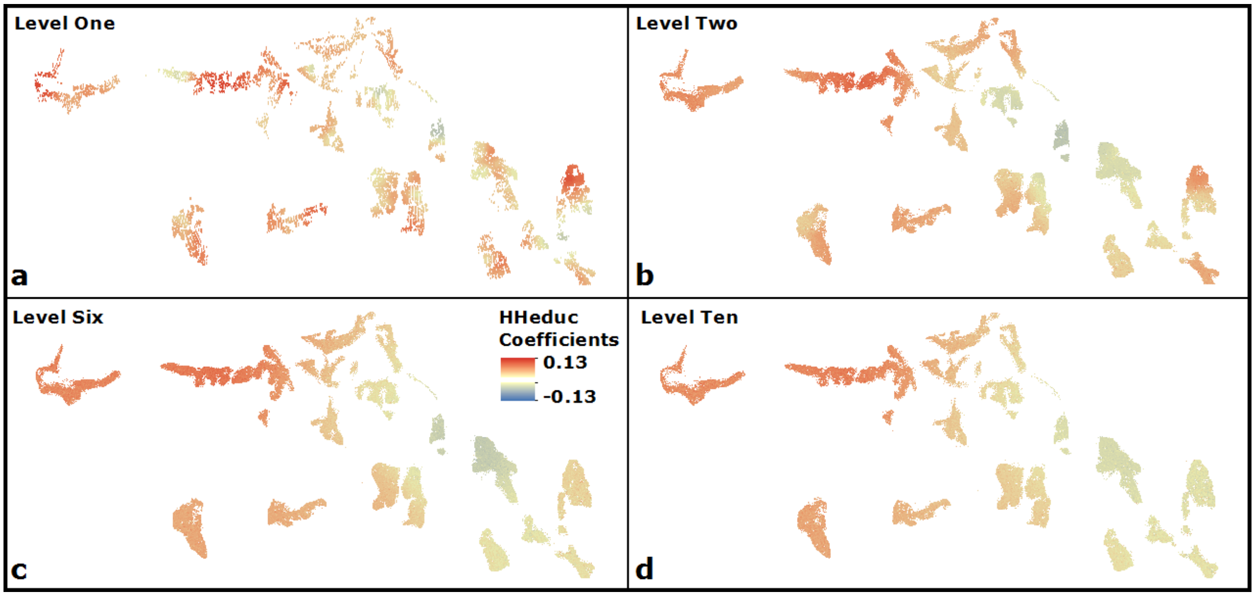

Figure 1. Spatial distribution of coefficient estimates for **household education** (HHeduc) from **local models** computed for household level (level one) and nine aggregation levels (levels two, six and ten shown).

