Using Spatial Entropy of Urban Vegetation to Measure Neighborhood Stability in Shrinking Cities

By

Zijun Yang

A thesis submitted

in partial fulfillment of the requirements

for the degree of

Mater of Science

(School for Environment and Sustainability)

at the University of Michigan

August 2018

Faculty advisors:

Professor Daniel G. Brown, Chair

Assistant Professor Meha Jain

Abstract

Previous studies on land-cover change have focused on urban growth and its consequences. However, urban shrinkage has also occurred as a consequence of global economic transformations. Urban shrinkage can have profound consequences and change the spatial patterns of urban vegetation. To detect and predict urban shrinkage is important for better urban planning and policy making. This study works on 1) determining the possible roles of spatial entropy, which represents the spatial configuration of urban vegetation, in combination with other socioeconomic variables, in predicting neighborhood stability and urban shrinkage, and 2) how the scale of defined neighborhoods may affect the relationship between spatial entropy and neighborhood stability. For the City of Detroit, MI, I adopted spectral mixture analysis of Landsat-8 imagery to yield moderate-resolution maps of urban vegetation proportion. I calculated spatial entropy for defined neighborhoods based on the vegetation information. Controlling for socioeconomic variables from parcel data and U.S. Census Data, I developed spatial models of the relationships between no-structure rate with neighorhoods, an indicator of urban shrinkage, and vegetation spatial entropy. Models were performed on two levels of neighborhoods: census block groups and census tracts. The results show that spatial entropy has the largest (negative) association with the no-structure rate compared with other predictors on both levels of neighborhoods. While high-resolution imagery or parcel-based data were not readily available, this study shows that moderate-resolution imagery can be an effective source for detecting and predicting urban shrinkage.

Keywords: shrinking cities, neighborhood stability, spatial entropy, remote sensing

Table of Contents

1 Introduction and Background

With the process of rapid urbanization, many studies on land-cover change have focused on urban growth and its consequences, such as impervious surfaces and urban sprawl(Bhatta et al. 2010b; Brueckner 2000; Weng 2001). While increases in global population and urban areas since the mid-20th century are expected to continue, urban shrinkage has also occurred as a consequence of global economic transformations, from cities in Europe and Japan to North America (Haase et al. 2014). Some of these cities are still undergoing population loss. Causes for this shrinking process can be complicated and include deindustrialization, economic crisis, suburbanization, and political factors.

Urban shrinkage can have profound consequences on various urban characteristics including business, employment, housing, and urban infrastructure. The decline in population can lead to further demographic change, declining population density, increasing residential and commercial vacancy, housing loss and demolition. It can also be aggravated by out-migration and population aging. Urban infrastructures may also be affected due to lower demand, declining investment and the high cost of maintenance (Bartholomae et al. 2017).

More specifically, urban shrinkage affects urban land use and change urban spatial patterns (Haase 2006). Depopulation might lead to residential vacancy, vacant industrial land, or housing loss. Moreover, urban shrinkage does not equally influence all neighborhoods; declines in population tend to be more impactful in relatively poorer areas (Guerrieri et al. 2012). This may imply a spatial differentiation pattern existing

within a shrinking city; that some more stable neighborhoods can manage to keep their residents while others that are shrinking tend to lose their resident population and housing, leading to further landscape change (Hoekveld 2014; Weaver and Bagchi-Sen 2013). In shrinking cities, some shrinking neighborhoods have high vacancy rates, where derelict lawns are overgrown, and empty parcels are completely reclaimed by shrubs and trees; the spatial patterns of land cover may be expected to be relatively more uneven. Thus, uneven vegetation cover may indicate the extent of shrinkage.

Might monitoring and predicting changes in urban neighborhood change in shrinking cities based on spatial patterns of urban vegetation in different neighborhoods help support urban planning and policy making? The green spaces in shrinking cities need to be well-managed. Open fields may serve as habitats and can potentially impact urban biodiversity. Also, urban vegetation has effects on urban ecology (Fritsche et al. 2007). While causes and impacts of urban shrinkage have been studied elsewhere (Haase et al. 2012; Schwarz et al. 2010), few studies focus on the patterns of urban vegetation and its monitoring as potential information about urban socioeconomic change.

Remote sensing (RS) and geographic information system (GIS) techniques have been used to quantitatively study the urban form, spatial configuration, and dynamics (Liu and Weng 2013). Using remote sensing imagery and image classification techniques, urban development monitoring has been conducted in the context of both growing and shrinking cities (Banzhaf et al. 2009). Based on time series remote sensing imagery and GIS TIGER road data, urban population growth rates can be modeled and estimated (Qiu et al. 2003). Simulation of land-use and urban form change with

modeling approaches has also been performed (Haase et al. 2012). Specifically, models of urban shrinkage found that residential vacancy was correlated with and could be modeled by some social, demographic, and spatial predictor variables (Kabisch et al. 2006). However, researchers also found that change of spatial patterns occurs with time lags (Reis et al. 2016), so it can be helpful to include real-time data sources like remote sensing imagery in studies on changes in shrinking cities.

A variety of spatial metrics based on RS and GIS techniques have been widely utilized for assessing and quantifying urban change (Ji et al. 2006; Reis et al. 2016; Siedentop and Fina 2010). These metrics can quantitatively describe and measure spatial patterns (Bhatta et al. 2010a). Several measurements like fragmentation, diversity, density, connectivity, and proximity have been developed and implemented in studies on urban forms (Knaap et al. 2007). Some landscape metrics are powerful tools for helping study urban landscape pattern and represent urban change characteristics. Diversity metrics, such as Shannon's entropy, can represent the distribution and composition of urban landscape quantitatively. Shannon's entropy was developed from the theory of information and was originally designed for measurement of information content. In geographical studies, entropy can be used to represent how evenly a geographical variable is distributed across the whole area (Kumar et al. 2007; Yeh and Li 2001). For example, in applications to land cover, the proportion of area in each of some number of classes can be distributed more evenly, such that each class covers similar, or dominated by a small subset of the classes. Larger values of entropy indicate a more even distribution (Jat et al. 2008; Li and Yeh 2004; Yeh and Li 2001). Thus, Shannon’s entropy should be a useful metric to identify a diverse spatial distribution of

geographical variables like vegetation abundance. Spatial entropy is scale dependent. For instance, zones that are too large may not be able to depict the differences in spatial configuration as the proportion of the land-cover type may become more balanced as a zone contains areas with different spatial patterns.

A number of studies have utilized spatial entropy to identify urban change. Using population data and land-use data, the trend of urban sprawl in Shanghai area was identified (Li et al. 2016). A study focusing on Pearl River Delta analyzed entropy values and compactness indices for ten cities and found uneven land development patterns in this area (Li and Yeh 2004). Renyi's entropy was also used for assessing the level of urban sprawl (Padmanaban et al. 2017). Using different forms of entropy to assess urban sprawl was studied as well (Yeh and Li 2001). However, most of these studies focused on the process of urban growth and conducted analysis at the level of the whole city instead of the level of neighborhoods; these studies showed an increasing trend in spatial entropy because of built-up areas sprawling across cites, while assessment of urban shrinkage and at the level of neighborhoods needs more studies.