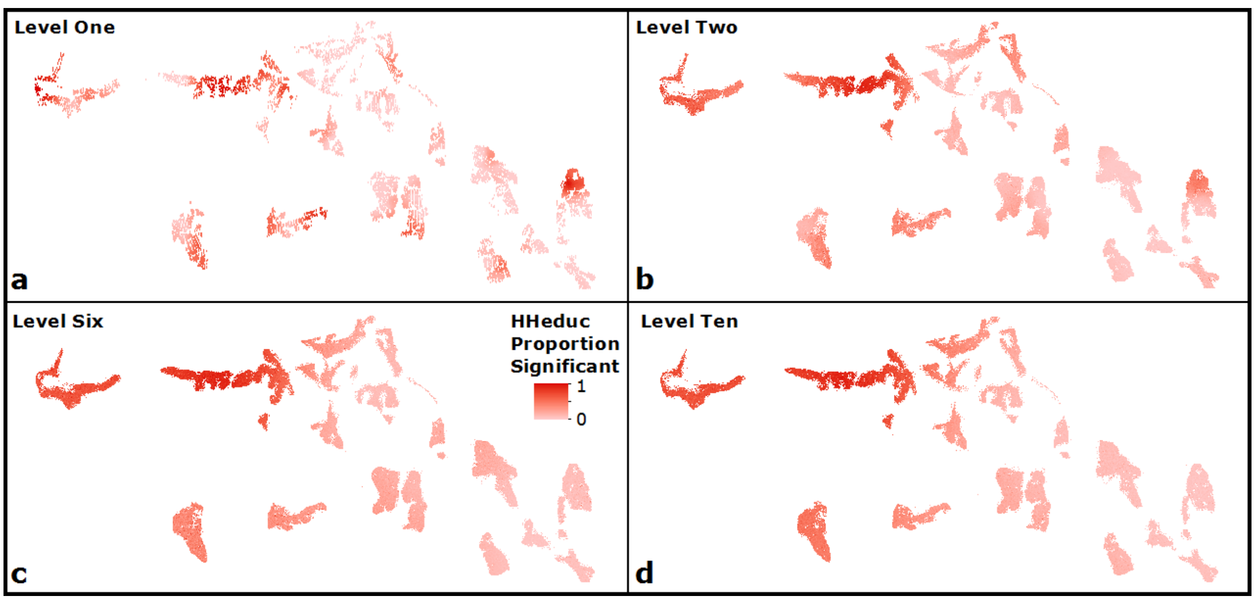

Figure 2. Spatial distributions of the proportions significant of the coefficient estimates from local models for the **HHeduc** variable across different levels of aggregation as shown in Figure 1.

For SES the patterns are related but differ to some extent. Very local pockets of positive high-valued coefficients (Figure 3) overlap with the highest proportions significant (Figure 4), as seen with HHeduc. However, the pattern of high spatial variation in the local model coefficients at level two breaks down much faster for both coefficient estimates and proportions significant with increasing aggregation. Both surfaces show a stronger smoothing effect than HHeduc indicating less stability of the surface at higher levels of aggregation. In contrast to the surfaces related to HHeduc, SES shows only a weak gradient at level ten across the study site which is accompanied by decreasing ranges in the boxplots for the values of both coefficients (Figure 6). This is also reflected in the spatial distribution of proportions significant (Figure 4). The decrease in the range of coefficient estimates and the low proportions significant indicate that SES is less stable and has lower predictive power across aggregation levels than HHeduc (Figure 5 and 6).

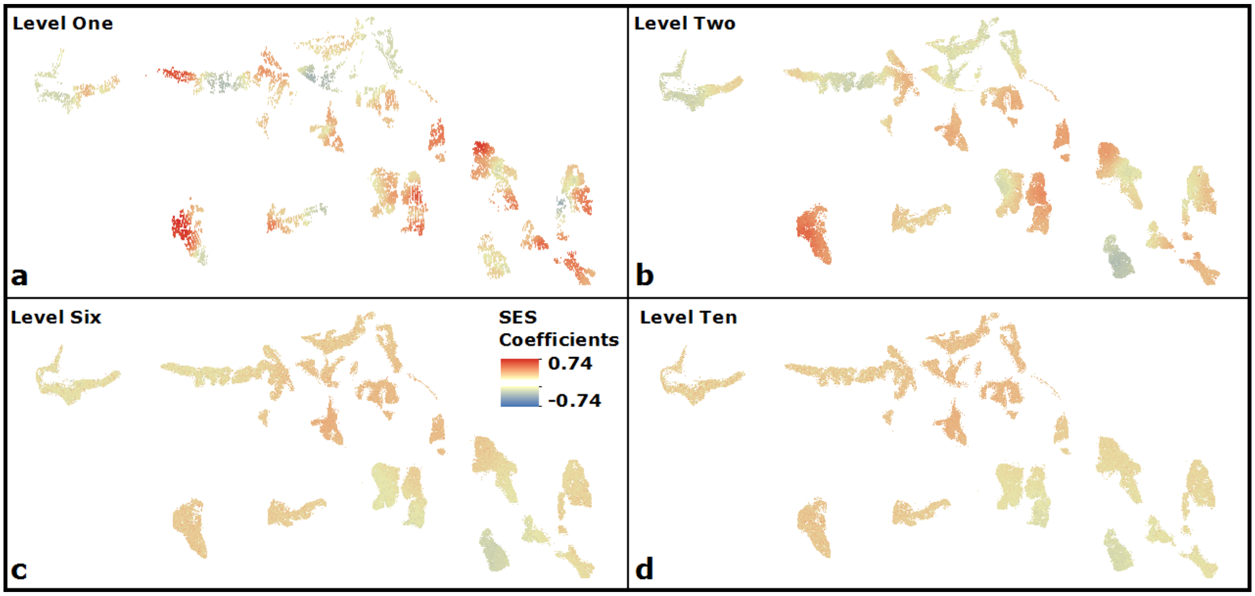

Figure 3. Spatial distribution of coefficient estimates **for socio-economic status** (SES) **from local models** computed for household level (level one) and nine aggregation levels (level two, six and ten shown).