As mentioned above, urban vegetation cover can contain valuable information implying shrinkage. In this study, I used Landsat-8 images, which are moderate-resolution images with 30m pixels. Images are freely available, and the dataset covers a long time period since 1972, which also enables potential time-series analysis. While GIS parcel-level data and high-resolution remote sensing data may be more accurate, they are usually not readily available especially for earlier years.

There are several methods for estimating and mapping the area of vegetation cover from moderate-resolution RS imagery. Most of these studies have utilized pixel-

based techniques, like per-pixel classification techniques, to investigate how urban land-use changes (Siedentop and Fina 2010). However, when using moderate spatial resolution imagery, it is hard to get accurate results with per-pixel classification because pixels tend to be mixed, such that two or more land-cover types can be contained in one pixel. Therefore, using sub-pixel estimation methods (e.g., spectral mixture analysis approaches), estimating a pixel's composition in terms of fractions of several sub-pixel categories (e.g., land-cover types) in the pixel can yield a more accurate representation of the land cover types and amounts physically present.

Spectral mixture analysis (SMA) is an effective method that has been widely utilized to estimate sub-pixel fractions (Deng and Wu 2013; Small and Milesi 2013). The spectral information of each pixel is a mixture of various materials' spectral signatures, which can be derived from other pixels in the scene. Mixing models can be linear and nonlinear. While the nonlinear mixing model's physical process can be complex, linear spectral mixture analysis (LSMA) has been widely used in spectral unmixing. In the LSMA model, every mixed pixel's spectral signature is assumed to be a linear combination of several pure spectral signatures of certain materials known as endmembers. Analysis of urban reflectance suggested that urban reflectance can be divided into three endmembers: high albedo, low albedo, and vegetation. This model has been effectively used for unmixing reflectance spectra and estimating vegetation fraction (Small 2002; Small and Milesi 2013). More recently, spatially adaptive SMA (SASMA) was developed and able to identify fraction with an accuracy of about 10% (mean absolute error of 8.50%, root mean square error of 15.25%) (Deng and Wu 2013).

In this study, I used spectral mixture analysis to identify vegetation cover. Shrinking neighborhoods were identified with the help of parcel data summarized as fraction of lots in a neighborhood that lack a structure (i.e., the no-structure rate), and spatial entropy was calculated based on vegetation fractions in these neighborhoods. Due to the increased out-migration and demolition rates, abandoned properties and empty parcels tend to be overgrown and covered by vegetation. Thus, shrinking neighborhoods tend to represent a more diverse spatial pattern and relatively uneven distribution of vegetation, which can lead to lower values for spatial entropy in shrinking neighborhoods compared to more stable neighborhoods. To assess how spatial entropy could be an effective indicator of urban spatial pattern change, we fitted spatial lag models of the no-structure rate in relation to spatial entropy of vegetation, while controlling for other predictors from demographic data and parcel data. We analyzed spatial entropy values for urban vegetation cover at the level of neighborhoods to see how this metric can reveal spatial differentiation pattern in different neighborhoods in Detroit City.

I investigated (1) how spatial entropy, a metric derived from moderate-resolution images representing the spatial configuration of land cover, relates to indicators of shrinkage in neighborhoods and helps differentiate neighborhood condition within a shrinking city. As previous studies (Batty et al. 2014; Bhatta et al. 2010a) have pointed out, spatial entropy is a scale-dependent metric so it is also important to investigate (2) what scales (i.e. size of neighborhoods, or spatial resolution of remote sensing imagery) used for entropy calculation yield results that best differentiate the urban spatial patterns of shrinking neighborhoods from those of stable neighborhoods.

2 Methods

2.1 Study Area

My study area was the City of Detroit, Michigan, which has experienced continuous demographic and economic decline during recent decades. The population of Detroit peaked in the 1950s as a result of the expansion of the auto industry and the industrialization of this city (Neill 2015). After World War II, Detroit gradually lost its advantages when faced with global market competition. Manufacturing job losses and suburbanization led to outmigration from the city. The subprime mortgage crisis in 2007 also aggravated the trend of population loss and economic decline. Deindustrialization and decentralizing trends of the auto industry have made Detroit probably the most famous shrinking city in the U.S (Xie et al. 2018). The population of Detroit City has dropped from 1.85 million in 1950 to 670,000 in 2015, which means the city's population has decreased by over 60% since 1950 (U.S. Census Bureau 2016).

Along with the demographic and economic decline, spatial patterns in Detroit City changed accordingly. Compared to urban growth, the spatial patterns of urban shrinkage may be less clear. In shrinking cities like Detroit, commonly observed spatial patterns include vacant land, large-scale demolition, and increasing open spaces. It is notable that even though the city is losing its population, different neighborhoods in the city present a different population trend as shown in Fig. 1. Some neighborhoods are losing their residents much quicker than others. In more stable neighborhoods where less population loss is observed, buildings/structures and vegetation are more evenly

distributed. In neighborhoods with more population loss, more houses and infrastructures are abandoned and demolished, and vegetation may grow in such places and then cover these areas. Thus, we may expect that houses and vegetation are distributed more unevenly in such shrinking neighborhoods.

Figure 1. (a) A shrinking neighborhood with overgrown vegetation and diversely distributed properties

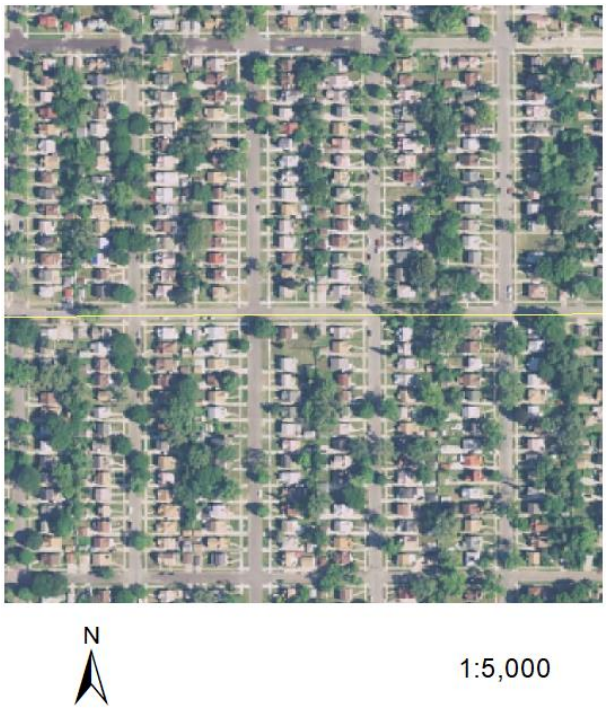

Figure 1. (b) A more stable neighborhood with regular distribution of houses and vegetation

2.2 Subpixel analysis (Spectral Mixture Analysis)

Vegetation information in this study was extracted from Landsat-8 imagery acquired on July 13, 2013 (Row 020 Col 031). The cloud-free image was acquired in the summertime for estimating vegetation abundance. The terrain-corrected surface reflectance product was download from the U.S. Geological Survey. Radiometric and atmospheric correction was performed on the image.

A limited spatial resolution will cause spectral mixture in remote sensing images, meaning that each pixel can contain various objects with different spectral information. The Landsat image we used in this study has a spatial resolution of 30m, which is larger than many structures in the city such as single houses. Thus, it is challenging to label each pixel as a single category of land use using per-pixel classification approaches. I employed Small's (Small 2002) model for spectral mixture analysis (SMA), which divides urban land cover into three categories: vegetation, high-albedo, and low-albedo. Spatially adaptive spectral mixture analysis (SASMA) algorithm was used because it can incorporate both spectral and spatial information to find endmembers and yield more accurate estimation.

For LSMA, the observed spectrum in an image is assumed to be a linear combination of spectra of several endmembers, which represent different land-cover types. Typically, the endmembers should have spectra of pure materials. However, finding the pure spectra of different land-cover types can sometimes be challenging. Instead, SASMA was designed to identify "most representative" endmembers with a spatially adaptive approach. A classification tree incorporating both spatial and spectral information was used for automatically extracting candidates of endmembers. For each mixed pixel in the scene, the algorithm will synthesize spectral signatures of all endmember candidates by inverse-distance-weighting (IDW) method within a local search window, to yield the final endmember spectra, which are considered as the most "representative" endmembers. Then, an LSMA method with the spectra was utilized for estimating the vegetation abundance within a pixel. To demonstrate SASMA's advantage, we also conduct a plain LSMA on the Detroit scene to see which method has

better results. To validate the results of SASMA and plain LSMA, we randomly sampled 180 validation sites, 90*90m in size, over the study area. Root mean square error (RMSE), mean absolute error (MAE), and systematic error (SE) were used for accuracy assessment. The reference dataset for validation is the National Agriculture Imagery Program (NAIP) Data. NAIP acquires high-resolution aerial imagery during the agricultural growing seasons. The spatial resolution for images of Michigan is ~0.6 meters. These aerial photographs were imported in ArcMap 10.3, and vegetation abundance for each site was calculated.

2.3 Spatial Entropy

Shannon's entropy has been used for quantifying the degree of spatial dispersion or concentration of a geographical variable x_i across n zones (pixels in this study) (Yeh and Li 2001). The relative entropy can be calculated as:

$$E = \sum_{i=1}^{n} p_i log\left(\frac{1}{p_i}\right) / log\,(n),\ p_i = x_i / \sum_{i=1}^{n} x_i,$$

where x_i is the value of the geographical variable of interest; in this analysis, the variable is the vegetation abundance. p_i is the share of vegetation in the i-th zone over the vegetation in all n pixels, The values of entropy range from 0 to 1. Smaller values of entropy indicate that the distribution of vegetation is more uneven while larger values indicate a more even distribution (Jat et al. 2008; Yeh and Li 2001). As shown in Fig. 1, shrinking neighborhoods have a relatively uneven distribution of vegetation. Some pixels are less vegetated with more building structures, while other pixels can be more vegetated. Thus, the hypothesis is that shrinking neighborhoods will have lower entropy

values in general; stable neighborhoods are vegetated more evenly, and proportions of vegetation in each pixel vary less, which results in higher entropy values.

The value of relative entropy varies with the number of pixels (n) within a neighborhood. Also, this value can still be affected by the size of pixels (Batty et al. 2014; Bhatta et al. 2010b). Thus, we experimented with two levels of neighborhoods: census tracts and census block groups. We also calculated spatial entropy at four levels of image spatial resolution: 30m, 60m, 120m, and 250m. Different levels of neighborhood and spatial resolution were tested to see on which level spatial entropy shows a stronger correlation with neighborhood shrinkage and instability.

2.4 Census and parcel variables

2.4.1 American community survey

The 2011-2015 American Community Survey (ACS) 5-year estimates data products at levels of census tract and census block group were acquired from the US Census Bureau. In the City of Detroit, there are 310 census tracts and 879 census block groups. The U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) has a product "TIGER/Line with Selected Demographic and Economic Data" which integrates geographic line dataset and ACS 5-year estimates. The dataset containing variables averaged from 2011 to 2015 was chosen to match the acquisition date (2013) of the Landsat image. Demographic and economic variables were selected as indicators of socioeconomic conditions of the neighborhoods. Selected variables include population in a neighborhood, population density of the neighborhood,

the proportion of the population with regular high school diploma, the population for whom poverty status is determined, the proportion of the population below the poverty level, and median income in the neighborhood.

2.4.2 Parcel data

The Motor City Mapping is a project dedicated to surveying every parcel within the city of Detroit. The survey was conducted from December 2013 to February 2014. The dataset has information of more than 370,000 parcels including their conditions, presence of structures, occupancy, and use. The City of Detroit's open data portal also has a parcel map dataset which includes the last sale prices of the parcels. The parcel-level datasets acquired from Data Driven Detroit data portal and the City of Detroit's open data portal were overlayed with the census tracts and census block groups' boundaries using ArcGIS. Variables at the two levels of neighborhoods were obtained from the overlayed data. The variables include the proportion of parcels encoded "No Structure" (no-structure rate), number of parcels in a neighborhood, average area of parcels within a neighborhood, area of a neighborhood, the proportion of commercial/industrial/institutional (CII) parcels, and the average price of residential parcels.

2.5 Spatial models

It is assumed that spatial autocorrelation is present in the study area. Thus, other than an ordinary least squares (OLS) model, modeling approaches accounting spatial autocorrelation are more suitable for modeling areal data. Appropriately estimating effects and their significance can be modeled in the presence of spatial autocorrelation using a spatial lag model or a spatial error model. The spatial lag model assumes that the dependent variable was affected not only by the values of predictors with the a given spatial unit, but also values of the dependent variable in neighboring locations. Impacts coming from neighboring regions are weighted by a spatial weights matrix. The spatial lag model takes the form as follow:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{e},$$

where **y** is the dependent variable, and **W** is the spatial weights matrix where the diagonal elements are zero, so one neighborhood's dependent variable will not appear on the right side of the formula (Bivand et al. 2008; Viton 2010).

The spatial error model assumes that the errors of the model are spatial autocorrelated, so it takes the form as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \mathbf{e},$$

W is defined as the in the spatial lag model (Bivand et al. 2008; Viton 2010).

Based on the variograms and results of Lagrange multiplier tests (Anselin 1988; Anselin et al. 1996) on the variables mentioned above, the spatial lag model should be more suitable for the dataset we have. The spatial models were fitted at the two levels of

neighborhoods (census tract and census block group) and the four levels of spatial resolution.

Spatial entropy and other variables from parcel data were log-transformed. As the spatial models require that the variables should be similar in scales and large scales will cause spatial models' inversion of asymptotic covariance matrix failed, variables were transformed or rescaled, so the spatial models were able to fit. Akaike's information criterion (AIC) was used for measuring the goodness of models and variable selection. Variable selection is a criterion-based procedure, and the model with the lowest AIC value was chosen as the final model.