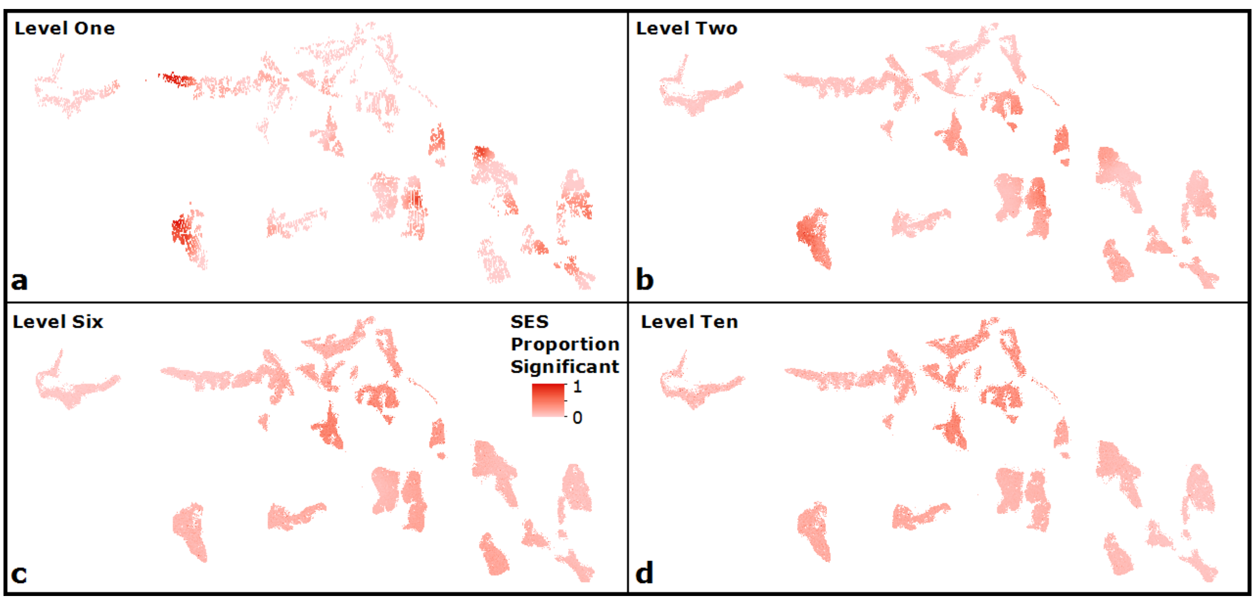

Figure 4. Spatial distributions of the proportions significant of the coefficient estimates from local models for the **SES** variable across different levels of aggregation as shown in Figure 3.

<u>*Error Structure:*</u> The spatial structure of regression residuals from local models shows a similar trend to the global models across aggregation levels but with systematically lower numbers of

local LISA clusters (Table 1), indicating that there is spatial non-stationarity in the underlying process (at the operational scale) and thus the need for a local model. This is further shown by the Moran's I statistic calculated on the residuals which is lower for local models than for global ones. An increasing trend is seen in both cases suggesting that spatial autocorrelation is more severe in the error surfaces at higher aggregation levels. The Moran's I statistic was significant ($p<0.05$) for all 500 model runs on all aggregation levels for both global and local models.