In spatial lag models, as the response variable is spatially lagged, so the response variable in neighborhood i is not only affected by the predictors in neighborhood i, but also by i's neighboring neighborhoods' response variables. Therefore, while the response variable in neighborhood i is directly affected by the predictors in neighborhood i, it is also indirectly affected by predictors of i's neighboring neighborhoods. Interpretation on the fitted coefficients β may be insufficient, as β does not account for the spatial spillover, while a change in any predictor variable of a single observation will influence not only the neighborhood itself (direct impact) but other neighborhoods (indirect impact). Impact measures were formulated to assess both direct impacts and indirect impacts. For a predictor r of two different neighborhoods i and j ($i \neq j$), in the OLS model, $\partial y_i/\partial x_{ir} = \beta_r$, $\partial y_i/\partial x_{jr} = 0$, while in the spatial lag model, $\partial y_i/\partial x_{jr} = ((\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{I}\beta_r)_{ij}$. Let $S_r(W) = ((\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{I}\beta_r)$, the average of diagonal elements in the $n \times n$ matrix $S_r(W)$ is the average direct impact, and the average total impact is calculated as the sum of all elements in the matrix divided by n.

The average indirect impact is the difference between the total impact and direct impact. The impact measure accounting both direct and indirect impacts of predictor variables is more suitable and used in this study to assess each predictor's effect on the response variable (Bivand et al. 2008; LeSage and Pace 2009; LeSage and Fischer 2008). All remote sensing image processing was done by ENVI 5.3. SASMA was implemented with Matlab R2015b. The spatial lag models were fitted in R with the 'spdep' package.

3 Results

3.1 Vegetation abundance

The result of vegetation abundance yielded by SASMA is shown as follow:

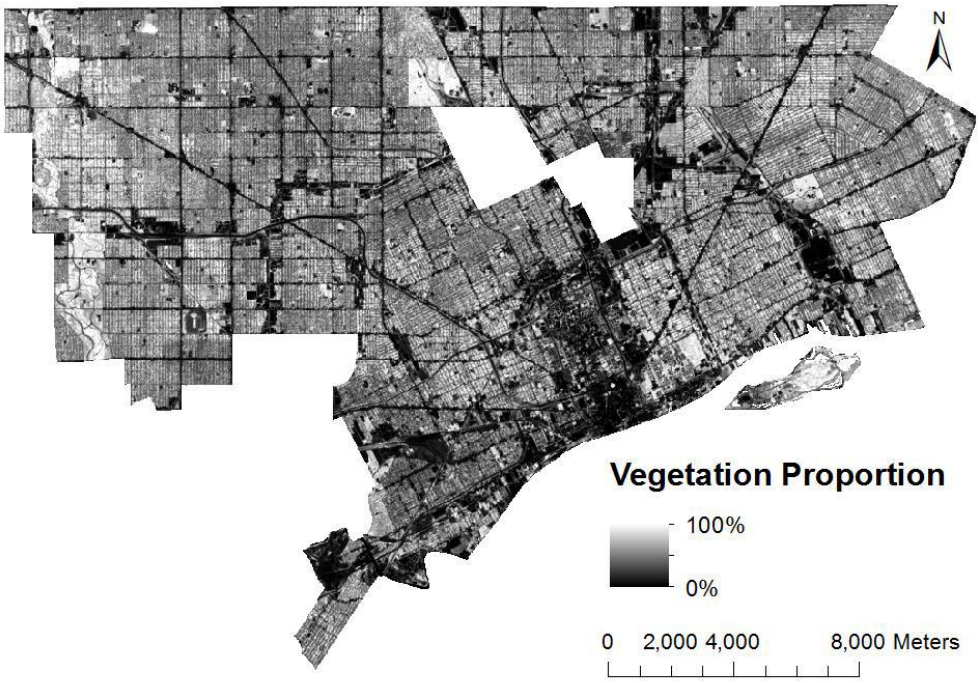

Figure 2. Vegetation abundance yielded by SASMA

As table 1 shows, SASMA achieves an RMSE of 7.9%, while non-adaptive LSMA has an RMSE of 14.01%. Also, SASMA has lower values in MAE and SE, which shows it is a more robust method compared to a plain LSMA in predicting subpixel vegetation abundance.

Table 1. Comparison of accuracy of vegetation abundance between SASMA and non-adaptive LSMA.

Method	RMSE	MAE	SE
SASMA	7.90%	6.04%	0.01%
non-adaptive LSMA	14.01%	11.05%	-0.09%

3.2 Spatial model results

3.2.1 Results at different levels of neighborhoods

Based on AIC, the selected variables at the level of census block group include spatial entropy, number of parcels, CII proportion, the average price of residential parcels, population, the proportion of the population with high school diploma, and median income in the neighborhood. Variables at the level of census tract are the same as variables at the block group level excluding the average price of residential parcels. Table 2 shows the models' fit of both OLS models and spatial lag models using the selected variables above at the two levels of neighborhoods. It is noted that even the OLS models have good fit (with R-squared of 0.54 and 0.50 at tract and block-group level respectively). Variance inflation factors are calculated and the result shows no serious collinearity among the predictors in these models. However, using spatial lag models improves the models' fit at both tract level and block-group level, which indicates the presence of spatial dependence in the response variable.

Table 2. Goodness of fit of OLS models and spatial lag models at tract and block-group levels. Null AIC is for models without predictor variables. R^2 for OLS models and pseudo-R^2 for spatial lag models.

	Tract		Block Group	
Model	OLS	Spatial Lag	OLS	Spatial Lag
null AIC	761.97	504.79	2402.69	1489.79
AIC	318.32	543.47	1818.13	1087.42
R^2/pseudo-R^2	0.54	0.82	0.5	0.81

Fig. 3 shows the impact measures of predictors at the two levels of neighborhoods. At block-group level, all predictors have significant impacts except the average price of residential parcels. At tract level, all predictors are significant.

At both levels, spatial entropy, median income, and population are the predictors that negatively impact on the response variable (no-structure rate), while high school proportion, CII proportion, and the number of parcels have positive impacts on no-structure rate.

In terms of the magnitude of the impacts, spatial entropy has the largest impact of -8.50 on the response variable. Median income (-3.54) and population (-2.74) are the second and third most influential predictors at this level. At the tract level, spatial entropy (-6.37) remains the most impactful predictor. Median income (-5.24) and population (-2.69) are the second and third most influential variable.

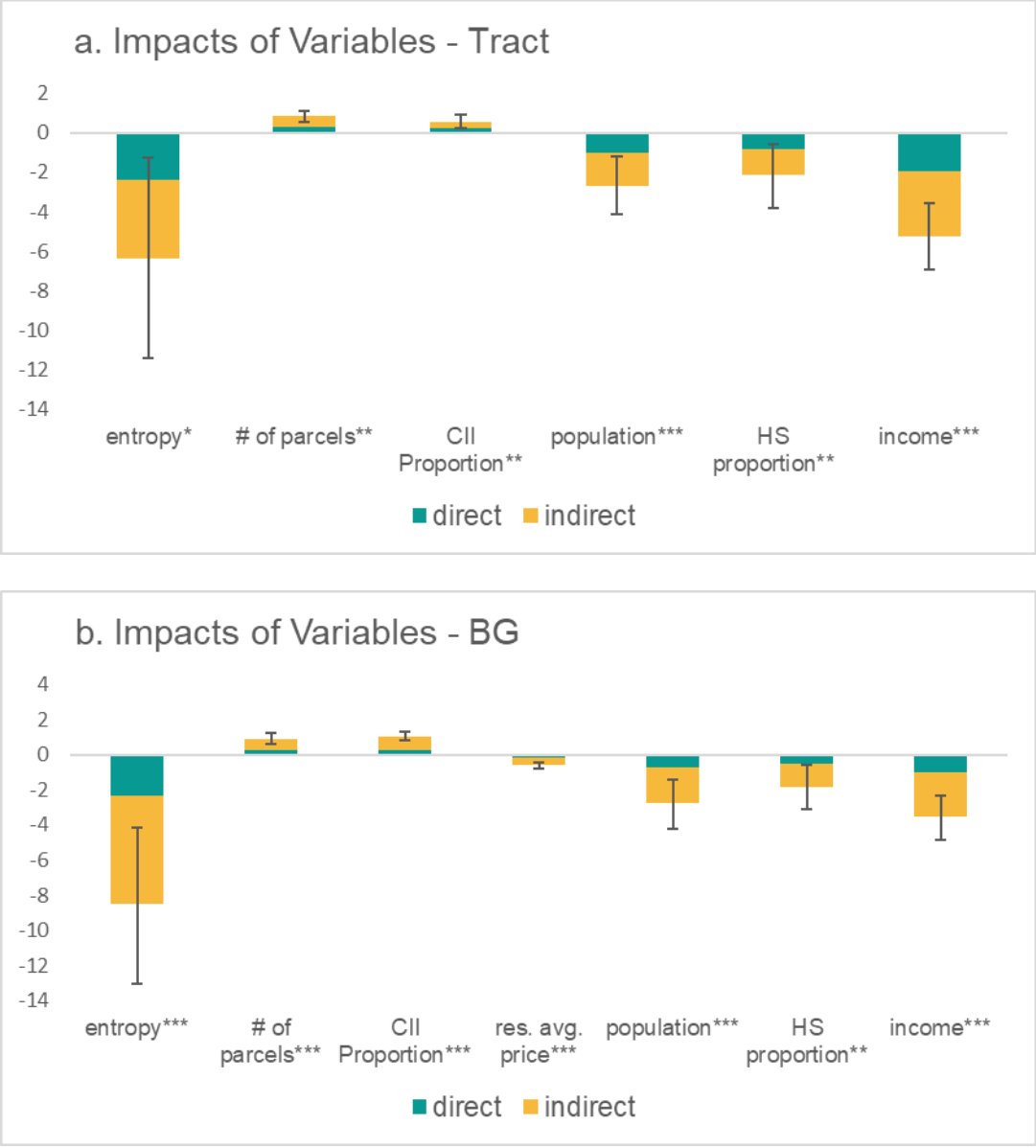

Figure 3. Impact measures of predictors at (a) tract and (b) block-group levels. Green bars represent direct impacts while orange indirect (described in 2.5 spatial models). Error bars stand for 95% confidence intervals. *p<.05, **p<.01, ***p<.001.

The residual spatial autocorrelation present in the OLS model was reduced as a problem in model estimation through use of the spatial model. The LM test (Breusch–Godfrey test) shows non-significant spatial autocorrelation in residuals for the spatial lag models at the tract (p=0.725) and block-group (0.380) levels, while significant

spatial autocorrelation is observed at both tract (p<0.001) and block-group (p<0.001) levels. The residual values of the block-group-level OLS model show strong clustering (Fig 4b), while, such clustering is not observed jn the residuals from the spatial lag model (Fig 4a). Lower values of residuals in the OLS model are clustered in the northwestern part and the northeastern corner of Detroit City. Thus, the spatial lag model is able to deal with the problem of spatial dependence in the dataset.

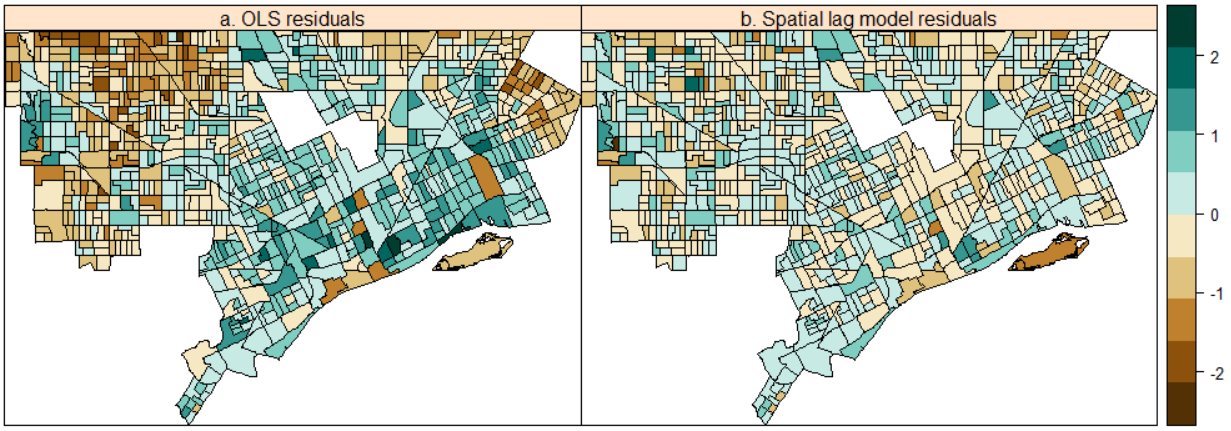

Figure 4. Residuals distribution of (a) OLS model and (b) spatial lag model at the block-group level.

Comparing models estimated for two scales (census block group and tract) and those with and without the social variables, we see a strong effect of vegetation spatial entropy on the no-structure rate (Figure 5). All models have reasonably high pseudo R-squared values, with the models at tract level having a slightly higher R-squared value than those at block-group level.

3.2.2 Results at multiple levels of image spatial resolution

AIC and impact values were used to compare models with spatial entropy values calculated based on different levels of spatial resolution (Table 3). At tract level, AIC

does not change much from spatial resolution of 30m to 250m, with the lowest AIC being 317.6 at the level of 120m. At block-group level, AIC values of the models at 30m, 60m, and 120m are similar, with the lowest AIC being 1084.7 at the level of 60m. However, when spatial resolution goes to 250m, the AIC value increases to 1097.2, which is much higher than the AIC values at other levels. Also, the impact measure of entropy based on the image of 250m resolution is no longer significant, which indicates the spatial resolution of 250m may not be appropriate for calculating spatial entropy at block-group level of neighborhoods.

Table 3. AIC of spatial models using entropy calculated from remotely sensed images with multiple levels of spatial resolution at tract and block-group levels. *p<.05, **p<.01, ***p<.001, n.s. = not significant.