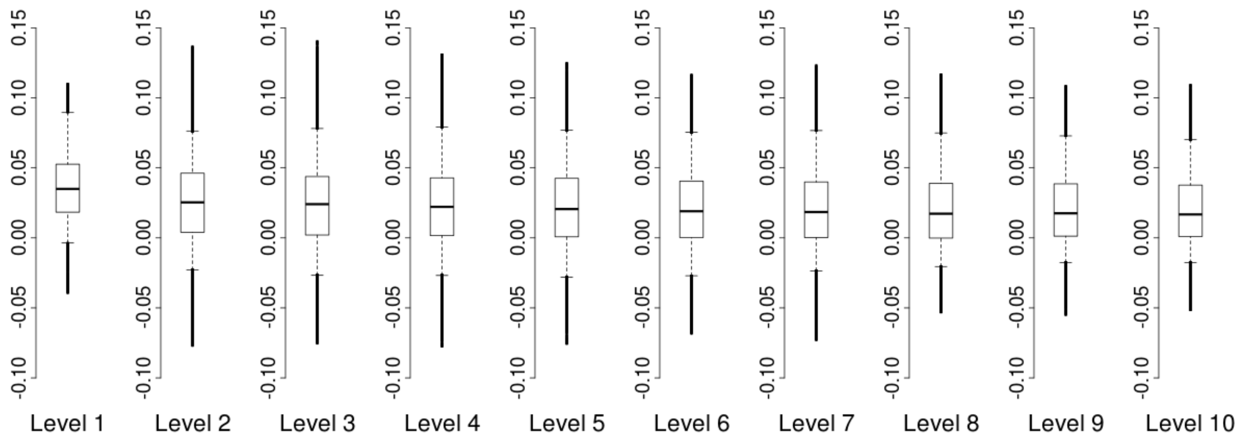

Figure 5. Boxplots of the local model coefficient estimates for the household education variable (HHeduc) at level one and across all 500 simulations for each aggregation level (two through ten).

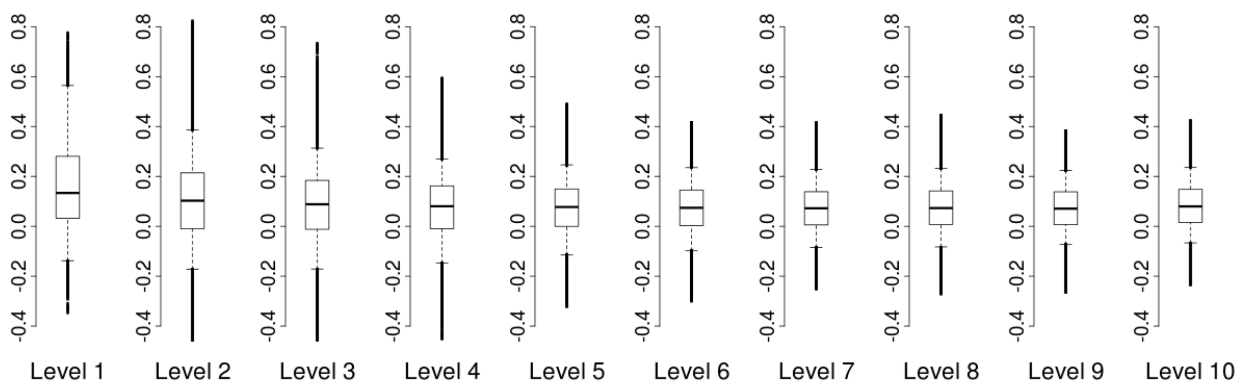

Figure 6. Boxplots of the local model coefficient estimates for the SES variable for level one and across all 500 simulations for each aggregation level (two through ten) for the local model.

Discussion and Concluding Remarks

While MAUP's aggregation effects have been examined in spatial analysis of demographic data extensively, researchers rarely have the opportunity to compare results from different aggregation levels to the scale at which the process of interest most logically takes place. In this study, household level demographic survey data from a rural area in South Africa including spatial identifiers are used to examine effects of aggregation on models of temporary outmigration in comparison to the operational scale of the process of interest, the household-level. Moreover this study examined aggregation effects within two different model frameworks, a global and a local approach. By including a local model it was possible to evaluate interactions between effects of aggregation and local migration-related associations while accounting for inherent spatial non-stationarity. Such interactions have not been investigated to-date since most studies on spatial non-stationarity use aggregated demographic data (e.g., census tracts or counties).

The results from the global models allow first insight into aggregation effects on statistical migration models if spatial non-stationarity is not considered. HHeduc remains a highly significant and stable predictor for temporary outmigration up to the highest level of aggregation. SES represents a significant predictor at scales close to the operational scale of the outcome, the household. In contrast, HHeduc appears to be a more general predictor, less sensitive to the underlying scale of analysis.

We argue that it is highly relevant to examine the effects of aggregation in a local modeling framework in comparison to this global approach. The importance of the evaluation of local model relationships has been demonstrated in a recent study (Leyk et al., 2012) which described the method used here for robust local coefficient estimation with statistical inference but did not address aggregation effects. The results from the local models allow for more in-depth interpretations: increasing levels of aggregation reduce the effects of spatial non-stationarity in the local model relationships evidenced by the observed decrease in local variation and the overall smoothing of coefficient surfaces. Looking closer at the two variables tested it becomes obvious they differ in what can be called “operational scale sensitivity”. This means that some variables represent statistically significant predictors at a wide range of scales (e.g., HHeduc) while others are significant only at scales close to the operational scale of the process of interest (e.g., SES). Thus a model’s performance (and estimated coefficients as well as their substantive interpretations) could depend on the selection of predictors used in relation to the given scale of analysis.