Block Group			Tract	
Resolution	Impact of Entropy	AIC	Impact of Entropy	AIC
30m	-8.50***	1087.4	-6.37*	318.32
60m	-8.01***	1084.7	-5.37*	318.71
120m	-4.92***	1087.6	-4.92*	317.6
250m	-1.10 n.s.	1097.2	-2.57*	319.52

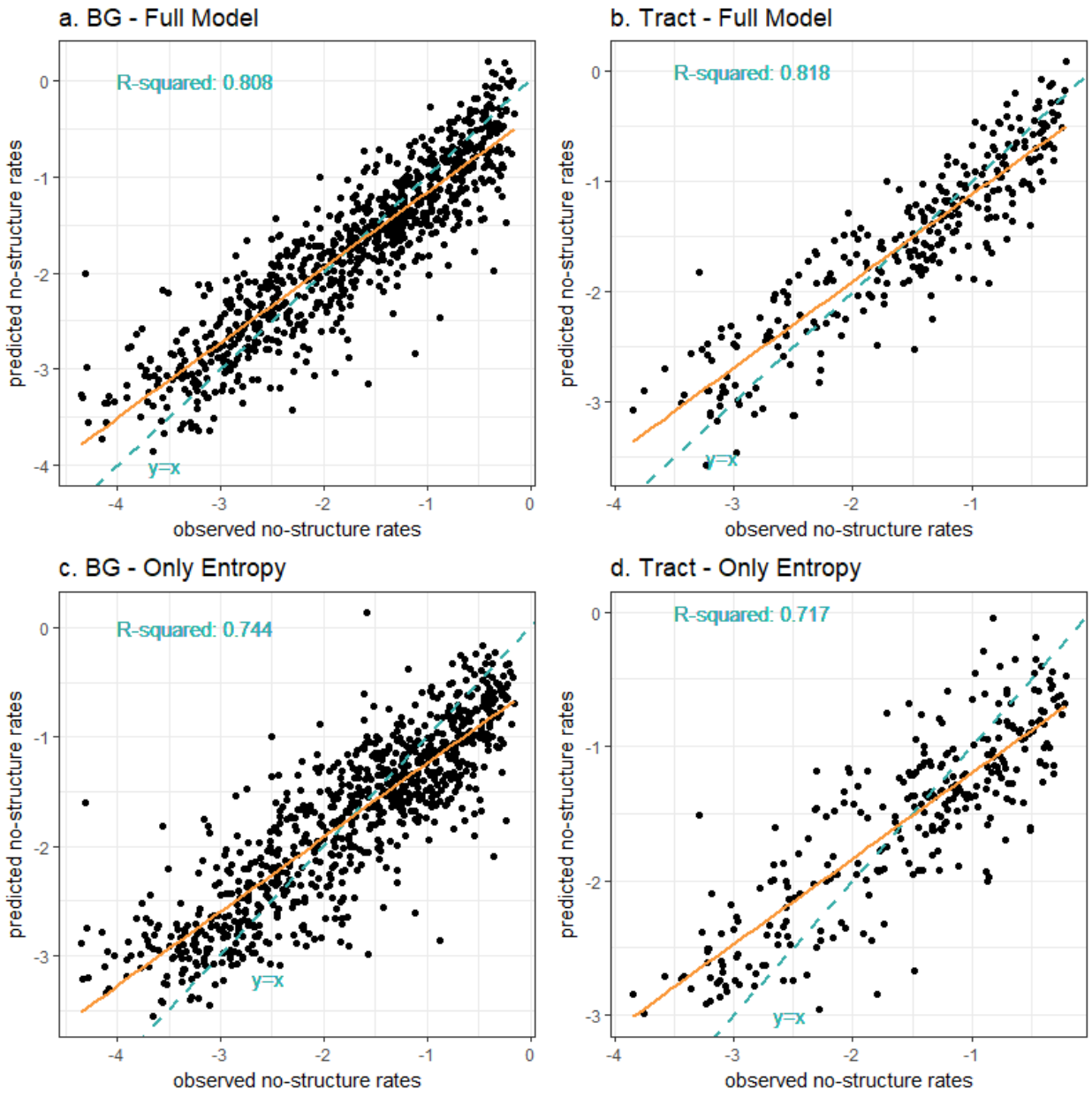

Figure 5. Comparison of observed values fitted values of no-structure rates at (a) & (c) block-group level and (b) & (d) tract level. (a) and (b) show the full models at block-group and tract levels with all predictors included, while (c) and (d) show the models with spatial entropy as the only predictor. The orange solid line is the regression line while the green dashed line is the reference line of 1:1. Each dot represents an individual neighborhood.

4 Discussion

4.1 Variable importance

To assess spatial entropy as a predictor of neighborhood structural conditions associated with shrinkage, we compared its impact with other social measures included in the models. Both tract and block-group level models had the same following predictors in addition to spatial entropy: number of parcels, CII proportion, population, the proportion of the population with high school diploma, and median income in the neighborhood, while the average price of residential parcels is only included in the block-group level model.

Spatial entropy had a significant effect on the no-structure rate at both tract and block-group levels even with these social variables included. In fact, its impact was largest among all of the variables. As expected, spatial entropy has a negative impact on no-structure rate, meaning that higher spatial entropy values are associated with lower no-structure rates. This relationship agrees with the hypothesis that spatial entropy is higher in neighborhoods with population and housing loss where no-structure rate tend to be higher and vegetation should be distributed unevenly. The second most impactful predictor at both levels was median income, which was also negatively associated with no-structure rate. Residents with higher income have more options to choose where they live, and they tend to live in more stable neighborhoods where structures are well maintained instead of shrinking neighborhoods where they may be neighbors to a number of empty lots. The third most influential predictor at both levels was population, and it is not surprising to find that population is negatively associated with no-structure rate. One of the most important characteristics of urban shrinkage is population loss. A

neighborhood with more population is more likely to be a stable neighborhood with fewer abandoned houses and vacant lots resulting in a lower no-structure rate. The proportion of the population with regular high school diploma was the fourth most impactful predictor in both levels' models. The more educated a neighborhood's residents are, the more likely is the neighborhood to have a low no-structure rate and, therefore, be more stable. The other two predictors, the number of parcels and CII proportion, are both statistically significant with relatively smaller impacts. These two predictors were also the only two that were positively associated with the response variable.

At the block-group level, the average price of residential parcels had the lowest impact value and was negatively associated with the response variable. It was surprising to find that the parcel price is not as influential as other predictors. Moreover, at tract level, parcel price was dropped based on AIC, and it was not statistically significant even when including it as a predictor at the tract level. It seems that parcel price should be a very useful variable in predicting a neighborhood's housing condition. More stable neighborhoods with better infrastructure and housing units should have higher parcel price. Its relatively low impact might be caused by the dataset. The average price of residential parcels is retrieved from the City of Detroit's open data portal, and the dataset records each parcel's price of its last sale. Thus, these price data can come from different years and in fact may not be comparable.

Overall, we find that the most impactful variable is spatial entropy based on remotely sensed images, and the second to fourth most influential variables all come from demographic data, while predictors from parcel data have the smallest impacts.

Also, Fig. 5(c) and (d) show that even the models can still perform well when including spatial entropy as the only one predictor. It is important to note that landscape changes may happen after (or lag in time) the social, economic, and demographic changes that are driving neighborhood change. However, possibility of enhancing studies of neighborhood change with the increased spatial and temporal resolution provided by remotely sensed imagery nonetheless suggests some value in using spatial entropy as a useful measure of housing loss and neighborhood stability.

4.2 Edge cases

Based on the results, spatial entropy is negatively associated with no-structure rate. However, we did find some edge cases in our study area with both low no-structure rate and low spatial entropy. Fig. 6 shows the high-spatial-resolution image (NAIP images) of such neighborhoods. These neighborhoods all have relatively large parcels whose use may be commercial, industrial, or institutional. Such commercial, industrial, and/or institutional areas are very different from residential areas in terms of spatial configuration, so the entropy value in such neighborhoods should decrease. In this study, we included the covariate CII proportion to control this factor. However, if parcel data are not available in the area of interest, we may still try using other metrics like distance to CBD, or distance to roads as a potential substitute for commercial proportion. Thus, spatial entropy should still be a practical and effective metric for prediction of neighborhood stability.

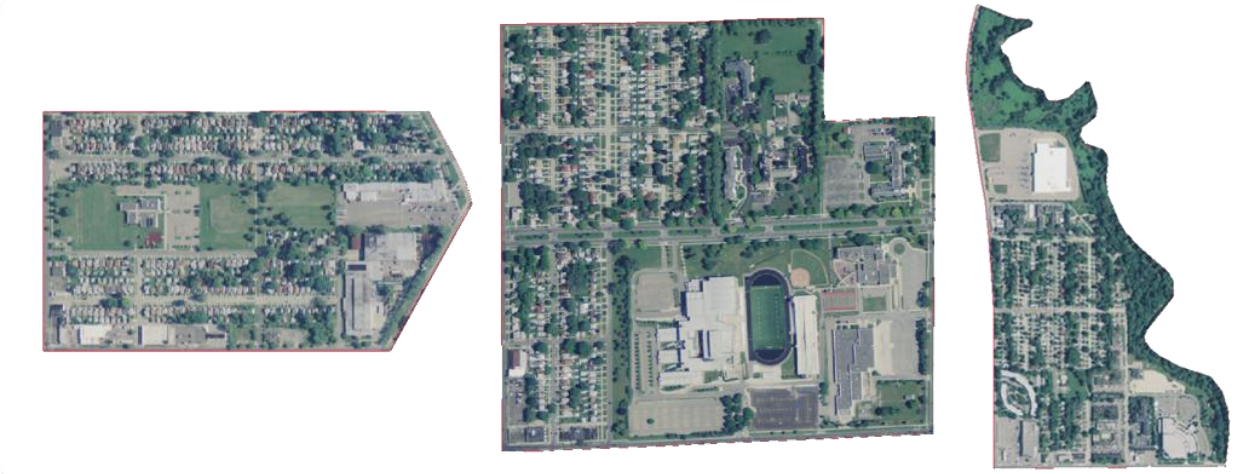

Figure 6. NAIP high-resolution images of 3 examples of "edge cases" neighborhoods both low no-structure rate and low spatial entropy.

4.3 Scale issue of spatial entropy

The scale issue is an interesting part of the use of spatial entropy in monitoring neighborhood shrinkage. Previous works have shown that the metric of entropy is scale-dependent, which means the size of pixels and the size of neighborhoods can both affect the results of spatial entropy (Batty et al. 2014). Table 3 shows how the model fit (measured by AIC) changes while tuning the size of pixels at both tract and block-group levels.

At the block-group level, the AIC values do not change much from the spatial resolution of 30m to 120m. However, when the spatial resolution goes to 250m, the AIC value significantly increases, which indicates a worse model fit.

Table 4 shows the average areas of census tracts and census block groups in Detroit City and the average numbers of pixels that are included in a neighborhood at multiple levels of spatial resolution. From Table 4 we can find that on average, each

neighborhood has less than seven pixels when the spatial resolution is 250m. We can also find that although spatial entropy is a scale-dependent metric, the model fit does not seem to be affected too much by pixel size unless the pixel size is relatively too large compared to the neighborhood size.

Table 4. Mean areas of all census block groups and census tracts in Detroit City, and mean # of pixels within a census block group and census tract at spatial resolutions of 30m, 60m, 120m, and 250m.

	Block Group	Tract
Mean area	409,234 m^2	1,160,375 m^2
Mean # of pixels–30m	448.7	1275.2
Mean # of pixels–60m	112.6	318.8
Mean # of pixels–120m	28.4	80
Mean # of pixels–250m	6.8	18.7

However, when using the 250m-resolution image, a neighborhood on average contains fewer than seven pixels at block-group level. Such neighborhoods with a really limited number of pixels will negatively affect the model fit. A previous study shows that the entropy value tends to increase quickly as the value of n increases when n is small (Batty et al. 2014). The rate of increase slows down when n gets larger. Thus, when neighborhood size is relatively small compared to the pixel size, the variation in neighborhood's entropy values may not mainly come from the difference in the characteristics of neighborhoods, but the difference in the number of pixels a neighborhood contains. This also explains why the total impact of spatial entropy at 250m and block-group level is not significant, because the correlation between spatial entropy and no-structure rate is weakened at the spatial resolution of 250m.

In terms of pixel size, we also hypothesize that the spatial resolution should not be too coarse. In this study, we are using spatial entropy to measure if urban vegetation is evenly distributed in a neighborhood. An image with coarser spatial resolution usually contains less information than one with finer resolution. When using larger pixels, it is similar to averaging the smaller pixels, which will decrease the variation of vegetation abundance in the pixels. Also, some uneven distribution of vegetation may only be observed at a finer spatial resolution. Table 3 also has similar results: the impact of spatial entropy does show a decreasing trend when pixel size gets larger.

It is also hypothesized that high spatial resolution images with too-small pixel sizes may bring in much noise and may not be suitable for calculating spatial entropy. We are also interested in how spatial entropy values would change as the spatial resolution of images gets finer than 30m. However, it is hard for us to experiment as 30m is the finest resolution we can get from the Landsat dataset.

Overall, to use spatial entropy as an indicator of urban shrinkage, it is better to use images with finer spatial resolution and make sure that each neighborhood contains an adequate number of pixels, so that the values of spatial entropy will not fluctuate too much.

4.4 Next Steps

While we find that spatial entropy is an effective indicator of neighborhood stability in Detroit City, our experiment is confined to a relatively small area. In future studies, we should experiment on whether the relationship between spatial entropy and

housing loss (no-structure rate) is consistent in other areas, like in suburban Detroit or other shrinking cities. Also, it will be interesting to do some time-series analysis, to see how spatial entropy changed over the years and try to understand how the shrinking process developed in the City of Detroit.

5 Conclusion

In this study, I estimated spatial lag models at tract and block-group levels to test if spatial entropy derived from moderate-resolution imagery is associated with a measure of neighborhood stability. Different from previous studies on urban change that used spatial entropy at the city level, we calculated spatial entropy values at the neighborhood level. Along with other demographic and parcel variables, the spatial lag models worked very well at both tract and block-group levels in predicting no-structure rates. Among all predictors, spatial entropy is the most impactful one, and even in models with the only one predictor being spatial entropy, the models can still perform well, which shows the importance of spatial entropy in predicting neighborhood stability. Thus, spatial entropy is an effective metric for monitoring the shrinkage of neighborhoods, and moderate-resolution imagery can be an effective source for monitoring and predicting urban shrinkage. Although spatial entropy is a scale-dependent metric, while allowing enough pixels in a neighborhood, the spatial resolution of remotely sensed images does not influence the models' fit much, and models can perform well at both tract and block-group levels.