Acknowledgments

We acknowledge the research contributions of Raphael Nawrotzki of the University of Colorado Boulder and Wayne Twine, Mark Collinson and Barend Erasmus of the University of the Witwatersrand, South Africa. Supported by NIH R03 HD061428, “Environmental Variability, Migration, and Rural Livelihoods.” The work has also benefited from the NICHD-funded University of Colorado Population Center (grant R21 HD51146) for research, administrative, and computing support. This work was also indirectly supported by the Wellcome Trust (grant 085477/Z/08/Z) through its support of the Agincourt Health and Demographic Surveillance System. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CUPC, Wellcome Trust, NIH, or NICHD.

References

Cho, S.-H., Lambert, D. M., & Chen, Z. (2010). Geographically weighted regression bandwidth selection and spatial autocorrelation: an empirical example using Chinese agriculture data. Applied Economics Letters, 17(8), 767e772.

Flowerdew, R., Geddes, A., Green, M. (2001). “Behaviour of Regression Models under Random Aggregation.” (Tate, Nicholas J., and Peter M. Atkinson, eds), 89-104. In *Modelling scale in geographical information science*. Wiley, Chichester.

Fotheringham, A. S., and D. W. S. Wong. (1991). “The Modifiable Areal Unit Problem in Multivariate Statistical Analysis.” Environment and Planning A 23, no. 7: 1025–1044.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Wiley, New York, 269 pp.Gray, C.L. and V. Mueller. 2012. “Natural disasters and population mobility in Bangladesh.” Proceedings of the National Academies of Sciences. April 17, 2012, 19(16): 6000-6005.

Griffith, D. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A* 40(11), 2751-2769.

Leyk S., Maclaurin G.J., Hunter L.M., Nawrotzki R., Twine W., Collinson M. and Erasmus B. (2012). Spatially and Temporally Varying Associations between Temporary Outmigration and Natural Resource Availability in Resource-Dependent Rural Communities in South Africa: A Modeling Framework. *Applied Geography* 34:559-568

O'Sullivan, D., & Unwin, D. (2010). *Geographic information analysis*. Wiley, Hoboken.

Openshaw S, Taylor P J. (1979). "A million or so correlation coefficients: three experiments on the modifiable areal unit problem." (Wrigley, N. ed), 127-144. In *Statistical Applications in the Spatial Sciences.* Pion, London.

Openshaw, S.. (1983). *The modifiable areal unit problem*. Norwick [Norfolk]: Geo Books.

Walsh, S. J., K. A. Crews-Meyer, T. W. Crawford, and W. F. Welsh. (2004). “Population and Environment Interactions: Spatial Considerations in Landscape Characterization and Modeling.” (Eric Sheppard and Robert B. McMaster, eds), 41–65. In *Scale and Geographic Inquiry*. Blackwell Publishing Ltd.

Wheeler, D. C. (2007). Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment & Planning A* 39(10), 2464-2481.

White, Michael J., and David P. Lindstrom. (2006). "Internal Migration." in D. Poston and M. Micklin eds., Handbook of Population. New York: Springer Publishers.

Wong, D.. (1995). “Aggregation Effects in Geo-referenced Data.” (Arlinghaus, Sandra L., ed), 83-106. In *Practical handbook of spatial statistics*. Boca Raton: CRC Press.

Wong, D.. (2009). “The modifiable areal unit problem (MAUP).” (Fotheringham, A. S. and P. A. Rogerson, eds), 105–123. In *The SAGE handbook of spatial analysis*. SAGE Publications. Los Angeles.

Galen J. Maclaurin, PhD Student, Department of Geography, University of Colorado Boulder, Boulder, CO, USA. Email <galen.maclaurin@colorado.edu>

Stefan Leyk, Assistant Professor, Department of Geography, University of Colorado Boulder, Boulder, CO, USA. Email <stefan.leyk@colorado.edu>

Lori M. Hunter, Associate Professor, Department of Sociology/Institute of Behavioral Science, CU Population Center, University of Colorado, Boulder, CO, USA, School of Public Health, University of the Witwatersrand, South Africa. Email <lori.hunder@colorado.edu>