References

Anselin, L. (1988). *Spatial econometrics: methods and models*. Springer Science & Business Media

Anselin, L., Bera, A.K., Florax, R., & Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional science and urban economics, 26,* 77-104

Banzhaf, E., Grescho, V., & Kindler, A. (2009). Monitoring urban to peri-urban development with integrated remote sensing and GIS information: a Leipzig, Germany case study. *International Journal of Remote Sensing, 30*, 1675-1696

Bartholomae, F., Woon Nam, C., & Schoenberg, A. (2017). Urban shrinkage and resurgence in Germany. *Urban Studies, 54*, 2701-2718

Batty, M., Morphet, R., Masucci, P., & Stanilov, K. (2014). Entropy, complexity, and spatial information. *Journal of geographical systems, 16*, 363-385

Bhatta, B., Saraswati, S., & Bandyopadhyay, D. (2010a). Quantifying the degree-of-freedom, degree-of-sprawl, and degree-of-goodness of urban growth from remote sensing data. *Applied Geography, 30*, 96-111

Bhatta, B., Saraswati, S., & Bandyopadhyay, D. (2010b). Urban sprawl measurement from remote sensing data. *Applied Geography, 30,* 731-740

Bivand, R.S., Pebesma, E.J., Gomez-Rubio, V., & Pebesma, E.J. (2008). *Applied spatial data analysis with R*. Springer

Brueckner, J.K. (2000). Urban sprawl: diagnosis and remedies. *International regional science review, 23,* 160-171

U.S. Census Bureau (2016). QuickFacts

Deng, C., & Wu, C. (2013). A spatially adaptive spectral mixture analysis for mapping subpixel urban impervious surface distribution. *Remote Sensing of Environment, 133*, 62-70

Fritsche, M., Langner, M., Köhler, H., Ruckes, A., Schüler, D., Zakirova, B., Appel, K., Contardo-Jara, V., Diermayer, E., & Hofmann, M. (2007). Shrinking cities–A new challenge for research in urban ecology. *Shrinking cities: Effects on urban ecology and challenges for urban development* (pp. 17-34): Peter Lang Publishing Group in association with GSE Research

Guerrieri, V., Hartley, D., & Hurst, E. (2012). Within-city variation in urban decline: The case of Detroit. *American Economic Review, 102*, 120-126

Haase, A., Rink, D., Grossmann, K., Bernt, M., & Mykhnenko, V. (2014). Conceptualizing urban shrinkage. *Environment and Planning A, 46*, 1519-1534

Haase, D. (2006). Derivation of predictor variables for spatial explicit modelling of 'urban shrinkage'at different scales. In, *45th Congress of the European Regional Science Association*

Haase, D., Haase, A., Kabisch, N., Kabisch, S., & Rink, D. (2012). Actors and factors in land-use simulation: The challenge of urban shrinkage. *Environmental Modelling & Software, 35*, 92-103

Hoekveld, J.J. (2014). Understanding spatial differentiation in urban decline levels. *European Planning Studies, 22*, 362-382

Jat, M.K., Garg, P.K., & Khare, D. (2008). Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. *International journal of Applied earth Observation and Geoinformation, 10*, 26-43

Ji, W., Ma, J., Twibell, R.W., & Underhill, K. (2006). Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics. *Computers, Environment and Urban Systems, 30*, 861-879

Kabisch, S., Haase, A., & Haase, D. (2006). Beyond growth–urban development in shrinking cities as a challenge for modeling approaches

Knaap, G.-J., Song, Y., & Nedovic-Budic, Z. (2007). Measuring patterns of urban development: new intelligence for the war on sprawl. *Local Environment, 12*, 239-257

Kumar, J.A.V., Pathan, S., & Bhanderi, R. (2007). Spatio-temporal analysis for monitoring urban growth–a case study of Indore city. *Journal of the Indian Society of Remote Sensing, 35*, 11-20

LeSage, J., & Pace, R.K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC

LeSage, J.P., & Fischer, M.M. (2008). Spatial growth regressions: model specification, estimation and interpretation. *Spatial Economic Analysis, 3*, 275-304

Li, J., Qiu, R., Xiong, L., & Xu, J. (2016). A gravity-spatial entropy model for the measurement of urban sprawl. *Science China Earth Sciences, 59*, 207-213

Li, X., & Yeh, A.G.-O. (2004). Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. *Landscape and Urban planning, 69*, 335-354

Liu, H., & Weng, Q. (2013). Landscape metrics for analysing urbanization-induced land use and land cover changes. *Geocarto International, 28*, 582-593

Neill, W.J. (2015). Carry on shrinking?: the bankruptcy of urban policy in Detroit. *Planning Practice and Research, 30*, 1-14

Padmanaban, R., Bhowmik, A.K., Cabral, P., Zamyatin, A., Almegdadi, O., & Wang, S. (2017). Modelling urban sprawl using remotely sensed data: A case study of Chennai city, Tamilnadu. *Entropy, 19*, 163

Qiu, F., Woller, K.L., & Briggs, R. (2003). Modeling urban population growth from remotely sensed imagery and TIGER GIS road data. *Photogrammetric Engineering & Remote Sensing, 69*, 1031-1042

Reis, J.P., Silva, E.A., & Pinho, P. (2016). Spatial metrics to study urban patterns in growing and shrinking cities. *Urban Geography, 37*, 246-271

Schwarz, N., Haase, D., & Seppelt, R. (2010). Omnipresent Sprawl? A Review of Urban Simulation Models with Respect to Urban Shrinkage. *Environment and Planning B: Planning and Design, 37*, 265-283

Siedentop, S., & Fina, S. (2010). Monitoring urban sprawl in Germany: towards a GIS-based measurement and assessment approach. *Journal of Land Use Science, 5*, 73-104

Small, C. (2002). Multitemporal analysis of urban reflectance. *Remote Sensing of Environment, 81*, 427-442

Small, C., & Milesi, C. (2013). Multi-scale standardized spectral mixture models. *Remote Sensing of Environment, 136*, 442-454

Viton, P.A. (2010). Notes on spatial econometric models. *City and regional planning, 870*, 2-17

Weaver, R.C., & Bagchi-Sen, S. (2013). Spatial analysis of urban decline: The geography of blight. *Applied Geography, 40*, 61-70

Weng, Q. (2001). Modeling urban growth effects on surface runoff with the integration of remote sensing and GIS. *Environmental management, 28*, 737-748

Xie, Y., Gong, H., Lan, H., & Zeng, S. (2018). Examining shrinking city of Detroit in the context of socio-spatial inequalities. *Landscape and Urban planning*

Yeh, A.G.-O., & Li, X. (2001). Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. *Photogrammetric engineering and remote sensing*